

Fickle Formulas

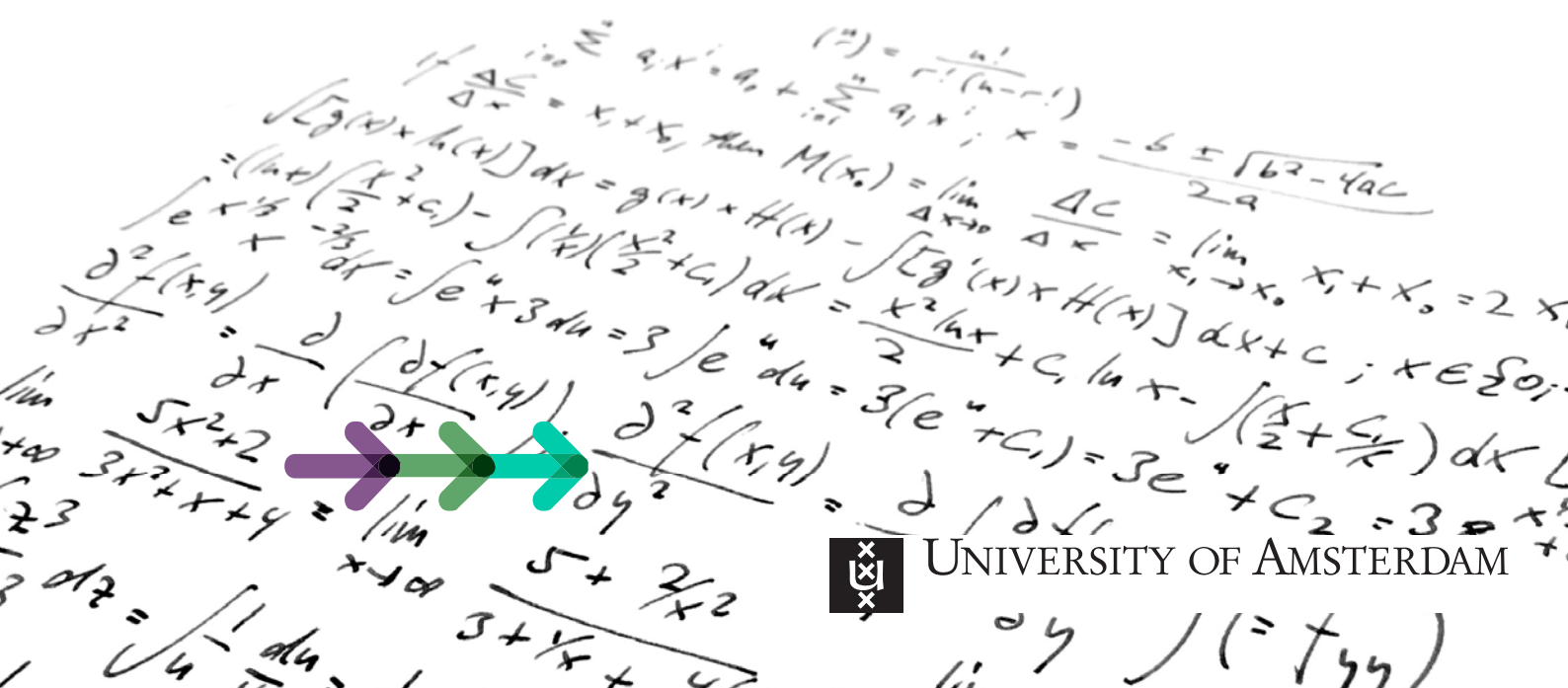
Globalization and the Growing Defects of International Economic Statistics

Working paper 2-2017.

Lukas Linsi and Daniel Mügge

University of Amsterdam

December 2017



Lukas Linsi

Department of Political Science
University of Amsterdam

l.a.linsi@uva.nl

Daniel K. Mügge

Department of Political Science
University of Amsterdam

d.k.muegge@uva.nl

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 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, with potentially serious implications for empirical research using this data. Our analysis identifies the principal reasons for these worrying trends and concludes with a first set of suggestions on how to address them in our research designs.

Keywords: Economic Measurement; Balance of Payments; Trade Flows; Capital Flows; Globalization.

Word count: 7,658



Introduction¹

Although the global economy is invisible to the naked eye, we discuss, research, and govern it day in, day out. To do so, we rely on macroeconomic statistics—numbers about trade, inflation, economic growth, foreign direct investment (FDI), and so on. What we know about the aggregate global economy, we know from spreadsheets that translate abstract concepts into concrete figures.²

Almost half a century ago, Sartori admonished researchers that “concept formation stands prior to quantification”.³ Although comparative political methodologists have generally heeded this advice and examined measurement inaccuracies and potential mismatches between their concepts and actual measures,⁴ economic indicators have generally 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,⁵ 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.⁶

¹ Earlier versions of this article have been presented at Georgetown, Cornell, the Enlighten/Fickle Formulas workshop in Santpoort, at EPSA2017, SASE2017, and IPES2017. For very helpful feedback we thank Peter Katzenstein, Raymond Mataloni, Boris Samuel, Mark Schwartz, Svend-Erik Skaaning, and our PETGOV colleagues at the University of Amsterdam (with special thanks to Brian Burgoon, Bart Stellinga, and Jonathan Zeitlin). Hanna Dose has provided valuable research assistance. Takeo David Hymans has edited and greatly clarified our prose. Our greatest thanks go to the many statisticians who generously shared their insider perspectives. This research was supported by the ERC Starting Grant FICKLEFORMS (grant # 637883) and the NWO Vidi project 016.145.395. Details are available on www.fickleformulas.org.

² Karabell 2014; Hirshman and Popp Berman 2014.

³ Sartori 1970, 1038.

⁴ Adcock and Collier 2001; Goertz 2006.

⁵ Cf. the critique by Herrera and Kapur (2007).

⁶ Kerner (2014) is an exception with his meticulous and critical discussion of FDI data.



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 data in official databases actually measure. What we call the *concept-measurement gap* is large and growing.

Statistical 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 21st century,⁷ when amorphous services trade, financial offshoring, and intangible assets cloud measurement and undermine the concept validity of many indicators. And despite capacity building efforts and drives towards international harmonization⁸, 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 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.

Far from a wholesale indictment of international economic statistics, we argue for greater awareness of problems with the data and their responsible use. Our analysis thus chimes with scholarship that has pushed the discipline to engage with

⁷ Dicken 2015.

⁸ Cf. Mosley 2003.



epistemological pitfalls that IPE analyses confront.⁹ We conclude this research note with three suggestions for data users: to buttress the robustness of our inferences, we need to familiarize ourselves more with specific measures used as proxies for broader concepts and better understand their limitations. Second, where alternative measurement approaches are available, we need to establish and explicitly argue which one best fits our hypotheses. And third, we need to conduct data sensitivity analyses to assess whether errors in specific BOP data series can be assumed to be randomly distributed, and think collectively about how to handle systematic biases in the data in cases in which they are not.

The measurement of economic life

The observation that there is more to quantitative data than meets the eye has a long history. Adam Smith disparaged the fashionable *political arithmetick* of the 18th century, arguing that data quality was too poor to allow for solid conclusions and that the putative hardness of numbers concealed behind-the-scenes politicking.¹⁰ In the 1940s, Simon Kuznets warned against reading too much into the national income indicator that he himself midwived,¹¹ while Oskar Morgenstern outlined the many limitations of popular macroeconomic measurements in his 1950 monograph *On the Accuracy of Economic Observations*.¹²

Macroeconomic data have nevertheless become indispensable to economic policy-making and academic research, with their role in social and political life like other types

⁹ Cf. Farrell and Finnemore 2009; Oatley 2011.

¹⁰ Dimand 1995.

¹¹ Coyle 2014; Fioramonti 2014.

¹² Morgenstern 1963 [1950].



of contestable political measurement¹³ triggering much criticism in recent scholarship. Gross domestic product (GDP) as a gauge for national welfare has attracted the most attention:¹⁴ critics have highlighted the gaps between casual, commonsense understandings of the measure and the narrowly economic dynamics GDP figures actually capture.¹⁵ 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.¹⁶ 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;¹⁷ users interested in the quality of a data set as a whole will prize completeness.¹⁸ While we appreciate such general priorities, we are more directly concerned with the quality of economic measurement itself, comprised of two dimensions.¹⁹ 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.

¹³ Finnemore 2013; Kelley and Simmons 2015; Snyder and Cooley 2015.

¹⁴ Stiglitz et al. 2010; Lepenies 2013; Fioramonti 2014; Philipsen 2015.

¹⁵ For example: the exclusion of unpaid labor from GDP, its ignorance of environmental destruction, or its inability to capture people's “happiness.”

¹⁶ E.g. Lequiller and Blades 2014.

¹⁷ Biemer et al. 2014.

¹⁸ For an insightful analysis of the politics of data *availability*, see Hollyer et al. 2011.

¹⁹ Goertz 2006; Herrera and Kapur 2007, 366.



Empirically, we focus on the International Monetary Fund's BOP statistics. Its first *Balance of Payments Manual* (BPM1, 1948)²⁰ contained template tables for member countries to fill in. An expanded BPM version with greater detail about what to include and exclude followed two years later.²¹ Since then, the enterprise has grown in size and ambition. While BOP statistics were originally collected "in whatever form the figures had been submitted,"²² 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 (2009) differs from its earlier editions in both substance and style.²³ It not only provides templates, but is organized as a didactic volume emphasizing the theoretical underpinnings and rules of the BOP system. While the 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*.²⁴

Data users' view of 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.²⁵ Although researchers concede, when pressed, that the data is far from perfect, how serious do they estimate the quality

²⁰ IMF 1948.

²¹ IMF 1950.

²² IMF Archives 1967, 3.

²³ IMF 2009.

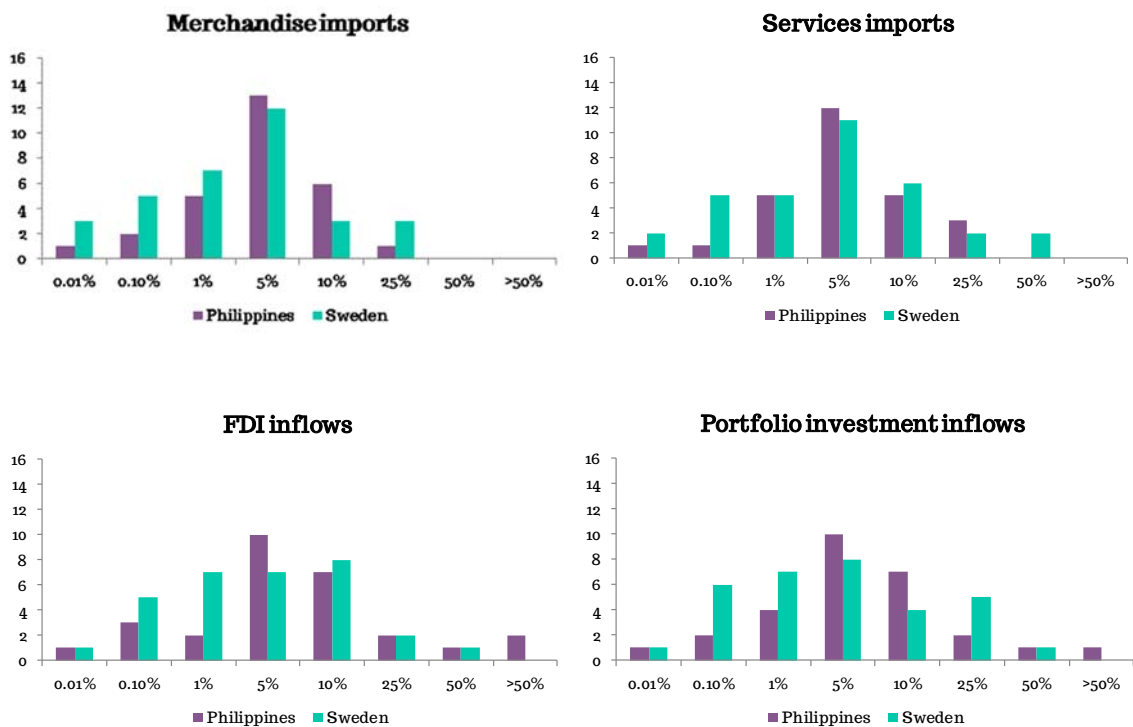
²⁴ IMF 2014.

²⁵ 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.



defects to be? To find out, we conducted an online survey among academic economists with a publication record in international economics.²⁶ 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.

Figure 1. Perceived error margins in BOP statistics



SOURCE: Own survey, details in text.

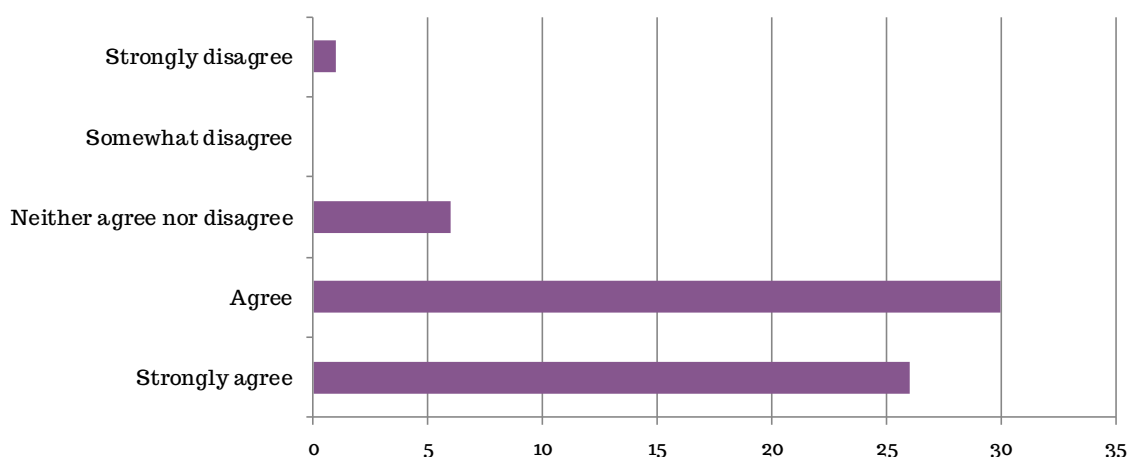
²⁶ We originally sent the survey in July/August 2017 to the 441 authors of all journal articles published between 2015 and 2017 indexed in the American Economic Association’s *EconLit* database, with a joint entry in either JEL codes F14 and F21, or F21 and F32. We received 71 complete answers.



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, for the Philippines.²⁷ Figure 1 shows that the majority of users consider the error margin to lie at about 5 percent for each BOP subcomponent. Changing the source of the data (Sweden or Philippines) did not substantively affect this general judgment.

Figure 2. Data users’ sense of quality developments

"Has the quality of international economic statistics improved over the past 20 years?"



SOURCE: Own survey, details in text.

²⁷ 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.



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 percent 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 balance of payments statistics

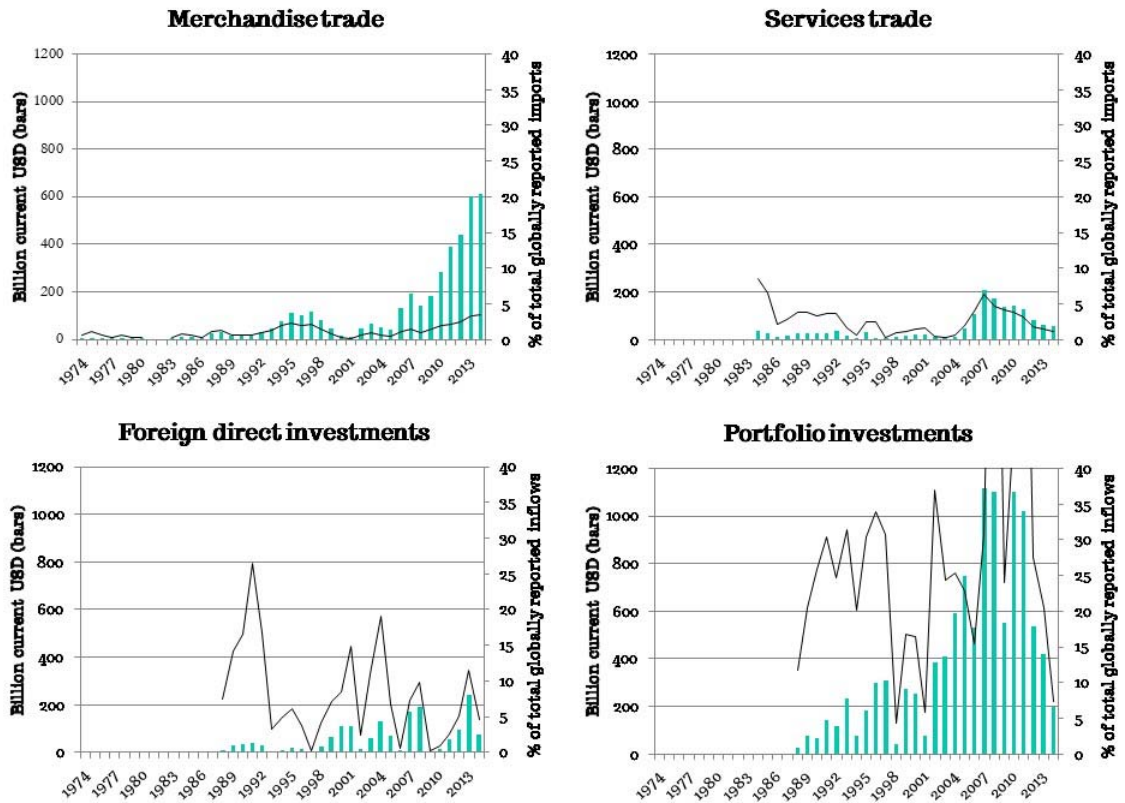
How do data users’ estimates of error margins compare to actual data quality? And to what extent is data quality actually improving? To answer these questions, we performed a series of measurement quality tests, reviewed the technical literature on national accounting, consulted archival records, and 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 globalization and the digitization of economic activity are eroding 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 3. The evolution of asymmetries in global mirror statistics over time



SOURCE: IMF Balance of Payment Statistics Yearbooks 1981, 1990-1992, 1994-2007, 2011-2015.

NOTE: Figures before 1984 are in billion SDRs, others in billion USD. Reported numbers are taken from YB closest to observation year (2014 figures are from 2015 YB, 2013 from 2014 YB, etc.)

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 subcomponents²⁸ (merchandise trade, services trade,

²⁸ In case of missing reported values, global estimates (provided separately in the BOP Yearbooks) use imputed data from other sources to match the number of reporters on both sides. Research interview with IMF statisticians, Washington D.C., 19 September 2017.



FDI, and portfolio investments (PFI)). 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 contain all monthly and annual data on bilateral merchandise trade flows reported by member countries since 1945.²⁹ 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.

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,

²⁹ In recent years, the database covers bidirectional merchandise trade among 150-170 countries.



one for only those dyads that have reported bilateral flows over the entire period (bottom). Lest outliers distort the graphical representation, we plot relative asymmetries relative to 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 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 thus not be identical. But costs, insurance, and freight rarely exceed ten percent of a good's value; often it is substantially lower.³⁰ The scatterplots include a line suggesting the error users might attribute to the c.i.f. vs. f.o.b. difference.³¹ Another line highlights the five percent error margin suggested by users (cf. Figure 1).³² The plots show that much of the asymmetry exceeds this "expected" error range. Frequently, the gaps between what A reports exporting to B and what B reports importing from A are stunning.

To illustrate these problems concretely, we calculated the 2014 US trade deficit with several key trading partners (cf. Table 1): according to official US data, the American merchandise trade deficit with Mexico amounted to \$54 billion; Mexican data put the figure roughly twice as high at \$104 billion. The deficit with China reached almost \$343 billion according to US authorities, but only \$243 billion according to Chinese records. The US Census bureau estimated the deficit with Canada to be \$35 billion; Canadian data showed it to be over \$88 billion. American authorities claim that imports from France exceeded US exports to that country by \$15 billion, while French sources indicate the difference to be less than \$3 billion, etc.

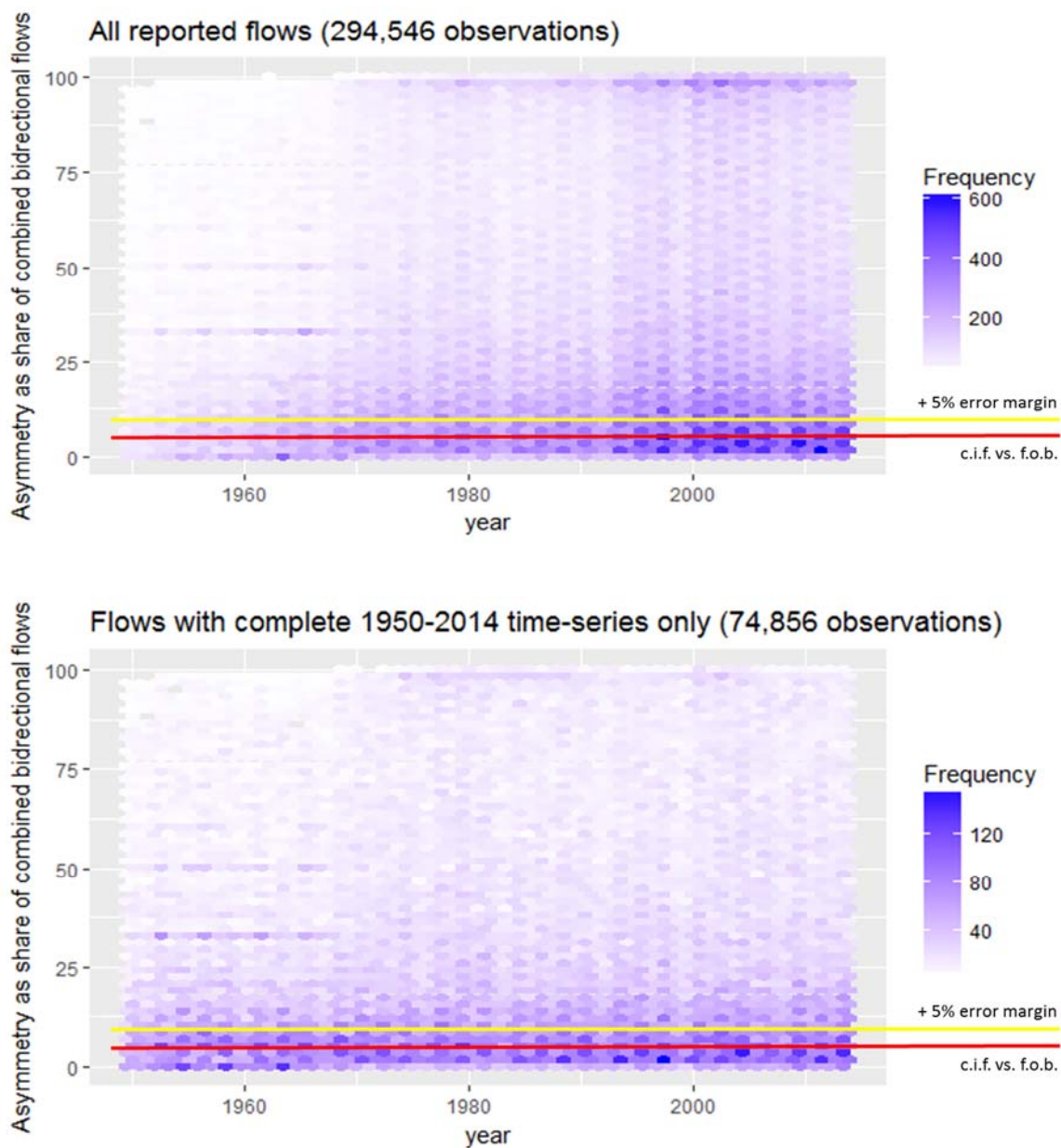
³⁰ Miao and Fortanier 2017.

³¹ This 10 percent higher valuation translates into roughly 5 percent for *combined* flows.

³² A 5 percent underestimation of a trade flow on one side and a simultaneous overestimation of 5 percent on the other side would result in a 5 percent difference in combined flows.



Figure 4. High density scatterplots of the relative size of mirror asymmetries in bilateral merchandise trade statistics



SOURCE: IMF Direction of Trade Statistics Database. NOTE: Further explanations in text.



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	467	397	124	154	-343	-243	100	16.9
Mexico	294	318	240	215	-54	-103	49	9.2
Canada	348	364	312	276	-36	-88	52	7.9
Germany	123	128	49	49	-74	-79	5	2.9
France	47	37	32	34	-15	-3	12	15.2
United Kingdom	54	60	54	53	0	-7	7	6.5

NOTE: All values in billion current USD. SOURCE: Own calculations based on IMF DOTs (version downloaded on 15 March 2017)

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 let it report higher deficits than its trading partners. But at least with the major US trading partners, this pattern does not hold. How can we explain the size of these measurement errors and their persistence?

The drivers of measurement inaccuracies

Political scientists often suspect deliberate data *manipulation* behind such inconsistencies. While data manipulation may certainly play a role,³³ 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.

³³ See, for example, Wallace 2016; Kerner et al. 2017.



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 at the OECD explained to us:

What you have to distinguish is that on the one hand you have the manuals, such as the SNA 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 compilation 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.³⁴

Even when national compilers agree on a common standard, implementation can diverge. International surveys on the collection of trade,³⁵ FDI,³⁶ or PFI³⁷ 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

³⁴ Research interview with Fabienne Fortanier, Head of Trade Statistics at OECD Statistics Directorate, Paris, 6 June 2017.

³⁵ United Nations Statistics Division 2006.

³⁶ IMF and OECD 2003.

³⁷ IMF Statistics Department 2000.



statistical manual or because transactions fall into a gray area.³⁸ They may use dissimilar valuation techniques to estimate non-market asset values (for example for unlisted FDI).³⁹ At-odds currency conversions or times at which transactions are recorded can further cloud statistics,⁴⁰ 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.⁴¹ In 2015, the Federal Reserve System commented on measurement problems in the US financial account:

[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.⁴²

Such dynamics affect all BOP components as ever-deeper *global value⁴³ and wealth⁴⁴ chains* 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.⁴⁵ *Intangible assets*, notoriously difficult to value for accounting purposes,⁴⁶ attract an important

³⁸ 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.

³⁹ Damgaard and Elkjaer 2014.

⁴⁰ United Nations Statistics Division 2006.

⁴¹ UNECE et al. 2011.

⁴² Federal Reserve Board 2015.

⁴³ Gereffi et al. 2005.

⁴⁴ Seabrooke and Wigan 2017.

⁴⁵ Shaxson 2012.

⁴⁶ Mügge and Stellinga 2015; Bryan et al. 2017.



share of corporate profits. *Financial liberalization and innovation* have boosted global capital flows, packaged into ever more complex products.⁴⁷ 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.⁴⁸

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. In 1966, the IMF's Assistant Chief of the Balance of Payments Division highlighted the challenge they pose:

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.⁴⁹

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 statistical literature has shown that this is a dangerous assumption. Rather than being "random", measurement errors are too systematic to be ignored but

⁴⁷ IMF 1992.

⁴⁸ *Ibid.*, 7.

⁴⁹ IMF Archives 1966, 25.



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 and digitization 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*.⁵⁰ 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.”⁵¹ But the gaps between the common scholarly understandings of BOP

⁵⁰ IMF 2009, 70-74.

⁵¹ IMF Archives 1956, 1970.



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 fledging 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.⁵² 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.⁵³ 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 *merchanted* 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

⁵² Farrell and Newman 2014.

⁵³ UNECE et al., chap. 5.



detect flows that never physically enter the country, such activity often remains unrecorded).⁵⁴

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 BPM6.⁵⁵

Statistics on services trade – which account for a continually increasing share of total global trade⁵⁶ – raise additional questions. Which activities should be included? Current standards aggregate four types of activity by mode of supply:⁵⁷ 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.

⁵⁴ Ibid., 85.

⁵⁵ 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 percent, from USD 189 to 126 billion. Sposi and Koech 2013.

⁵⁶ A recent paper estimates the volume of services exports as a share of total exports having increased from less than ten percent in 1970 to close to twenty percent in 2014. Loungani et al. 2017, 8.

⁵⁷ WTO 2017.



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.⁵⁸ 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. Large chunks of services “trade” may consist of purely domestic sales booked abroad for tax purposes “phantom international [trade] flows”⁵⁹ in the words of Robert Lipsey. Without serious consideration of such issues, measures of cross-border flows risk to “lose their meaning.”⁶⁰

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.⁶¹ 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. Since then, BOP statistics from most countries⁶² classify cross-border investments of at least 10 percent 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

⁵⁸ Palan et al. 2009; Shaxson 2012.

⁵⁹ Lipsey 2006.

⁶⁰ Ibid., 50.

⁶¹ IMF Archives 1956.

⁶² In practice, a few countries still use other (usually higher) thresholds. See IMF and OECD 2003.



investment purpose since at least the 1950s,⁶³ the issue has become particularly acute now that offshore holding structures increasingly obfuscate ownership.⁶⁴

Recent estimates by the US Bureau of Economic Analysis⁶⁵ indicate that holding companies' share of the US outward direct investment position has grown from less than ten percent in 1982 to close to fifty percent in 2012. Figures from Eurostat point into the same direction.⁶⁶ 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.⁶⁷ 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 twenty percent of global FDI flows in 2015).⁶⁸ For researchers interested in investment flows between countries, such issues (should) take center stage, as Andrew Kerner has shown through a replication exercise.⁶⁹

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.⁷⁰ 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

⁶³ IMF Archives 1956.

⁶⁴ Garcia-Bernardo et al. 2017.

⁶⁵ Ibarra-Caton and Mataloni 2014.

⁶⁶ Eurostat 2016.

⁶⁷ Cf. Blanchard and Acalin 2016.

⁶⁸ UNCTAD 2016, 3.

⁶⁹ Kerner 2014.

⁷⁰ Bertaut et al. 2006.



Islands.⁷¹ Conceptually, the lines between “foreign” and “domestic,” “long-term” and “short-term” investments are blurred. Hard figures of the kind we find in statistical yearbooks do nothing to change this fundamental ambiguity; indeed, they suggest clarity 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.⁷² 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. While the US FED implemented methodological changes to track stock swaps separately in 2012,⁷³ such issues remain unaddressed in nearly all other jurisdictions.

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.

⁷¹ Fichtner 2016.

⁷² Griever et al. 2001.

⁷³ Research interview with US BEA economists, Washington D.C., 20 September 2017.



Implications

The growing gap between what the data aspires to represent and what it actually captures 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.⁷⁴ 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, balance of payments 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.⁷⁵ Although replication studies with GDP⁷⁶ or FDI⁷⁷ data have shown that much-cited research results are sensitive to measurement errors and ambiguities in conceptual definitions, admonitions have generally fallen on deaf ears. All too frequently, the precision with which economic statistics appear in data sets makes us oblivious to the fact that they remain “human-made estimates... not true values.”⁷⁸ Nearly all statisticians we interviewed advised cautious interpretation of international economic statistics. In their minds, the primary

⁷⁴ Cf. Mosley 2000.

⁷⁵ Cf. Manski 2015.

⁷⁶ Johnson et al. 2013.

⁷⁷ Kerner 2014.

⁷⁸ Research interview with senior WTO statistician, Geneva, 22 August 2017.



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.”⁷⁹

Statistical compendia do not trumpet data problems on their covers. But footnotes or appendices normally do mention limitations, often obliquely. Statisticians are certainly aware of them and would, we learned in our interviews, tackle them today rather than tomorrow were it possible.

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.⁸⁰ 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.⁸¹

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.

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. Nevertheless, we wish to conclude with three suggestions that together constitute an however imperfect data integrity check. First, to strengthen the robustness of our

⁷⁹ Ibid.

⁸⁰ 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.

⁸¹ Research interview with WTO statistician, Geneva, 22 August 2017.



results, researchers need to become familiar with the specific measurements used in our data sets as proxies for broader concepts. In particular, we need to acquaint ourselves with the measurement problems of data sets to determine whether these endanger our inferences. We also need to examine to what degree the reported data actually reflect what they say on the outside think for example of reported US outward PFI that resulted from foreign takeovers of American firms.

Second, where alternative measurement approaches are available, we should make clear which approach most closely fits our hypotheses. For some theories involving trade, gross trade may be the relevant concept; for others, re-exports are irrelevant and should ideally be excluded from the data. Whether a data set is appropriate for analysis depends not only on the concepts it covers, but on whether the specific measurement approach actually fits underlying intuitions.

Third, all macroeconomic indicators, independent of their conceptual fit, suffer from measurement errors which we cannot simply assume to be random. Factors that can systematically bias a country's BOP data include deliberate data manipulation, statistical capacity, the structure of the national economy (for example its level of development or the size of its digital sector), and its specific function in the global economic network (e.g. trading hub, financial center, tax haven, offshore center, and so on). As researchers, we inevitably have to deal with imperfect data on international economic trends. The most promising way forward, we argue, leads through careful reflection about potential biases in our data and robustness tests that evaluate how sensitive our findings are to the unescapable defects of economic statistics.



Bibliography

- Adcock, Robert, and David Collier. 2001. Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *American Political Science Review* 95 (3):529-546.
- Bertaut, Carol C., William L. Grier, and Ralph W. Tryon. 2006. Understanding U.S. Cross-Border Securities Data. *Federal Reserve Bulletin*.
<http://ticdata.treasury.gov/Publish/frbul2006.pdf>.
- Biemer, Paul, Dennis Trewin, Heather Bergdahl, and Lilli Japac. 2014. A System for Managing the Quality of Official Statistics. *Journal of Official Statistics* 30 (3):381-415.
- Blanchard, Olivier, and Julien Acalin. 2016. What Does Measured FDI Actually Measure? Policy Brief, Peterson Institute for International Economics.
<https://piie.com/system/files/documents/pb16-17.pdf>.
- Bryan, Dick, Michael Rafferty, and Duncan Wigan. 2017. Capital Unchained: Finance, Intangible Assets and the Double Life of Capital in the Offshore World. *Review of International Political Economy* 24 (1):56-86.
- Coyle, Diane. 2014. *GDP: A Brief but Affectionate History*. Princeton, N.J.: Princeton University Press.
- Damgaard, Jannick, and Thomas Elkjaer. 2014. Foreign Direct Investment and the External Wealth of Nations: How Important Is Valuation? *Review of Income and Wealth* 60 (2): 245-60.
- Deutsche Bundesbank. 1997. Comparison of Bilateral Balance of Payments Between Portugal and Germany. IMF Committee on Balance of Payments Statistics Working Papers.
<https://www.imf.org/external/pubs/ft/bop/pdf/bpwp9702.pdf>.
- Dicken, Peter. 2015. *Global Shift: Mapping the Changing Contours of the World Economy*. New York: Guilford Press.
- Dimand, Robert. 1995. "I have no great faith in Political Arithmetick." Adam Smith and Quantitative Political Economy. In *Measurement, Quantification and Economic Analysis. Numeracy in Economics*, edited by Ingrid Rima, 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, Henry, and Martha Finnemore. 2009. Ontology, Methodology and Causation in the American School of International Political Economy. *Review of International Political Economy* 16 (1):58 71.

Farrell, Henry, and Abraham L. Newman. 2014. Domestic Institutions beyond the Nation-State: Charting the New Interdependence Approach. *World Politics* 66 (2): 331 363.

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, Jan. 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 63.

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

Fioramonti, Lorenzo. 2014. *How Numbers Rule the World: The Use and Abuse of Statistics in Global Politics*. London: Zed Books.

Garcia-Bernardo, Javier, Jan Fichtner, Frank W. Takes, and Eelke K. Heemskerk. 2017. Uncovering Offshore Financial Centers: Conduits and Sinks in the Global Corporate Ownership Network." *Scientific Reports* 7.

Gereffi, Gary, John Humphrey, and Timothy J. Sturgeon. 2005. The Governance of Global Value Chains. *Review of International Political Economy* 12 (1):78 104. doi:10.1080/09692290500049805.

Goertz, Gary. *Social Science Concepts: A User's Guide*. 2006. Princeton, N.J.: Princeton University Press.

Griever, William L., Gary A. Lee, and Francis E. Warncock. 2001. The U.S. System for Measuring Cross-Border Investment in Securities: A Primer with a Discussion of Recent Developments. *Federal Reserve Bulletin*, October.

Herrera, Yoshiko M., and Devesh Kapur. 2007. Improving Data Quality: Actors, Incentives, and Capabilities. *Political Analysis* 15:365 86.

Hirschman, Daniel, and Elizabeth Popp Berman. 2014. Do Economists Make Policies? On the Political Effects of Economics. *Socio-Economic Review* 12 (4):779 811.



Hollyer, James, Peter Rosendorff, and James Vreeland. 2011. Democracy and Transparency. *Journal of Politics* 73 (4):1191-1205.

Ibarra-Caton, Marilyn, and Raymond J. Jr. Mataloni. 2014. "Direct Investment Position for 2013." https://bea.gov/scb/pdf/2014/07_July/0714_direct_investment_positions.pdf.

IMF, and 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 - First Edition*. Washington, D.C.: IMF.

. 1950. *Balance of Payments Manual - Second Edition*. Washington, D.C.: IMF.

. 1992. *Report on the Measurement of International Capital Flows*. Washington, D.C.: IMF.

. 2009. *Balance of Payments and International Investment Position Manual*. Washington, D.C.: IMF.

. 2014. *BPM6 Compilation Guide: Companion Document to the Sixth Edition of the Balance of Payments and International Investment Position Manual*. Washington, D.C.: IMF.

International Monetary Fund, and Organisation for Economic Co-operation and Development. 2003. Foreign Direct Investment Statistics: How Countries Measure FDI 2001. <https://www.imf.org/external/pubs/ft/fdis/2003/fdistat.pdf>.

IMF Archives. "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.

. 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, 1966.

. 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, 1967.



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

Johnson, Simon, William Larson, Chris Papageorgiou, and Arvind Subramanian. 2013. Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates. *Journal of Monetary Economics* 60 (2):255-74.

Karabell, Zachary. 2014. *The Leading Indicators: A Short History of the Numbers that Rule Our World*. New York: Simon & Schuster.

Kelley, Judith G., and Beth A. Simmons. 2015. Politics by Number: Indicators as Social Pressure in International Relations. *American Journal of Political Science* 59 (1):55-70.

Kerner, Andrew. 2014. What We Talk About When We Talk About Foreign Direct Investment. *International Studies Quarterly* 58 (4):804-15. doi:10.1111/isqu.12147.

Kerner, Andrew, Morten Jerven, and Alison Beatty. 2017. Does It Pay to Be Poor? Testing for Systematically Underreported GNI Estimates. *Review of International Organizations* 12 (1):1-38.

Lepenes, Philipp. 2013. *Die Macht der einen Zahl: eine politische Geschichte des Bruttoinlandsprodukts*. *Politische Geschichte*

des Bruttoinlandsprodukts. Erste Aufl. Berlin: Suhrkamp.

Lequiller, Francois, and Derek Blades. 2014. *Understanding National Accounts*. Paris: OECD.

Lipsev, Robert E. Measuring International Trade in Services. 2006. *National Bureau of Economic Research Working Paper Series* No. 12271.

<http://www.nber.org/papers/w12271> <http://www.nber.org/papers/w12271.pdf>.

Loungani, Prakash, Saurabh Mishra, Chris Papageorgiou, and Ke Wang. 2017. *World Trade in Services: Evidence from A New Dataset*. IMF Working Papers. Washington, D.C. Available at <https://www.imf.org/~media/Files/Publications/WP/2017/wp1777.ashx>.

Manski, Charles F. 2015. Communicating Uncertainty in Official Economic Statistics: An Appraisal Fifty Years after Morgenstern. *Journal of Economic Literature* 53 (3):631-53.

Miao, Guannan, and Fabienne Fortanier. 2017. Estimating Transport and Insurance Costs of International Trade. *OECD Statistics Working Papers*, no. 4. <http://dx.doi.org/10.1787/18152031>.

Morgenstern, Oskar. 1963. *On the Accuracy of Economic Observations*. Second edition. Princeton, N.J.: Princeton University Press.



Mosley, Layna. 2000. Room to Move: International Financial Markets and National Welfare States. *International Organization* 54 (4):737-73.

. 2003. Attempting Global Standards: National Governments, International Finance, and the IMF's Data Regime. *Review of International Political Economy* 10 (2):331-62.

Muegge, Daniel K., and Bart Stellinga. 2015. The Unstable Core of Global Finance: Contingent Valuation and Governance of International Accounting Standards. *Regulation & Governance* 9 (1):47-62.

Oatley, Thomas. 2011. The Reductionist Gamble: Open Economy Politics in the Global Economy. *International Organization* 65 (2):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, Ronen, Richard Murphy, and Christian Chavagneux. 2009. *Tax Havens: How Globalization Really Works*. Ithaca, N.Y.: Cornell University Press.

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

Sartori, Giovanni. 1970. Concept Misformation in Comparative Politics. *American Political Science Review* 64 (4):1033-1053.

Seabrooke, Leonard, and Duncan Wigan. 2017. The Governance of Global Wealth Chains. *Review of International Political Economy* 24 (1):1-29. doi:10.1080/09692290.2016.1268189.

Shaxson, Nicholas. 2012. *Treasure Islands: Uncovering the Damage of Offshore Banking and Tax Havens*. New York: Palgrave Macmillan.

Snyder, Jack L, and Alexander Cooley, eds. 2015. *Ranking the World: Grading States as a Tool of Global Governance*. Cambridge, U.K.: Cambridge University Press.

Sposi, Michael, and Janet Koech. 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, Joseph E, Amartya Sen, and Jean-Paul Fitoussi. 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, Jeremy L. 2016. Juking the Stats? Authoritarian Information Problems in China. *British Journal of Political Science* 46 (1):11-29.

Ward, Michael. 2004. *Quantifying the World: UN Ideas and Statistics*. Bloomington, In: 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]

