Globalization and the growing defects of international economic statistics

Linsi, L.; Mügge, D.K.

Link to publication

Routledge
Taylor & Francis Group

Globalization and the growing defects of international economic statistics

Lukas Linsi and Daniel K. Mügge

Department of Political Science, University of Amsterdam, Amsterdam, Netherlands

ABSTRACT

Official international economic statistics are generally considered accurate and meaningful gauges of cross-border flows of trade and capital. Most data users also assume that the quality of the underlying data keeps improving over time. Through an extensive review of the national accounting literature, archival research, two dozen interviews with high-level statisticians, and a series of data quality tests, we evaluate this common view for the primary source of data on trade and capital flows: the International Monetary Fund's Balance of Payments (BOP) Statistics. Our assessment paints a less rosy picture: reported figures are far less accurate than they are typically imagined to be and often do not correspond to the theoretical concepts with which users associate them. At the same time, measurement quality deteriorates over time as the transnationalization of economic production gradually undermines the validity of BOP statistics. Our findings raise serious questions about the widespread use of these numbers, with their deceptive pretense to accuracy, in scholarly research and public debate about the international political economy.

Introduction

Although the global economy is invisible to the naked eye, we discuss, research, and govern it day in, day out. To do so, we often rely on macroeconomic statistics—numbers about trade, inflation, economic growth, foreign direct investment (FDI), and so on. Much of what we know about the aggregate global economy, we know from spreadsheets that translate abstract concepts into concrete figures (Karabell, 2014; Hirschman and Berman, 2014).

Almost half a century ago, Sartori admonished researchers that "concept formation stands prior to quantification" (Sartori, 1970, p. 1038). Although comparative political methodologists have often heeded this advice and examined measurement inaccuracies and potential mismatches between their concepts and actual measures (e.g. Adcock

CONTACT Lukas Linsi ✉ l.a.linsi@uva.nl 📧 Department of Political Science, University of Amsterdam, Nieuwe Achtergracht 166, Amsterdam 1018WV, Netherlands

and Collier, 2001; Goertz, 2006), economic indicators have mostly escaped such scrutiny. Produced by government agencies, macroeconomic statistics—unlike, say, democracy indices—carry the authority of being "official" numbers. While most researchers realize that economic statistics are less than perfect (cf. Herrera & Kapur, 2007), users of statistics in policymaking, politics, and academia generally assume that the data are not too bad to begin with, and that they are improving.[1]

We argue that both assumptions are unwarranted. Analyses of error margins in international economic statistics, interviews with high-level statisticians, and archival records reveal that measurement uncertainty is worryingly large. We also find significant gaps between the concepts we wish to capture with international economic data and what figures in official databases actually measure. What we call the *concept-measurement gap* is big—and growing.

Indicators derived from a country's national accounts and balance of payments (BOP) depict distinct national economies interacting across clearly identifiable borders. But this neatly inter-national image corresponds less and less to the economic realities of the twenty first century (Dicken, 2015; Baldwin, 2016), when amorphous services trade, financial offshoring, mushrooming use of complex derivatives and intangible assets cloud measurement and undermine the concept validity of many indicators.[2] Despite capacity building efforts and drives towards international harmonization (cf. Mosley, 2003), measurement accuracy has hardly improved over the past decades.

As a result, the measurement quality of BOP statistics is deteriorating, and we cannot simply assume that the data suit our analytical purposes. Belying their clear separation in statistics, FDI flows are frequently impossible to distinguish from short-term capital flows; domestic sales can end up registered as cross-border services "trade"; foreign takeovers of domestic firms appear as portfolio capital "outflows", and so on. Given the stickiness of international statistical standards in the face of accelerating economic change, these problems are only likely to get worse.

The tension between an increasingly globalized economy and economic statistics that stick to an inter-national template chimes with scholarship challenging IPE to confront its epistemological pitfalls (cf. Farrell & Finnemore, 2009; Oatley, 2011). To be sure, data is never perfect, and it would be nonsense to suggest that nothing useful can be gleaned from official international statistics. As we argue below, the usefulness of data for example about trade or FDI clearly depends on the specific questions we ask as researchers. Yet the serious measurement problems that underlie these figures demand more careful attention than most analysts admit. Some deficiencies can be remedied sufficiently to render the data suitable for analytical purposes; we offer some examples in the concluding section. Other problems are too intractable to suggest obvious 'solutions'. It is then incumbent upon the users of the defective figures to argue why they are still useful. Rather than assuming that some numbers are better than no numbers, the burden of proof that data is appropriate for analytical ends rests with analysts themselves. As we show in this article, that principle also extends to official international economic statistics.

The measurement of economic life

The observation that there is more to quantitative data than meets the eye has a long history. Smith disparaged the fashionable *political artithmetick* of the

eighteenth century, arguing that data quality was too poor to allow solid conclusions and that the putative hardness of numbers concealed behind-the-scenes politicking (Dimand, 1995). In the 1940s, Simon Kuznets warned against reading too much into the national income indicator that he himself midwifed (Coyle, 2014; Fioramonti, 2014), while Oskar Morgenstern (1963) outlined the many limitations of popular macroeconomic measurements in his monograph *On the Accuracy of Economic Observations* (1963 [1950]).

Macroeconomic data have nevertheless become indispensable to economic policymaking and academic research, with their role in social and political life—like other types of contestable political measurement (Finnemore, 2013; Broome & Quirk, 2015; Kelley & Simmons, 2015; Snyder & Cooley, 2015; Broome, Homolar & Kranke, 2018)—triggering much criticism in recent scholarship. Gross domestic product (GDP) as a gauge for national welfare has attracted the most attention (Stiglitz, Sen, & Fitoussi, 2010; Lepenies, 2013; Fioramonti, 2014; Philipsen, 2015): critics have highlighted the gaps between casual, commonsense understandings of the measure and the narrowly economic dynamics GDP figures actually capture.[3] These criticisms have merit, but they focus on careless data *interpretation*, not on data problems themselves. Statisticians readily acknowledge that GDP is a production measure that may reveal little about societal welfare, let alone wellbeing or happiness (e.g. Lequiller and Blades, 2014). In what follows, we sidestep the question of whether these indicators chime with people's normative ambitions and ask instead whether they actually capture what they purport to do.

Data "quality" has multiple dimensions. For example, policymakers and investors often privilege timeliness (Biemer, Trewin, Bergdahl, & Japec, 2014); users interested in the quality of a data set as a whole will prize completeness. While we appreciate such general priorities, we are more directly concerned with the quality of economic measurement itself, comprised of two dimensions (cf. Goertz, 2006; Herrera & Kapur, 2007, p. 366). First, an indicator's *accuracy* points to (random and non-random) measurement errors. Second, the *concept-measurement gap* tracks how well the data corresponds to what the indicator purports to measure—whether what it says on the (statistical indicator) box accurately describes what's inside of it.

Empirically, we focus on the International Monetary Fund's BOP statistics. Its first *Balance of Payments Manual* (IMF, 1948) contained template tables for member countries to fill in. An expanded version with greater detail about what to include and exclude followed two years later (IMF, 1950). Since then, the enterprise has grown in size and ambition. While BOP statistics were originally collected "in whatever form the figures had been submitted" (IMF Archives, 1967, p. 3), international statisticians over the past decades have worked to harmonize statistical standards, building a sophisticated framework to integrate and systematize all BOP components and encouraging countries to follow the same data collection and presentation guidelines. The most recent, sixth edition of the BPM (IMF, 2009) differs from its earlier editions in both substance and style. It not only provides templates, but is organized as a didactic volume emphasizing the theoretical underpinnings and rules of the BOP system. While BPM1 was less than 50 pages, the latest version has grown into an authoritative document of nearly 400 pages, accompanied by a 600-page *Compilation Guide* (IMF, 2014).

Data users' views on the quality of BOP statistics

The IMF's BOP statistics are the source of data on international trade and capital flows that policymakers and researchers probably use most.[4] Although researchers concede, when pressed, that the data is far from perfect, how serious do they estimate the quality defects to be? To find out, we conducted an online survey among academic economists with a publication record in international economics.[5] Rather than aiming for full representativeness, our aim was simply to get a sense of the kinds and magnitude of problems that data users perceive, and whether they see these problems as decreasing or growing.

We presented the economists with a series of actual IMF BOP statistics from 2012 (a country's total imports of merchandise and services, bilateral merchandise imports from the USA, total inflows of foreign direct and portfolio investments) and asked them about their "intuitive best guess of the error margin inherent in this number". Half of the respondents, randomly selected, saw the figures for Sweden; the other half those for the Philippines.[6] Figure 1 shows that the majority of users consider the error margin to lie at about 5% for each BOP subcomponent. Changing the source of the data (Sweden or Philippines) did not substantively affect this general judgment.

We also asked respondents to indicate their level of agreement with the following statement: "The quality of international economic statistics has generally improved over the past 20 years." As Figure 2 shows, close to 90% of respondents either agreed or strongly agreed, revealing near consensus among academic economists that international economic data are improving over time.

Evaluating the measurement quality of BOP statistics

How do data users' estimates of error margins compare to actual data quality? And to what extent is data quality actually improving? To answer these questions, we performed a series of measurement quality tests, reviewed the technical literature on national accounting, consulted archival records, and conducted two dozen semi-structured interviews with high-level statisticians at international organizations and national statistical offices. Our findings contradict common assumptions about data quality: measurement errors are persistent and significantly larger than widely acknowledged, while economic globalization erodes the validity of BOP concepts and the measurement quality of BOP statistics.

Accuracy

All cross-border flows measured in BOP statistics are in principle recorded twice: once by the sending economy and once by the receiving one. Asymmetries between these two quantities—which in theory should be identical—can indicate measurement problems. Errors in subcomponents can cancel each other out at the aggregate level and transactions missed by both sender and receiver do not show up on either side. Mirror analyses therefore underestimate "actual" measurement errors. But they do suggest a lower-bound estimate of such problems and their evolution over time.
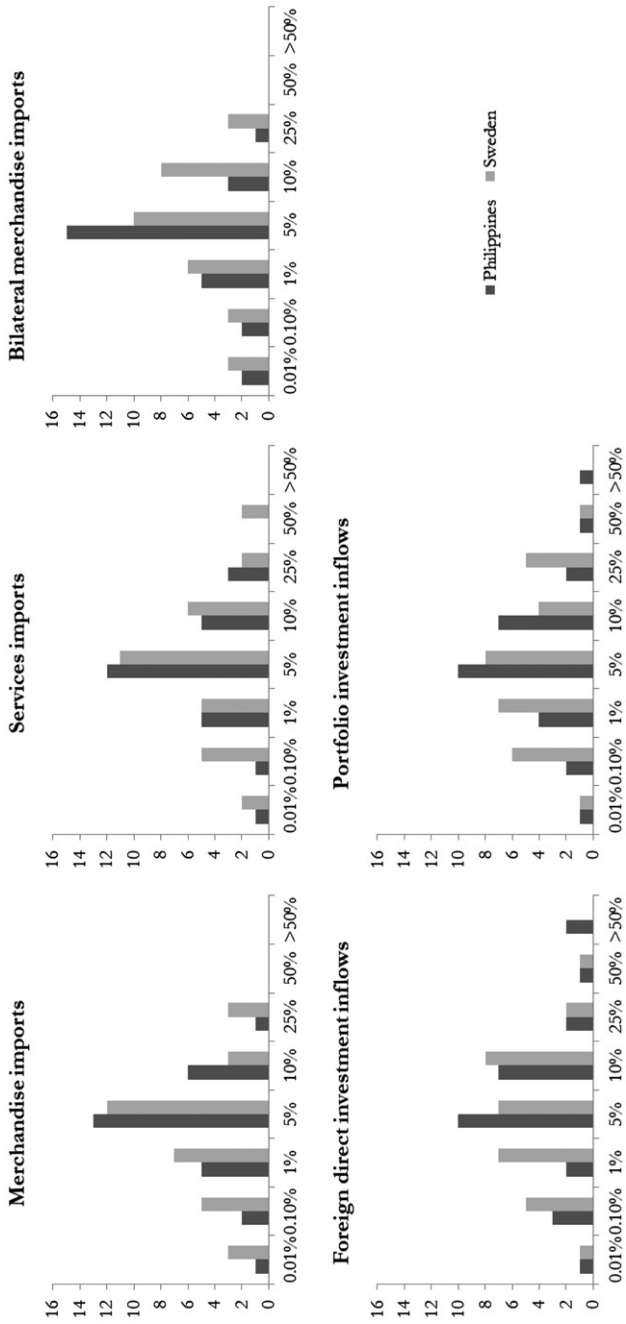
Figure 1. Perceived error margins in BOP statistics. Source: Own survey. Details in text.
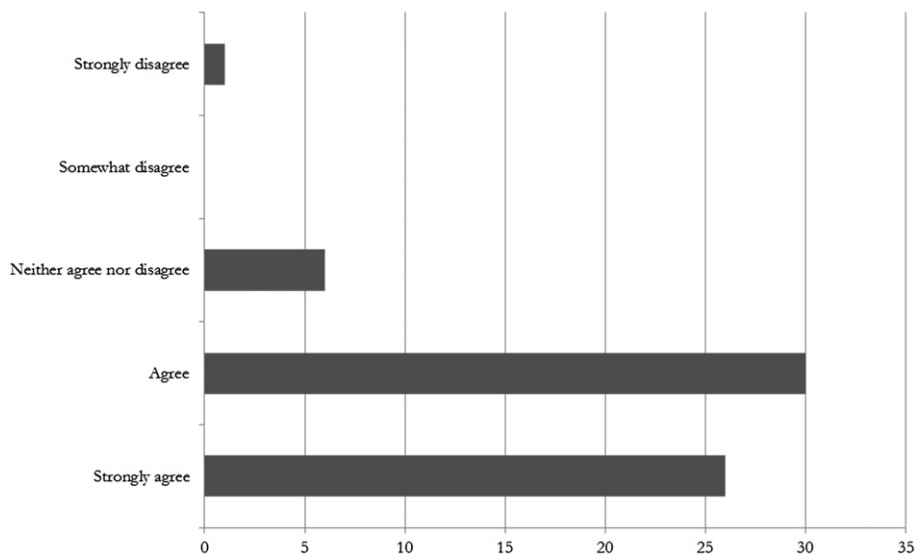
Figure 2. Data users' level of agreement with statement "The quality of international economic statistics has generally improved over the past 20 years". Source: Own survey. Details in text.

We use this approach for two complementary analyses: first, at the highest level of aggregation, we compare the size of reported total *global* inflows with reported total global outflows for four key BOP subcomponents: merchandise trade, services trade, FDI, and portfolio investments (PFI).[7] Second, we use the IMF's Direction of Trade Statistics (DOTS) to analyze *bilateral* asymmetries in merchandise trade statistics.

Figure 3 tracks absolute mirror asymmetries for merchandise trade, services trade, FDI, and PFI (bars) as well as how they compare to total reported inflows for each (line). Two aspects are noteworthy: first, despite decades of work by the IMF and others to align countries' methodologies, there is no indication of measurement errors decreasing. They may in fact be increasing. Second, we find marked differences between the various BOP subcomponents, belying users' sense that measurement errors are roughly similar across them (cf. Figure 2). They are much more sizable for FDI than for trade and stunningly large for PFI flows—where the discrepancy was nearly as large as total reported inflows in 2008 and 2011.

We performed a similar exercise for bilateral merchandise trade statistics, which are more developed than other bilateral data sets. The IMF's Direction of Trade Statistics database contains all monthly and annual data on bilateral merchandise trade flows reported by member countries since 1945.[8] We matched annual dyadic import and export records to calculate the reported trade flow from country A to B, first according to data from A, then from B. This allows us to calculate the mirror asymmetry between the two flows.

We dropped all dyad-years for which the IMF indicated the use of partner records to impute missing mirror values and ignored all dyadic observations in which one of the values is equal to zero to avoid an inflation of asymmetries (statistical offices sometimes substitute zero for missing values). This leaves us with 294,546 cases in which two countries have separately reported the same flow.

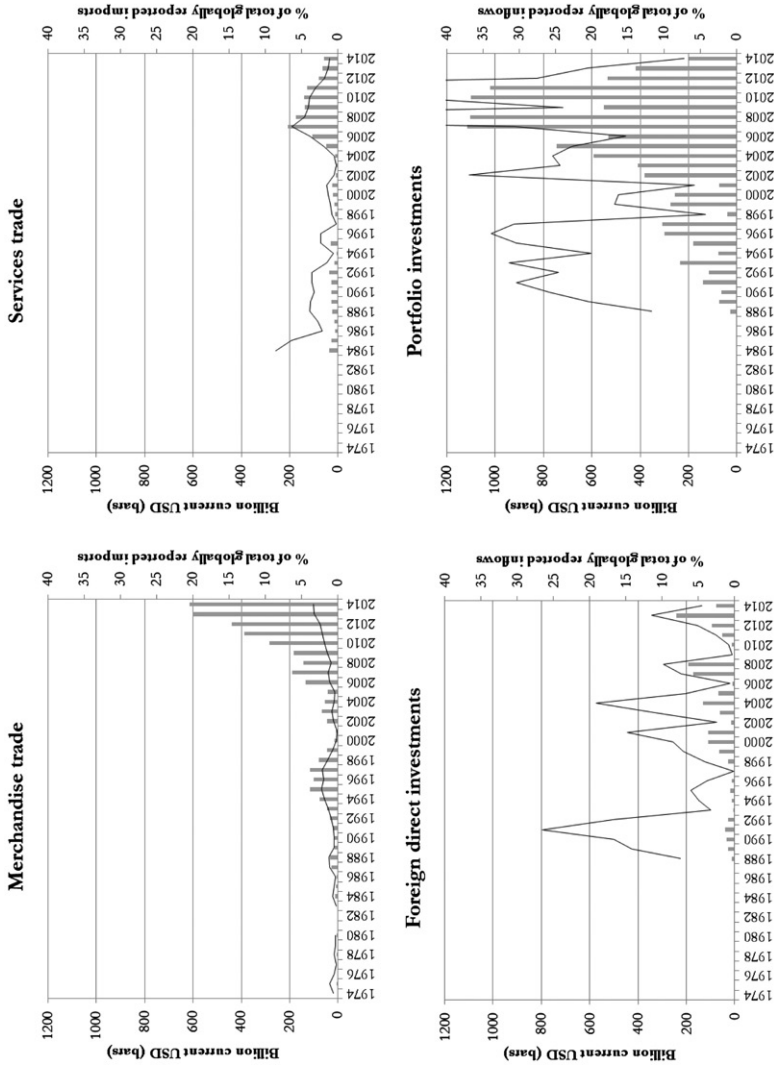**Figure 3.** The evolution of asymmetries in global mirror statistics over time. Source: Print version of IMF Balance of Payment Statistics Yearbooks 1981, 1990–1992, 1994–2007, 2011–2015. Note: Figures before 1984 are in billion SDRs, all others in billion USD. Reported numbers correspond to figures in YB closest to year of observation (i.e. figures for 2014 are from 2015YB, 2013 from 2014YB, etc.).
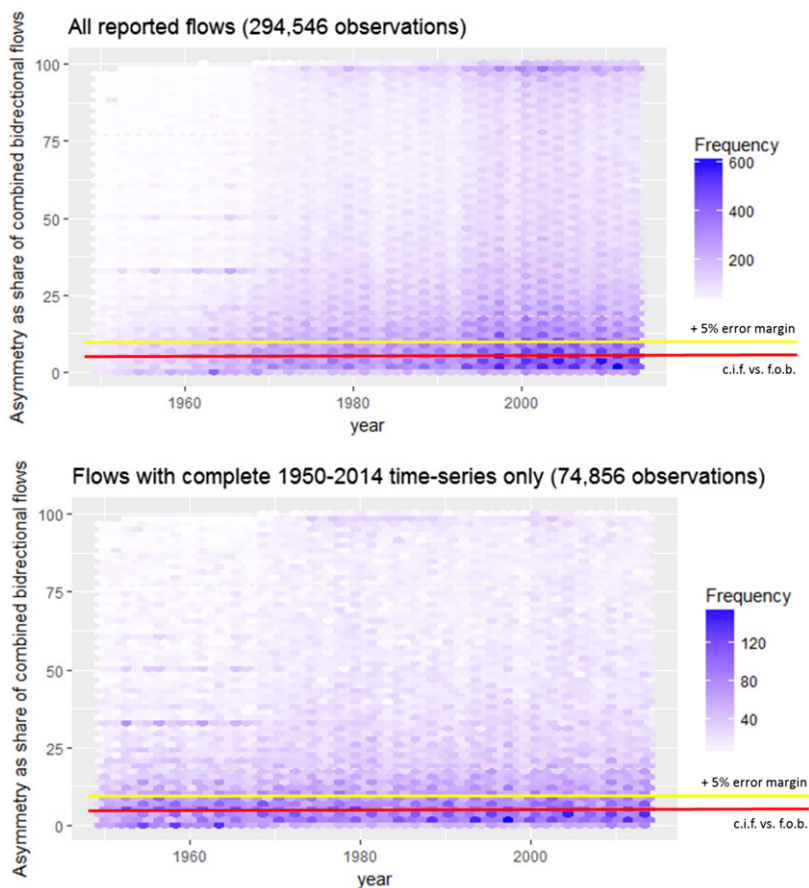
Figure 4. Mirror asymmetries in bilateral merchandise trade statistics. Source: IMF Direction of Trade Statistics Database. Further explanations in text.

To report the results, we create two high-density scatterplots: one for all reporters (top of Figure 4) and—to discount the consequences of the addition of new reporters over time—one for only those dyads that have consistently reported bilateral flows over the entire period (bottom of Figure 4). Lest outliers distort the graphical representation, we plot relative asymmetries as a share of combined flows (i.e. the sum of the flow from A to B reported by A and the one reported by B). This bounds the maximum size of the asymmetry at 100.

Import values typically include cost, insurance, and freight (c.i.f. valuation) while export values do not (free-on-board or f.o.b. valuation). Mirror flow values will therefore not be identical. But costs, insurance, and freight rarely exceed 10% of a good's value; in most cases it is substantially lower (Miao & Fortanier, 2017). The scatterplots include a line suggesting the error one might attribute to the c.i.f. vs. f.o.b. difference.[9] Another line highlights the 5% error margin suggested by users (cf. Figure 1 above).[10] The plots show that much of the asymmetry exceeds this "expected" range of error. Frequently, the differences between what A reports exporting to B and what B reports importing from A are stunning.

Table 1. US merchandise trade balance with its main trading partners according to US and partner country records in 2014.

Trade partner	A: Imports, US records	B: Imports, partner records	C: Exports, US records	D: Exports, partner records	US trade balance, US records (C–A)	US trade balance, partner records (D–B)	Absolute difference	Difference as % of combined trade flows, US records
China	444	397	124	146	−320	−251	69	12.1
Mexico	291	318	240	213	−51	−105	54	10.2
Canada	345	364	312	273	−33	−91	58	8.8
Germany	121	128	49	48	−72	−80	8	4.7
France	46	37	32	33	−14	−4	10	12.8
United Kingdom	53	60	54	52	1	−8	9	8.4

Note: All values in billion current USD. Source: Own calculations based on IMF DOTS (version downloaded on 15 March 2017), adjusted for trade-weighted cif-fob margins provided in Miao and Fortanier 2017: 18.

To illustrate these problems in concrete terms, we calculated the US trade deficit with several key trading partners in 2014 (cf. Table 1):[11] according to official US data, the American merchandise trade deficit with Mexico amounted to $51 billion; Mexican data put the figure roughly twice as high at $105 billion. The deficit with China reached almost $320 billion according to US authorities, but only $251 billion according to Chinese records. The US Census Bureau estimated the deficit with Canada to be $33 billion; Canadian data showed it to be over $91 billion. American authorities claim that imports from France exceeded US exports to that country by $14 billion, while French sources indicate the difference to be less than $4 billion, and so on.

The plots above show that discrepancies of such magnitude are not cherry-picked outliers; they are the rule rather than the exception. We find no indication that measurement errors are getting smaller over time. In the case of the USA—the most prominent trade deficit country—we might have expected political bias that would lead it to report higher deficits than its trading partners. But at least with the major US trading partners, this pattern does not hold.

Not surprisingly, the size of these discrepancies is so large that it can substantively affect findings of regression analyses using that data. While we address this specific issue—and strategies to mitigate these problems in econometric models—more systematically in separate work, one brief example from our replication efforts may be helpful to illustrate the seriousness of this issue for scholarly research.

Relying on bilateral IMF DOTS data, a widely cited study by Andrew Rose published in the *American Economic Review* (Rose, 2004)[12] reported the surprising finding that countries' accession to formal[13] GATT/WTO membership did not lead to any notable increase in trade, and that the effect may even be negative. We replicated Rose's finding with the latest version of the IMF DOTS data. We first followed standard practice in bilateral trade studies in using import data from both sides to measure bilateral flows (i.e. we use import records from A to proxy trade flows going from B to A, and import records from B to measure flows going in the opposite direction); then we ran exactly the same model, but use the export 'mirror' records (i.e. export records of A proxy flows from A to B and export records from B flows from B to A). The results are striking: while the import data confirms the puzzling negative association of formal GATT membership with

Table 2. The contrasting effects of formal GATT/WTO accession in mirror statistics.

	Bilateral trade data based on import records	Bilateral trade data based on export records
Both formal GATT/WTO members	−0.116 (0.0366)***	0.499 (0.0946)***
Only one formal GATT/ WTO member	−0.194 (0.0367)***	0.367 (0.0966)***
Control variables included?	Yes	Yes
Year-FE	Yes	Yes
Dyad-FE	No	No
N	198,144	198,144
R-square	0.65	0.38

Note: Robust standard errors in parentheses; *** $p < 0.01$; **$p < 0.05$; *$p < 0.1$; data and commands adapted from replication files provided by Goldstein, Rivers, and Tomz (2007); full results in Table 3 in Appendix.

bilateral trading flows, we find a strongly positive and statistically significant relationship when using the exports mirror (cf. Table 2; full results in Table 3 in Appendix). This observation neither suggests any deliberate misrepresentation of findings in previous studies, nor does it show the 'actual' effect of formal WTO accession to be positive. Instead, our re-analysis raises doubts about whether the available data is of sufficient quality in order to answer the empirical question at hand with as much confidence as previous studies have done.

Given the puzzling size and persistence of these measurement errors and their material importance for economic analyses, what are the forces underlying them?

The drivers of measurement inaccuracies

Political scientists often suspect deliberate data manipulation behind such inconsistencies. While data manipulation may certainly play a role (see Wallace, 2016; Kerner, Jerven, & Beatty, 2017), we see no indications of its systematic importance. Instead, measurement errors stem primarily from structural limitations to the harmonization of statistical practices and the growing complexity of economic processes.

Over the decades, international statistical communities have built an impressive intellectual framework supporting BOP statistics and have pushed hard for internationally harmonized concepts. But there are limits to the harmonization of actual statistical output. As a senior statistician explained to us (research interview with Fabienne Fortanier, Head of Trade Statistics at OECD Statistics Directorate, Paris, 6 June 2017):

> What you have to distinguish is that on the one hand you have the manuals, such as the SNA [Systems of National Accounts] 2008, or BPM6, which are conceptual manuals. They define the concepts, what's included and what's excluded, and how these concepts are related: for example, which elements add up to the current account balance. Or what transactions should be treated as a good or a service, etc. This is the international manuals. … But the compilation of the statistics is done nationally. And countries differ quite a bit in terms of the data sources and resources that are nationally available, in terms of their legal system and the legal context (…) in their methods for conducting surveys. (…) That is, the compilations of the data that underpin the concepts defined in the manuals differ across countries, which generates differences across countries. (…) So … the concepts are exactly the same, but the ways in which they are measured in practice can be different.

Even when national compilers agree on a common standard, implementation can diverge. International surveys on the collection of trade (United Nations Statistics Division, 2006), FDI (IMF & OECD, 2003), or PFI (IMF Statistics Department, 2000) data and bilateral reconciliation exercises reveal common challenges. Countries rely on different sources: some statistical offices have the legal powers to survey enterprises, others rely on subsamples of voluntary responses; some supplement customs data with administrative tax records, others do not. National compilers may interpret classifications differently, for example because they adhere to different editions of a statistical manual or because transactions fall into a gray area.[14] They may use dissimilar valuation techniques to estimate non-market asset values (for example for unlisted FDI; see Damgaard & Elkjaer, 2014). At-odds currency conversions or times at which transactions are recorded can further cloud statistics (United Nations Statistics Division, 2006), as well as unclear origins and destinations of merchandise that passes through several jurisdictions.

Such practical limitations cause substantial measurement errors, but they are only part of the story. The growing complexity of the global economy has enormously complicated the accurate recording of transactions (UNECE, Eurostat, & OECD, 2011). In 2015, the Federal Reserve Board commented on measurement problems in the US financial account (Federal Reserve Board of Governors, 2015):

> [T]he recent increase in statistical discrepancy most likely is the result of a shift in the sources of net financial inflows, from easier-to-measure purchases of securities by foreign official investors to activities across a range of instruments and by a range of private investors that in totality are more difficult to track.

Such dynamics affect all BOP components as ever-deeper global value and wealth chains (Gereffi, Humphrey, & Sturgeon, 2005; Seabrooke & Wigan, 2017) obscure national ownership. They spawn transactions at odds with the BOP's conceptual framework: merchanting trade, e-commerce, and capital flows channeled through impenetrable holding companies hidden in secrecy jurisdictions (Shaxson, 2012). Intangible assets, notoriously difficult to value for accounting purposes (Mügge and Stellinga, 2015; Bryan, Rafferty, & Wigan, 2017a), attract an important share of corporate profits. Financial liberalization and innovation have boosted global capital flows, packaged into ever more complex products (IMF, 1992). At the same time, budget cuts and eroding border controls have undermined traditionally important data collection systems, such as border customs inspections or exchange control systems (Ibid., p. 7).

Although international organizations have narrowed national compilers' room for interpretation and discretion in data gathering and reporting, the structural transformations outlined above have undercut economic measurement. Errors have persisted or grown worse despite ambitious harmonization programs. Already in 1966, the IMF's Assistant Chief of the BOP Division highlighted the challenge they pose (IMF Archives, 1966, p. 25):

> The fact that the statistics appear unreliable to an extent and in a manner that cannot always be fully assessed may in itself be a conclusion of considerable importance to analysts who are obliged to work with them. (…) [T]he best data now available are sometimes conflicting or otherwise obviously deficient and thus require cautious handling. (…) The interpretation of developments may be substantially affected by the choice made between alternative data sources and by the assumptions made about the causes of observed discrepancies.

To our mind, the admonition has lost nothing of its import. Yet as empirical researchers we too often disregard these problems, assuming that measurement errors are randomly distributed (leading, in the worst case, merely to attenuation bias). The review of the national accounting literature has shown that this is a dangerous assumption to make. Rather than being "random", measurement errors are too systematic to be ignored but not systematic enough to allow straightforward statistical treatment. Rather than being assumed away, they deserve our serious attention.

The concept-measurement gap

Measurement accuracy is obviously an important attribute of economic data. But for academic research, which frequently seeks to test theoretical arguments, validity problems are even more consequential. Irrespective of measurement accuracy, data will mislead academic inquiry if what it actually measures systematically differs from what it wants to capture. It is here that the globalization of economic activity is most worrisome.

Social scientists mostly use BOP data to study the determinants or effects of cross-border flows of goods, services, or capital. It entails something crossing a border in some meaningful sense, and often also a corresponding change in the nationality of asset ownership—say, bank deposits were "in" Germany and are now "in" Switzerland. Normally, data usage also implies that flows originate in reported sending country A and are destined to reported receiving country B.

But these "something moves from A to B" dynamics are not necessarily what BOP data record. Rather than aiming to identify the "nationality" of asset ownership, it uses the criterion of legal residence (IMF, 2009, p. 70–74). It also does not track flows from origins to ultimate destinations, but merely those among immediate partner countries.

These tensions are hardly new. Already in the 1950s, BOP technicians debated the treatment of "re-exporting" trade flows or how to assign capital flows routed via "paper companies" (IMF Archives, 1956, 1970). But the gaps between the common scholarly understandings of BOP concepts and the numbers entering BOP databases has widened significantly in recent years.

Merchandise trade statistics fail to distinguish clearly between the places of consignment of imports and exports and where goods are actually produced or consumed. In contrast to the fledgling WTO/OECD Trade in Value Added (TiVA) initiative, BOP statistics have traditionally treated every border crossing equally. As global production chains deepen, this statistical blend of conceptually distinct trade flows may increasingly distort the interpretation of trade data. For example, merchandise trade is commonly seen as an important dimension of economic interdependence, which may induce inter-country cooperation (cf. Farrell & Newman, 2014). But whether exports from A to B create meaningful interdependence depends on whether A actually produces the goods or merely passes them on.

BPM5 still counted goods that enter a country only for processing before onward shipment as 'conventional' imports. In BPM6, the IMF recommends ignoring the gross value of these flows and recording the processing fee in the trade in services accounts (UNECE et al., 2011, Chapter 5). If implemented, this approach would reveal a completely different image of world trade, with trading nations (in

contrast to those producing for export) becoming much smaller players in the global economy.

In the case of merchanting—transactions in which a resident entity re-sells a good acquired abroad in a third country, without the product ever physically entering the resident's economy—BPM6 recommends recording the difference between the gross export and import values in the goods account rather than the merchant's profits as a service export as in BPM5. (That said, because merchant resident countries struggle to detect flows that never physically enter the country, such activity often remains unrecorded; ibid., p. 85).

Taken together, the fragmentation of global production chains necessitates careful differentiation; whether one is interested in "gross" or "net" flows ultimately depends on the conceptual or theoretical question at hand. In any case, analysts need to assess whether the data suit their purposes, which may hinge on largely unheeded details such as whether a country follows BPM5 or BPM6.[15]

Statistics on services trade—which account for a continually increasing share of total global trade[16]—raise additional questions. Which activities should be included? Current standards aggregate four types of activity by mode of supply (WTO, 2017): cross-border delivery of services incorporated in physical products; consumption of non-residents while abroad (including tourism or foreign student tuition fees); provision of services via companies' foreign affiliates; and services provided internationally through the cross-border movement of natural persons (such as jet-setting consultants). Which of these should fall inside the researcher's purview? Mass-tourism is likely to have different political economy implications from, say, banking service provision through foreign affiliates. Researchers need to choose based on the question at hand; the choice should not be left to presentational conventions in statistical yearbooks.

Services trade statistics also struggle to distinguish actual "cross-national" transactions from MNE-internal accounting procedures. To minimize tax payments, multinational enterprises often create special purpose vehicles in low-tax jurisdictions where they "book" profits on intellectual property (Palan, Murphy, & Chavagneux, 2009; Shaxson, 2012). BOP statistics are based on an entity's formal legal residency rather than the nationality of its ultimate owners and hence do not adjust for the "re-routed" trade in services.[17] Large chunks of services "trade" may consist of purely domestic sales booked abroad for tax purposes—"phantom international [trade] flows" in the words of Robert Lipsey (2006). Without serious consideration of such issues, measures of cross-border flows risk to "lose their meaning" (ibid., p. 50).

While global corporate restructuring poses serious questions for trade statistics, its challenges to capital flow statistics are graver still. Statisticians have long struggled to distinguish long-term investments involving managerial control from short-term capital allocations of a more speculative nature (IMF Archives, 1956). In the 1980s, the IMF opted to err on the side of the latter with a "hard" threshold rule over national statisticians' qualitative judgments to distinguish FDI from PFI flows (Linsi, 2018). Since then, BOP statistics from most countries[18] classify cross-border investments of at least 10% of a company's equity as FDI; investments below that threshold are recorded as PFI. Although BOP technicians have debated the sensibility of a mechanical threshold rule to capture investment purpose since at least the 1950s (IMF Archives, 1956), the issue has become particularly acute

now that offshore holding structures increasingly obfuscate ownership (Haberly & Wojcik 2014, 2015; Garcia-Bernardo, Fichtner, Takes, & Heemskerk, 2017).

Recent estimates by the US Bureau of Economic Analysis (Ibarra-Caton & Mataloni, 2014) indicate that holding companies' share of the US outward direct investment position has grown from less than 10% in 1982 to close to 50% in 2012. Figures from Eurostat point into the same direction (Eurostat, 2016). This seriously challenges the usefulness of BOP FDI statistics, clouding not only the ultimate origin or destination but also the purpose of a majority of measured global FDI flows. We can no longer distinguish between long-term investments and the speculative investments of for example private equity or hedge funds (cf. Blanchard & Acalin, 2016). Funds may be destined for a recipient in a third country or be re-routed to the country of origin, for example for corporate inversions (which UNCTAD estimates to have accounted for nearly 20% of global FDI flows in 2015; UNCTAD, 2016, p. 3). For researchers interested in investment flows between countries, such issues (should) take center stage, as Andrew Kerner (2014) has shown through a replication exercise.

BOP data on PFI flows is plagued by similar issues. Complex chains of financial intermediaries distort the geographical image of short-term capital flows in favor of custodian centers such as Luxembourg and Switzerland—even when they are mere conduits and funds never "touch ground" in any meaningful way (Bertaut, Griever, & Tryon, 2006). It is simply unclear how to measure residents' equity and debt positions when assets and liabilities are concentrated in SPEs incorporated in offshore financial centers such as the Cayman Islands (Fichtner, 2016). The rapid growth of financial innovations since the 1980s (Miller, 1986; Scholes, 1998) have created intractable challenges for BOP statistics. Derivatives contracts have effectively turned "nationality" from a physical location into "a tradable attribute of an asset" (Bryan, Rafferty, & Wigan, 2017b, p. 52). Introduced to manage investors' exposure to one country's interest rate or exchange rate risks, derivatives are expressly designed to blur the "nationality" of financial products by driving a wedge between the financial product and the location of the underlying asset. It becomes impossible to distinguish "foreign" from "domestic" investments, and "long-term" from "short-term" ones. In the words of Peter Garber (1998: 33), "derivative products … make a mockery of the use of capital account categories".

Hard figures of the kind we find in statistical yearbooks do nothing to change these fundamental ambiguities; indeed, they suggest certainty where none exists. For instance, an early 2000s US Treasury analysis found that nearly two-thirds of total registered portfolio equity "outflows" from the US in the 1990s were in fact stock swaps resulting from foreign takeovers of US firms (Griever, Lee, & Warnock, 2001). More than half of the money that looked like foreign investments by US residents never left the US economy; it only entered official statistics that way because the US-based companies in which they were held changed legal residence.

In short, the de-nationalization of economic production and consumption and the growing complexity and opacity of corporate and financial structures have not only impaired progress towards the harmonization of statistical standards. Much more fundamentally, they have undermined the validity and hence usefulness of the statistical constructs themselves. Patterns of production, trade, and financial flows no longer conform to textbook images in which country A sends a

domestically produced good to country B and in return receives a payment that can be traced to consumers in that country. As multinational enterprises, obscure special purpose entities, highly fragmented production chains, and complex patterns of debts and credits proliferate, national accounting templates that assume simple economic relationships capture current realities less and less well.

Political implications

These growing defects of BOP statistics can be consequential for global politics in at least three ways. First, the apparent solidity of BOP statistics can mask the considerable uncertainty underlying them. It not only generates an unwarranted sense of confidence about our ability to monitor global economic transactions. It also bestows disproportionate power on those actors in the global economy whose authority relies on quantitative economic assessments. This concerns international organizations such as the IMF or the World Bank as much as credit rating agencies, whose data-based verdicts can shape nations' economic fortunes. Rising current account deficits, for example, are taken as unmistakable signs of economic troubles, just as waning external debt evidences a real improvement in the management of public finances. The possibility that such changes are grounded in nothing more than mismeasurements are rarely considered. Excessive faith in BOP statistics can also mute skepticism and undercut efforts to understand economic developments more thoroughly. For example, in the run-up to the global financial crisis financial regulators referenced (in fact inadequate) BOP data to argue that matters were under control—leaving "many countries … surprised by their vulnerability to collateralized securities and other risky instruments" (Moulton and van de Ven, 2018, p. 16) once things turned sour.

Second, the concept-measurement gap can distort policy analyses when the indicators feeding policy assessments don't neatly capture what policymakers think they do. Equating FDI with long-term greenfield investments, politicians and policymakers frequently draw on BOP statistics as a gauge of an economy's 'competitiveness'—even though present-day FDI statistics are more sensitive to corporate inversions, derivative constructions and tax avoidance schemes than to the building of new factories (Linsi, 2016). Trade statistics can tempt politicians to base their strategies on analyses of bilateral trading relationships—even if those fail to capture the difference between an actual exchange of national products and the mere passing on of third-country merchandise. The resulting absurdities surfaced when in 2017 both the US and the UK boasted a trade surplus with the other country (Romei and Cocco, 2017). Regulators may assess debt sustainability in light of 'national' savings, although derivative structures risk turning the latter into a largely meaningless concept (Bryan et al., 2017b, 56–57). Even matters at the heart of big powers' "grand strategies" can be affected by the concept-measurement gap. A question as central to international economic policy as the US economy's net position towards the rest of the world depends on valuation choices and the statistical operationalization of what counts as part of the "US economy": while official BOP measures show the US as a clear debtor, alternative valuations of foreign assets held by US investors might well turn it into a net creditor (Hausmann and Sturzenegger, 2007).

BOP-related measurement problems can spill over beyond the realm of international economic policy. For example, BOP statistics' inability to track intra-firm profit-shifting can severely undermine estimates of industrial production and hence productivity (Guvenen, Mataloni, Rassier, & Ruhl, 2017). Ireland's official 2015 GDP growth rate of 26% was fueled by the restructuring of foreign multinationals headquartered there, triggering controversial debates about the meaningfulness of GDP figures in a globalized economy (Boland, 2017). The Irish example is extreme, bordering on the ridiculous. But it highlights a more pervasive problem: BOP statistics systematically attribute economic production and investment to jurisdictions in ways that are beneficial to the owners of capital. They therefore hide actual value-creation and productive capacity especially in poor, heavily export-dependent countries and recast economic relationships borne out of power differentials as objective economic realities (cf. Smith 2012).

Third, most worryingly, distorted analyses can feed misguided policy responses. Local content requirements imposed by trade negotiators may have unintended consequences if they lack a clear view of how such regulations ripple through the supply chains. Credit rating agencies that build country risk assessments on skewed current account figures (Sinclair, 2005; Afonso, Gomes, & Rother, 2007) can distort governments' access to global capital markets. The prominent role of low-quality 'external sector' statistics in IMF analyses of countries' economic performance (Bryan et al., 2017b, p. 56) can turn mismeasurements into investment decisions. And the reliance of BOP statistics can motivate governments under speculative attack to fire their bullets in the wrong direction. To quote Peter Garber once more (Garber, 1998, p. 2–3):

> In the presence of derivatives, … [balance of payments] data can generate false inferences about the sources of a crisis and lead to misinformed policy prescriptions. They confound the sources of the crisis: whether it stems from foreign speculators, panicked green-screen traders, or domestic insiders armed with knowledge about weak fundamentals. In addition, in the presence of large volumes of derivatives, claims that crises are generated by such inappropriate policies as an excessively short maturity of the public debt can be mirages of on-balance sheet accounting.

Because the errors in BOP data are so manifold and stem from diverse sources, specific political implications will depend on the case in question. The pointers above should make clear, however, that while not universal in their effect, the political implications of defective BOP statistics are too momentous to be disregarded.

Conclusions

Although users of BOP statistics are aware that they are not perfect, measurement problems are widely seen as minor. As we have shown from various angles, such optimism is unwarranted. Measurement quality of official BOP statistics is low, and economic globalization is further undermining the ability of nationally based statistics to capture economic activity in a meaningful way.

This is worrying. Credit rating agencies, investors, and international organizations rely on macroeconomic data in their country assessments and surveillance, often with material consequences for those countries (cf. Mosley, 2000). Such data can also inform international judicial deliberations, for example in WTO arbitration panels, and carry hard-wired legal consequences. They feed analyses by policy

analysts and journalists, nurturing constructions of broader narratives about macroeconomic trends and development trajectories. And every now and then, BOP figures become directly politicized, for example in spats about American trade relations with China, Mexico, or Germany, or the trading position of Germany within the European Union.

Ignorance of measurement problems is just as problematic within academia itself. As researchers, we frequently build strong causal inferential claims, disregarding that measurement uncertainty may easily be large enough to make the difference between statistically significant and insignificant findings (cf. Manski, 2015). Our replication example with bilateral trade data – like other replication studies with GDP (Johnson, Larson, Papageorgiou, & Subramanian, 2013) or FDI (Kerner, 2014) data—shows how much-cited research results are sensitive to measurement errors and ambiguities in conceptual definitions.

We do not seek to indict quantitative scholarship per se. Rather, we believe it is a sign of disciplinary maturity to look squarely at the limitations of our data and decide for what purposes and with what caveats we can plausibly use them. It is therefore regrettable that warnings about data quality have generally fallen on deaf ears. The superficial precision of economic statistics obscures that they remain "human-made estimates … not true values" (research interview with senior WTO statistician, Geneva, 22 August 2017). Nearly all statisticians we interviewed advised cautious interpretation of international economic statistics. In their minds, the primary goal of their work is not to enable academic researchers to draw statistical inferences at some threshold of statistical significance, but "to give policymakers a merely descriptive picture of broad trends" (ibid.).

Statistical compendia do not trumpet data problems on their covers. That said, footnotes or appendices frequently do mention data limitations, even if obliquely. Statisticians are certainly aware of them and would, we learned in our interviews, tackle them today rather than tomorrow were it possible. Yet invariably, the problems defy easy solutions. The mismatch between a globalized economy and the statistics that depict it in inter-national terms is here to stay. Commitment to international harmonization means that existing statistical standards are hard to amend.[19] Even when definitional and conceptual issues are less thorny, building new data sets requires heroic effort and a great deal of time. Statistical standards will thus always lag behind developments in the real economy. The more rapidly the economy changes, the larger the gap becomes (ibid.).

Just as statisticians have no easy fixes for the problems we have outlined, there are no off-the-shelf solutions for the academic users of international economic data. There are strategies to mitigate some problems: rather than sweep mirror asymmetries under the carpet, we can leverage them to improve inferences by running models using both import and export data to evaluate the robustness of results; or, alternatively, we might include the difference between the two as an independent variable to model measurement errors explicitly. We can and should establish how sensitive analytical results are to the exclusion of well-known trading hubs, offshore financial centers, nodes for SPEs or countries with low statistical capacity. US BEA data on the size of American MNCs' staff abroad or fixed capital expenditures by majority-owned affiliates offer valuable checks of BOP FDI data (cf. Kerner, 2014), and so on.

While such remedies may improve data sufficiently for some analytical ends, they do not solve deeper underlying issues. The politics behind the costly collection of economic statistics mean that nearly all figures are compiled by agencies that have a mandate to produce 'national' data. But intense and highly complex economic cross-border interactions sit uneasily with nationally based data, and it is frequently not clear whether 'nation-states' are the most appropriate unit of analysis for studies of the international political economy.[20] In addition, some measurement problems go beyond technical and logistical challenges. Many economic stocks and transactions do not have obvious values, of the kind that we could readily read off a price tag attached to international merchandise trade. Even when such monetary values are available, we can legitimately ask whether they represent some inherent value of the good in question or an actor's ability to appropriate surplus (Smith, 2012). Either way, the growing defects of international economic data call for greater data skepticism and urges IPE researchers to strengthen the inferential foundations of quantitative analyses through routinized and careful examination of whether official statistics suit our analytical ends.

Funding

Notes

1. A notable exception is Kerner (2014).
2. Our analysis resonates with that of Bryan, who in 2001 already noted the contradiction between globalization on the one hand and an increasing importance attached to national economic performance on the other (Bryan, 2001).
3. For example: the exclusion of unpaid labor from GDP, its ignorance of environmental destruction, or its inability to capture people's "happiness."
4. They are also the main source for trade and capital flow statistics disseminated through the World Bank's World Development Indicators (WDI) database. Nearly two-thirds of the academic economists we surveyed for this article indicated WDI as the database which they most frequently use for research purposes.
5. We originally sent the survey in July/August 2017 to the 441 authors of all journal articles published between 2015 and 2017 indexed in the American Economic Association's *EconLit* database, with a joint entry in either JEL codes F14 and F21, or F21 and F32. We received 71 complete answers.
6. We chose these two countries to evoke images of "typical" advanced/developing economies. We do not believe there are strong *a priori* reasons for respondents to adopt extreme views on the quality of statistics these countries produce.
7. In case of missing reported values, global estimates (provided separately in the BOP Yearbooks) use imputed data from other sources to ensure consistency in the number of reporters on both sides. Research interview with IMF statisticians, Washington D.C., 19 September 2017.
8. In most recent years, the database covers bidirectional merchandise trading flows among 150–170 countries.
9. Note that a 10% higher valuation of imports vs. exports corresponds to a roughly 5% difference in *combined* flows.
10. A 5% underestimation of a trade flow on one side and a simultaneous overestimation of 5% on the other side would result in a 5% difference in combined flows.

11. To net out valuation differences, we convert import data to f.o.b. values using dyad-specific c.i.f.-f.o.b. margins from the OECD. The margins are 1% in US trade with Canada and Mexico; 2% with Western European countries; and 5% with China. See Miao & Fortanier, 2017, p. 18.
12. We work with the replication materials provided by Goldstein, Rivers, and Tomz (2007) and base our analysis on their replication of Rose's findings, which they present in Model 1, Table 1, page 53.
13. The study by Goldstein, Rivers, and Tomz (2007) challenges Rose's findings through the introduction of a more fine-grained categorization of countries' GATT/WTO membership.
14. For instance, is the purchase of an e-book from a foreign provider to be classified as a "good" or "service" import? Cf. Ward, 2004.
15. Analysis by the Dallas FED suggests, for instance, that correcting trade balances for value added reduces the US trade deficit with China in 2009 by 33%, from USD 189 to 126 billion. Sposi & Koech, 2013.
16. A recent paper estimates the volume of services exports as a share of total exports having increased from less than 10% in 1970 to close to 20% in 2014 (Loungani, Mishra, Papageorgiou, & Wang, 2017, p. 8).
17. A recent analysis by Goldman Sachs suggests the US trade deficit to shrink from 3 to 1.5% of GDP if trade figures are adjusted for profit-shifting (Smith, 2017).
18. In practice, a few countries still use other (usually higher) thresholds (IMF & OECD, 2003).
19. The OECD's Trade in Value-Added (TiVA) data will be very welcome as an attempt to depict value-creation more accurately. It remains to be seen, however, whether it succeeds to put some of the hard conceptual conundrums to rest.
20. A dilemma that goes back to the earliest days of national economic statistics when the founding fathers of GDP grappled with this difficult question: "Should it be individual entrepreneurs? Climate zones? Ethnic subgroups? Economic social classes? Religious denominations? Kuznets rejected all these options in favor of the nation-state because the available data were organized and maintained by sovereign states." (in Fogel, Fogel, Guglielmo, & Grotte, 2013, p. 67)

Acknowledgments

References

Adcock, R., & Collier, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review*, 95(03), 529–546.

Afonso, A., Gomes, P., & Rother, P. (2007). What "hides" behind sovereign debt ratings? *ECB Working Paper Series*, No. 711. https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp711.pdf?6dcb8c9f1f85e3a95e9687ddd523a402.

Baldwin, R. (2016). *The great convergence: Information technology and the new globalization*. Cambridge, MA: Harvard University Press.

Bertaut, C. C., Griever, W. L., & Tryon, R. W. (2006). Understanding U.S. Cross-Border Securities Data. Federal Reserve Bulletin. http://ticdata.treasury.gov/Publish/frbul2006.pdf.

Biemer, P., Trewin, D., Bergdahl, H., & Japec, L. (2014). A system for managing the quality of official statistics. *Journal of Official Statistics*, 30(3), 381–415.

Blanchard, O., & Acalin, J. (2016). What does measured FDI actually measure? Policy Brief, Peterson Institute for International Economics. https://piie.com/system/files/documents/pb16-17.pdf.

Boland, V. (2017). Ireland's 'de-globalised' data calculate a smaller economy. *Financial Times*. 18 July.

Broome, A., Homolar, A., & Kranke, M. 2018. Bad science: International organizations and the indirect power of global benchmarking. *European Journal of International Relations*, *24*(3), 514–539.

Broome, A., & Quirk, J. (2015). Governing the world at a distance: The practice of global benchmarking. *Review of International Studies*, *41*(05), 819–841.

Bryan, D. (2001). Global accumulation and accounting for national economic identity. *Review of Radical Political Economics*, *33*(1), 57–77.

Bryan, D., Rafferty, M., & Wigan, D. (2017a). Capital unchained: Finance, intangible assets and the double life of capital in the offshore world. *Review of International Political Economy*, *24*(1), 56–86. doi:10.1080/09692290.2016.1262446

Bryan, D., Rafferty, M., & Wigan, D. (2017b). From Time-Space Compression to Spatial Spreads. In AndrewLeyshon and BrettChristophers (Eds.), *Money and finance after the crisis: Critical thinking for uncertain times* (pp.43–67). New York: Wiley.

Coyle, D. (2014). *GDP: A brief but affectionate history*. Princeton, N.J.: Princeton University Press.

Damgaard, J., & Elkjaer, T. (2014). Foreign direct investment and the external wealth of nations: How important is valuation? *Review of Income and Wealth*, *60*(2), 245–260. doi:10.1111/roiw.12098

Dicken, P. (2015). *Global shift: Mapping the changing contours of the world economy*. New York: Guilford Press.

Dimand, R. (1995). "I have no great faith in Political Arithmetick." Adam Smith and quantitative political economy. In IngridRima (Ed.) *Measurement, quantification and economic analysis. Numeracy in economics* (pp.22–30). London: Routledge.

Eurostat. (2016). Special Purpose Entities within EU Direct Investment Statistics. 29th Meeting of the IMF Committee on Balance of Payments Statistics. https://www.imf.org/external/pubs/ft/bop/2016/pdf/16-05.pdf.

Farrell, H., & Finnemore, M. (2009). Ontology, methodology and causation in the American School of International Political Economy. *Review of International Political Economy*, *16*(1), 58–71. doi:10.1080/09692290802524075

Farrell, H., & Newman, A. L. (2014). Domestic institutions beyond the nation-state: Charting the new interdependence approach. *World Politics*, *66*(02), 331–363. doi:10.1017/S0043887114000057

Federal Reserve Board of Governors. (2015). The USA's plans and difficulties in Compiling the CPIS Enhancements. 28th Meeting of the IMF Committee on Balance of Payments Statistics. https://www.imf.org/external/pubs/ft/bop/2015/pdf/15-09d.pdf.

Fichtner, J. (2016). The anatomy of the Cayman Islands offshore financial center: Anglo-America, Japan, and the role of hedge funds. *Review of International Political Economy*, *23*(6), 1034–1063. doi:10.1080/09692290.2016.1243143

Finnemore, M. (2013). Constructing Statistics for Global Governance. Unpublished manuscript, presented at the Yale University Macmillan International Relations Workshop, 13 February.

Fioramonti, L. (2014). *How numbers rule the world: The use and abuse of statistics in global politics*. London: Zed Books.

Fogel, R. W., Fogel, E. M., Guglielmo, M., & Grotte, N. (2013). *Political arithmetic: Simon Kuznets and the empirical tradition in economics*. Chicago: University of Chicago Press.

Garber, P. (1998). Derivatives in International Capital Flow. NBER Working Paper Series, No. 6623.

Garcia-Bernardo, J., Fichtner, J., Takes, F. W., & Heemskerk, E. K. (2017). Uncovering offshore financial centers: Conduits and sinks in the global corporate ownership network." *Scientific Reports 7*.

Gereffi, G., Humphrey, J., & Sturgeon, T. J. (2005). The governance of global value chains. *Review of International Political Economy*, *12*(1), 78–104. doi:10.1080/09692290500049805

Goertz, G. (2006). *Social science concepts: A user's guide*. Princeton, N.J.: Princeton University Press.

Goldstein, J., Rivers, D., & Tomz, M. (2007). Institutions in international relations: Understanding the effects of the GATT and the WTO on world trade. *International Organization*, *61*(1), 37–67.

Griever, W. L., Lee, G. A., & Warncock, F. E. (2001). The U.S. system for measuring cross-border investment in securities: A primer with a discussion of recent developments. Federal Reserve Bulletin, October.

Guvenen, F., Mataloni, R. J., Rassier, D. G., & Ruhl, K. R. (2017). Offshore profit shifting and domestic productivity measurement. NBER Working Paper Series, No. 23324.

Haberly, D., & Wojcik, D. (2014). Regional blocks and imperial legacies: Mapping the global offshore FDI network. *Economic Geography*, *91*(3), 251–280. doi:10.1111/ecge.12078

Haberly, D., & Wojcik, D. (2015). Tax havens and the production of offshore FDI: An empirical analysis. *Journal of Economic Geography*, *15*(1), 75–101. doi:10.1093/jeg/lbu003

Hausmann, R., & Sturzenegger, F. (2007). The missing dark matter in the Wealth of Nations and implications for global imbalances. *Economic Policy*, *22*(51), 470–518. doi:10.1111/j.1468-0327.2007.00182.x

Herrera, Y. M., & Kapur, D. (2007). Improving data quality: Actors, incentives, and capabilities. *Political Analysis*, *15*(04), 365–386. doi:10.1093/pan/mpm007

Hirschman, D., & Berman, E. P. (2014). Do economists make policies? On the political effects of economics. *Socio-Economic Review*, *12*(4), 779–811. doi:10.1093/ser/mwu017

Ibarra-Caton, M., & Mataloni, R. J. Jr. (2014). "Direct investment position for 2013." https://bea.gov/scb/pdf/2014/07 July/0714_direct_investment_positions.pdf.

IMF; OECD. (2003). Foreign direct investment statistics: How countries measure FDI. http://www.imf.org/external/pubs/ft/fdis/2003/fdistat.pdf.

IMF Statistics Department (2000). *Analysis of 1997 coordinated portfolio investment survey results and plans for the 2001 survey*. Washington, D.C.: IMF.

International Monetary Fund. (1948). *Balance of payments manual* (1st ed.). Washington, D.C.: IMF.

International Monetary Fund. (1950). *Balance of payments manual* (2nd ed.). Washington, D.C.: IMF.

International Monetary Fund. (1992). *Report on the measurement of international capital flows*. Washington, D.C.: IMF.

International Monetary Fund. (2009). *Balance of payments and international investment position manual*. Washington, D.C.: IMF.

International Monetary Fund. (2014). *BPM6 compilation guide: Companion document to the sixth edition of the balance of payments and international investment position manual*. Washington, D.C.: IMF.

IMF Archives. (1956). "Verbatim report of the international monetary fund meeting of fund statistical correspondents on balance of payments discussions held at the Burgundy Room – Sheraton-Park Hotel, Washington, D.C. on Thursday, September 27, 1956 at 2:30 p.m." Assistant Director Arie Bouter Files, Box #1 File #1, IMF Archives.

IMF Archives. (1966). Research and Statistics Department, Assistant Director Arie Bouter Files, Box # 1 File #8. John S. Smith, "Asymmetries and Errors in Reported Balance of Payments Statistics", DM/66/67, December 5,

IMF Archives (1967). Research and Statistics Department, Assistant Director Arie Bouter Files, Box # 1 File #8. John Alves, "Progress Towards Uniformity in Balance of Payments Presentation", DM/67/26, April 19,

IMF Archives (1970). Research Department, Balance of Payments Division Subject Files, Box #2 File #9. "Report on Meeting of Balance of Payments Experts", April 16,

Johnson, S., Larson, W., Papageorgiou, C., & Subramanian, A. (2013). Is newer better? Penn world table revisions and their impact on growth estimates. *Journal of Monetary Economics*, *60*(2), 255–274. doi:10.1016/j.jmoneco.2012.10.022

Karabell, Z. (2014). *The leading indicators: A short history of the numbers that rule our world*. New York: Simon & Schuster.

Kelley, J. G., & Simmons, B. A. (2015). Politics by number: Indicators as social pressure in international relations. *American Journal of Political Science*, *59*(1), 55–70. doi:10.1111/ajps.12119

Kerner, A. (2014). What we talk about when we talk about foreign direct investment. *International Studies Quarterly*, *58*(4), 804–815. doi:10.1111/isqu.12147

Kerner, A., Jerven, M., & Beatty, A. (2017). Does it pay to be poor? Testing for systematically underreported GNI estimates. *The Review of International Organizations*, *12*(1), 1–38. doi: 10.1007/s11558-015-9239-3

Lepenies, P. (2013). *Die Macht der einen Zahl: eine politische Geschichte des Bruttoinlandsprodukts. Politische Geschichte des Bruttoinlandsprodukts*. Erste Aufl. Berlin: Suhrkamp.

Lequiller, F., & Blades, D. (2014). *Understanding national accounts*. Paris: OECD.

Linsi, L. (2016). Less compelling than it seems: rethinking the relationship between aggregate FDI inflows and national competitiveness. *Columbia FDI Perspectives,* No. 184, October 10. http://ccsi.columbia.edu/files/2013/10/No-184-Linsi-FINAL.pdf

Linsi, L. (2018). Foreign Direct Investment. In K.Dingwerth and C.Weinhardt (EDs.), *The language of world trade politics: Unpacking the terms of trade* (pp.50–63). London: Routledge.

Lipsey, R. E. Measuring International Trade in Services. (2006). *National Bureau of Economic Research Working Paper Series* No. 12271. http://www.nber.org/papers/w12271%5Cnhttp://www.nber.org/papers/w12271.pdf.

Loungani, P., Mishra, S., Papageorgiou, C., & Wang, K. (2017). *World trade in services: Evidence from a new dataset*. IMF Working Papers. Washington, D.C. https://www.imf.org/~/media/Files/Publications/WP/2017/wp1777.ashx.

Manski, C. F. (2015). Communicating uncertainty in official economic statistics: An appraisal fifty years after Morgenstern. *Journal of Economic Literature*, *53*(3), 631–653. doi:10.1257/jel.53.3.631

Miao, G., & Fortanier, F. (2017). Estimating transport and insurance costs of international trade. OECD Statistics Working Papers, no. 4. http://dx.doi.org/10.1787/18152031.

Miller, M. H. (1986). Financial innovation: The last twenty years and the next. *Journal of Financial and Quantitative Analysis*, *21*(4), 459–471. doi:10.2307/2330693

Morgenstern, O. (1963). *On the accuracy of economic observations* (2nd ed.). Princeton, NJ: Princeton University Press.

Mosley, L. (2000). Room to move: International financial markets and national welfare states. *International Organization*, *54*(4), 737–773. doi:10.1162/002081800551352

Mosley, L. (2003). Attempting global standards: National governments, international finance, and the IMF's data regime. *Review of International Political Economy*, *10*(2), 331–362.

Moulton, B. R., & van de Ven, P. (2018). Addressing the challenge of globalization in national accounts. Paper presented at CRIW Conference. Bethesda, MD, March 9–10. http://papers.nber.org/conf_papers/f100570/f100570.pdf

Mügge, D. K., & Stellinga, B. (2015). The unstable core of global finance: Contingent valuation and governance of international accounting standards. *Regulation & Governance*, *9*(1), 47–62.

Oatley, T. (2011). The reductionist gamble: Open economy politics in the global economy. *International Organization*, *65*(02), 311–341.

OECD. (2016). Reducing Discrepancies and Asymmetries: What Can Direct Investment and Trade in Services Statisticians Learn from Each Other? 29th Meeting of the IMF Committee on Balance of Payments Statistics. https://www.imf.org/external/pubs/ft/bop/2016/pdf/16-11.pdf.

Palan, R., Murphy, R., & Chavagneux, C. (2009). *Tax Havens: How Globalization Really Works*. Ithaca, N.Y.: Cornell University Press.

Philipsen, D. (2015). *The Little Big Number: How GDP Came to Rule the World and What to Do about It*. Princeton, N.J.: Princeton University Press.

Romei, V., & Cocco, F. (2017). UK and US report trade surplus with each other. Financial Times. 24 September. https://www.ft.com/content/82ebed88-9ede-11e7-8cd4-932067fbf946.

Rose, A. K. (2004). Do we really know that the WTO increases trade? *American Economic Review*, *94*(1), 98–114.

Sartori, G. (1970). Concept misformation in comparative politics. *American Political Science Review*, *64*(04), 1033–1053.

Scholes, M. S. (1998). Derivatives in a dynamic environment. *The American Economic Review*, *88*(3), 350–370.

Seabrooke, L., & Wigan, D. (2017). The governance of global wealth chains. *Review of International Political Economy*, *24*(1), 1–29.

Shaxson, N. (2012). *Treasure islands: Uncovering the damage of offshore banking and tax havens*. New York: Palgrave Macmillan.

Sinclair, T. (2005). *The new masters of capital: American bond rating agencies and the politics of creditworthiness*. Ithaca, NY: Cornell University Press.

Smith, J. (2012). The GDP illusion: Value added versus value capture. *Monthly Review*, *64*(3), 86–102. doi:10.14452/MR-064-03-2012-07_6

Smith, N. (2017). Suddenly, America's trade deficit isn't so awful. Bloomberg Business Week, 22 December 2017.

Snyder, Jack L., & A.Cooley (Eds.) (2015). *Ranking the world: Grading states as a tool of global governance*. Cambridge, UK: Cambridge University Press.

Sposi, M., & Koech, J. (2013). Value-added data recast the U.S.-China trade deficit. *Dallas FED: Economic Letter*. http://www.dallasfed.org/assets/documents/research/eclett/2013/el1305.pdf.

Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2010). *Mismeasuring our lives: Why GDP doesn't add up*. New York: New Press.

UNCTAD. (2016). *World investment report 2016 - investor nationality: Policy challenges*. Geneva: United Nations Publications.

UNECE, Eurostat, and OECD. (2011). *The impact of globalization on national accounts*. New York: 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.

Wallace, J. L. (2016). Juking the stats? Authoritarian information problems in China. *British Journal of Political Science*, *46*(01), 11–29. doi:10.1017/S0007123414000106

Ward, M. (2004). *Quantifying the world: UN ideas and statistics*. Bloomington: Indiana University Press.

WTO. (2017). Basic purpose and concepts. *GATS Training Module*. https://www.wto.org/english/tratop_e/serv_e/cbt_course_e/c1s3p1_e.htm. [last accessed: 11 October 2017]