



UvA-DARE (Digital Academic Repository)

The politics of global statistical harmonization

DeRock, D.J.

Publication date

2022

Document Version

Final published version

[Link to publication](#)

Citation for published version (APA):

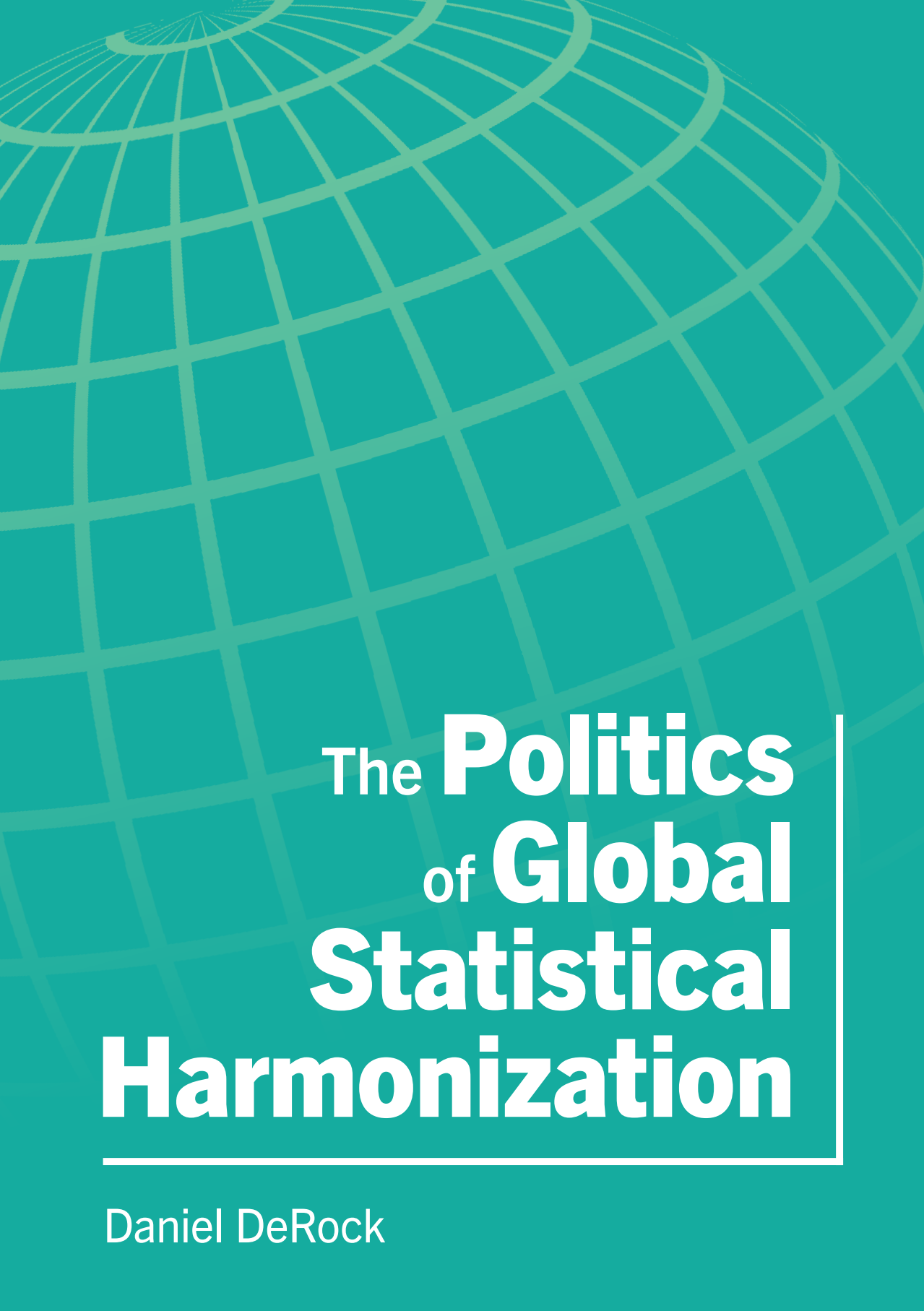
DeRock, D. J. (2022). *The politics of global statistical harmonization*.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.



The **Politics**
of **Global**
Statistical
Harmonization

Daniel DeRock

The Politics of Global Statistical Harmonization

Daniel DeRock

This research was supported by the European Research Council (ERC) under the grant 'FICKLEFORMS' (Grant #637883).

Cover & Layout: Douwe Oppewal
Print: Ipskamp Printing, Enschede

© Copyright 2021: Daniel DeRock

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, by photocopying, recording, or otherwise, without the prior written permission of the author.

The Politics of Global Statistical Harmonization

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het
College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in
de Aula der Universiteit
op woensdag 19 januari 2022, te 14.00 uur

door Daniel Joseph Derock
geboren te Illinois

Promotiecommissie

Promotor: prof. dr. D.K. Mügge Universiteit van Amsterdam

Copromotor: prof. dr. B.M. Burgoon Universiteit van Amsterdam

Overige leden: prof. dr. M. Jacobs Universiteit van Amsterdam
 prof. dr. F.H.B. Biermann Universiteit van Amsterdam
 prof. dr. A. Freyberg-Inan Universiteit van Amsterdam
 prof. dr. M. de Goede Universiteit van Amsterdam
 dr. N.R.M. Pouw Universiteit van Amsterdam

Faculteit der Maatschappij- en Gedragwetenschappen

Table of contents

Overview of articles in the thesis	p. VI
Acknowledgements	p. VII
1. Introduction.....	p. 2
2. Theorizing the politics of statistical harmonization.....	p. 16
3. Hidden in plain sight: unpaid household services and the politics of GDP measurement	p. 28
4. The statistical trilemma: built-in limitations of international economic statistics	p. 46
5. Bounded autonomy in the global governance of economic statistics	p. 66
6. How GDP spread to China: the experimental diffusion of macroeconomic measurement	p. 86
7. Conclusion	p. 108
Bibliography.....	p. 118
Appendix: List of interviews.....	p. 142
Summary.....	p. 148
Samenvatting.....	p. 151

Overview of articles in the thesis

This thesis is based on four stand-alone articles, two of which have already been published. Two of the articles are co-authored. For the co-authored articles, both authors contributed in equal manner. The thesis also builds upon two online publications and one working paper, all single-authored.

The bibliographic details of the published articles are as follows:

DeRock, D. (2021). Hidden in Plain Sight: Unpaid Household Services and the Politics of GDP Measurement. *New Political Economy*, 26(1), 20-35.

Van Heijster, J. & DeRock, D. (2020). How GDP spread to China: the experimental diffusion of macroeconomic measurement. *Review of International Political Economy*. <https://doi.org/10.1080/09692290.2020.1835690>.

The two remaining articles are both currently under review. The details of these articles are as follows:

DeRock, D. & Mügge, D. (2021). The statistical trilemma: built-in limitations of international economic statistics. *Under review*.

DeRock, D. (2021). Bounded autonomy in the global governance of economic statistics. *Under review*.

Online publications (not peer-reviewed):

DeRock, D. (2018). The Dangerous Depoliticization of Economic Numbers. The Policy Corner. <https://www.policycorner.org/en/2018/02/26/the-dangerous-depoliticization-of-economic-numbers/>

DeRock, D. (2019). Unpaid Work and the Governance of GDP Measurement. E-International Relations. <https://www.e-ir.info/2019/01/11/unpaid-work-and-the-governance-of-gdp-measurement/>

Working paper:

DeRock, D. (2020). *Leaving no data behind? International organizations and the conflicted global governance of monitoring the Sustainable Development Goals*. Fickle Formulas Working Paper 01-2020.

Acknowledgements

This PhD has been quite an adventure, and I haven't done it alone. I will get around to thanking the people who supported me on this journey in just a moment. First, though, I want to quote one of my favorite passages from one of my favorite books. The book is *The Dispossessed* by Ursula K. Le Guin, and the passage is one I've been trying, unsuccessfully, to include somewhere in the dissertation. This acknowledgements section is the last thing I'm writing before it all goes to the printer, so it's now or never.

“He tried to read an elementary economics text; it bored him past endurance, it was like listening to somebody interminably recounting a long and stupid dream. He could not force himself to understand how banks functioned and so forth, because all the operations of capitalism were as meaningless to him as the rites of a primitive religion, as barbaric, as elaborate, and as unnecessary. In a human sacrifice to deity there might be at least a mistaken and terrible beauty; in the rites of the moneychangers, where greed, laziness, and envy were assumed to move all men's acts, even the terrible became banal”
(Le Guin 1974).

I'm tremendously grateful to have spent a large part of my life making sense of “the operations of capitalism” in a way that Shevek – the “he” in the above quote – would find tolerable. And this brings me to the first of many acknowledgements. I would like to thank Angela Wigger for sparking my passion for political economy as a master's student at Radboud University and for encouraging me at every step of my career. I'm equally grateful to many others at Radboud, especially Bertjan Verbeek and Gerry van der Kamp-Alons. Their support during my master's program gave me the knowledge and confidence to pursue a PhD.

Moving on to the PhD: I couldn't have asked for better supervisors than Daniel Mügge and Brian Burgoon. Along with being my promotor, Daniel had the official title of “daily supervisor.” He was, and still is, truly there for me on a daily (more or less) basis. From academic writing skills to new ways of understanding the world, I've learned more from Daniel than I ever could have anticipated. Brian, too, has been an incredibly supportive and insightful supervisor since the first day. After every meeting with Brian, I walked away feeling not only much closer to finishing, but also smarter.

This difficult process would not have been nearly as enjoyable without my Fickle Formulas teammates. Jessica, Joan, Roberto: I'm so lucky to have you as teammates and as friends. I've learned a lot from your brilliant minds. We've had some great times at the workshops and conferences, not to mention countless days at the office thinking about and discussing indicators. I'm very grateful to the rest of the Fickle Formulas team, as well. Thank you, Lukas, Juliette, Tobias, Ana, Francisca, and Roel!

I've made so many friends throughout the PhD that I'm worried to leave out some names unintentionally. To avoid that, I'm going to skip the names altogether. You know who you are! Thank you so much to all the AISSR PhDs I met along the way, especially the Rebels Without a Clue of room B9.01 and the CORPNET office next door.

I also want to thank everyone in PETGOV. I'm honored to have been part of such an impressive group of scholars. There are too many to name, but I'm thankful for the support and encouragement of Luc Fransen, Ursula Daxecker, Annette Freyberg-Inan, Julian Gruin, Eelke Heemskerk, Philip Schleifer, Abbey Steele, and many others. Many thanks, also, to Naziha and others in the Political Science secretariat.

My dissertation would not have been possible without the generosity and expertise of the interviewees. Each of them took time out of very busy schedules and provided the most important data upon which my research is supported. The interviews I conducted are by far the most memorable experiences of my PhD. As a fledgling PhD student, I was amazed that such high-profile policy experts would invite me to ask them questions. To any interviewees reading this: thank you very much. You made this research possible.

Finally, last but not at all least, thanks to my friends and family who supported me through the PhD and all the years that came before. I'm so grateful for the unending support of my parents and of my brother and sister, Eric and Carrie. I could write an entire chapter and it wouldn't be enough to give proper acknowledgement to my parents. Thank you, Mom and Dad. And thanks to all my friends all over the world who have kept me grounded.

Thank you, Chrissy, for everything. You've been my biggest supporter and my best friend throughout all of this. You celebrated with me when I made big steps in the PhD, kept me in balance when things got too stressful, gave me confidence when I had imposter syndrome. Most importantly, you bring so much happiness to my life outside of academia. Despite both of us being busy and achieving huge goals, we even found time to get married.

And how could I forget my dogs, Ivy and Juno? Ivy has helped keep my sanity in check for the past ten years. Juno, the newest family member, added tons of joy and energy to the strenuous final stage of the PhD.

1

Introduction

Much of what we know about our economies and societies comes to us through statistics. We expect statistics to tell us the size and growth rate of a given country, how many people are unemployed or impoverished, the inflation rate, and much more. Policymakers, investors, and development agencies, among others, rely on these figures to make sense out of a complex world. The kinds of statistics that are available, and the particular ways they are measured, shape actors' perceptions and thereby influence policy decisions from behind the scenes.

Because the information provided by economic statistics is so crucial for so many actors – from policymakers, to institutional investors, to NGOs – a great deal of effort goes into making statistics available, comparable, and reliable. Since the late 1940s, international organizations (IOs) such as the United Nations, the International Monetary Fund (IMF), and the World Bank have been at the forefront of efforts to harmonize economic statistics around the world. The international statistical standards that they have created, like the System of National Accounts (SNA) and the Balance of Payments Manual (BPM), are intended to ensure that official statistics are trustworthy and comparable. The hope is that these statistics provide an evidence-based foundation for policy and global economic relations. As the current director of the IMF Statistics Department has put it, “Without standards, there are no consistent and comparable data and without data there are no good policies” (IMF 2016, p. 1).

As the world’s foremost embodiment of multilateralism, and also as technocratic institutions reliant on data for their own work, IOs take the lead in promoting common rules and norms of economic measurement. However, there are gaps between the ambitions of IOs and the actual, on-the-ground statistical practices of member countries. For one thing, international statistical standards often clash with socioeconomic realities within countries (Alenda-Demoutiez and Mügge 2019; Morgan 2009). While countries and sub-national regions are diverse – for example in terms of the size of the informal sector, the functioning of the financial system, or the structure of labor markets – statistical standards demand that national economies are measured more or less uniformly. Such one-size-fits-all measurement practices reduce complex socioeconomic realities to the deceptively simple figures we see in databases and newspaper headlines (Espeland and Stevens 2008). Such issues are particularly problematic in the developing world, where statistical practices have often been ‘imported’ from the highly industrialized countries in which international standards originated. For example, international standards for unemployment statistics were first designed for industrialized countries when factory work was considered the norm. Yet, the ILO estimates that informal work constitutes more than 90 percent of labor in many sub-Saharan African countries (International Labour Office 2018, p. 85ff).

Moreover, the production chain of official statistics is full of potential bottlenecks (Herrera and Kapur 2007). Before economic statistics end up in databases, government statisticians and civil servants must first run costly household surveys, collect data from government ministries and businesses, make this raw data consistent with internationally harmonized methodologies, and report the resulting statistics on time and according to quality standards. Many countries, especially poor ones, lack the budget and capacity to meet these demands (Jerven 2013). As a result, data is often missing or of poor quality, and compliance with standards might only be superficial or partial. By region, Africa and Oceania (which includes many small island states) perform substantially lower than other

regions (PARIS21 2019, p. 25). In many countries, staff and budgets are extremely limited, as a UN Statistics Division statistician specialized in capacity-building explains:

I work with some very small countries – you know, in the Caribbean or elsewhere – and they might have one person doing all the economic statistics, CPI, national accounts, everything. And if that one person leaves, you're right back to square one. (Interview 1)

What is more, especially in countries where these capacity constraints are present, the importance placed on internationally harmonized economic statistics can potentially crowd out other data-gathering that would be more suitable to local policy priorities. For instance, should a country with limited statistical capacity prioritize fiscal indicators over data collection for the Sustainable Development Goals (SDGs)? Prioritizing certain kinds of statistics over others can force statisticians to make unreliable estimates or limit policymakers' knowledge of pressing social issues if they are not measured at all.

These problems are all the more troublesome because statistics are often taken at face value. They are widely seen as technocratic, and thus non-political. This makes statistics powerful, but largely misunderstood, objects in politics and economics. Given the degree to which statistical practices shape perceptions and policies, combined with the remarkable challenge of reducing so much diversity into shared global measurement standards, it is not at all clear why global harmonization efforts take the form that they do. These observations and concerns lead to the following central research question:

What are the drivers of international harmonization of economic statistics?

I conceive of statistical harmonization as a process containing multiple steps. Even if measurement practices are formally institutionalized in international standards, it is not given that they will be agreed upon and implemented globally, especially when standards clash with domestic socioeconomic contexts. The dissertation investigates three levels of harmonization: standard-setting, diffusion, and implementation. Thus, the central question can be broken up into three sub-questions:

1. Which factors shape the design of international statistical standards?
2. What explains the spread of international statistical standards and practices around the world, especially to countries where standards have clashed with local economic conditions?
3. How are international statistical standards and practices adopted and implemented domestically within national statistical systems?

The sub-questions correspond to the three levels of statistical harmonization indicated above (see also Figure 1 below). Regarding the first sub-question, I argue that statistical

standard-setting is driven by experts (Chapter 3), but also constrained by limitations inherent to official statistics (Chapter 4). Regarding the second sub-question, the spread (or diffusion) of standards and practices can be either top-down or bottom-up. Chapter 6 argues that the diffusion of GDP to China was a bottom-up process driven by domestic preferences and ideas. This is a crucial case, but also an atypical one. For most developing countries, the top-down efforts of international organizations often exert strong pressure to comply with standards for economic statistics (Chapter 5). Finally, regarding the third sub-question, I find that implementation varies between countries, and also that implementation efforts vary between IOs. In China, implementation of the SNA was impeded by severe mismatches between the standards and domestic conditions (Chapter 6). In that case, IOs took a hands-off approach to promoting adoption of the SNA. In many developing countries, low statistical capacity is the biggest barrier to compliance with IOs' demands, and compliance is often only partial. As chapter 5 demonstrates, IOs attempt to promote compliance with statistical standards through a combination of formal and informal means.

The dissertation builds upon a burgeoning literature on the politics of economic statistics and indicators. It engages with and refines international political economy (IPE) and international relations (IR) theories related to international standards, norm and policy diffusion, and international organizations. The empirical chapters shed light on many of the key actors and institutions that govern the international statistical system. They also aim to explain why the system takes the form that it does and how it has evolved over time.

This introductory chapter proceeds first with a brief overview of gaps in the existing literature (Chapter 2 includes a more detailed literature review). The remaining sections give a description of the research focus and an overview of the following chapters.

Gaps in the existing literature

Social science literature about statistics has a long history, and it has grown significantly in recent years (Mügge 2019). Literature on the philosophy, history, and sociology of science has contributed important insights about the social aspects of numbers (Alonso and Starr 1987; Camargo 2009; Desrosières 2000; Hacking 1981; Porter 1995). More recently, IPE and IR literature has examined the roles of statistics, indicators, and benchmarks in global governance (e.g. Cooley and Snyder 2015). A smaller but growing body of literature narrows in on the political economy of economic statistics (e.g. Herrera 2010; Jerven 2013; Mügge and Linsi 2020; Yarrow 2020).

Three broad strands in this literature relate directly to the dissertation. The first cluster focuses on problems in economic measurement. For example, several authors have pointed

out biases and shortcomings in GDP methodology and the hazards of using the indicator as a yardstick for societal welfare (e.g. Coyle 2014; Fioramonti 2013; Philipson 2014; Stiglitz, Sen, and Fitoussi 2010). This literature argues that many of the statistics we often take for granted should at least be scrutinized, if not abandoned altogether. However, we still know little about why indicators like GDP are measured the way they are. Chapters 3 and 4 of the dissertation take up this question. They focus on the actual governance of statistical standards and the forms of agency and constraints that shape them. Both chapters also identify causal mechanisms behind particular outcomes of standard-setting, such as limitations in GDP measurement. This approach builds upon critiques of economic statistics but shows that solutions to measurement problems are more complicated than often assumed.

A second cluster of literature focuses on the power and roles of numbers in global governance (Davis, Kingsbury, and Merry 2012; Fukuda-Parr, Yamin, and Greenstein 2014; Kelley and Simmons 2015). This literature sheds light, for example, on how indicators enable “governance at a distance” (Hansen and Porter 2012). It also shows that quantitative benchmarks, such as the World Bank’s Doing Business Indicator, can be misleading (Broome, Homolar, and Kranke 2018). The literature demonstrates that technocratic governance through numbers has far-reaching effects. But it generally does not ask how these quantitative governance tools have spread around the world in the first place. This diffusion is a prerequisite for numbers having the power that they do. Chapter 6 provides a theoretical framework for the diffusion and localization of statistical practices, with emphasis on domestic variables. Chapter 5 focuses on transnational variables that drive the (attempted) diffusion of statistical practices.

The third relevant cluster focuses on statistical capacity problems in developing countries (Dargent, Lotta, Mejía, and Moncada 2018; Jerven 2013; Taylor 2016). It details the constraints on underfunded and understaffed national statistical offices. The literature on statistical capacity challenges the image of official statistics as accurate representations of economic reality. There are distributional consequences, given that bad data can result in wrongheaded policy (Devarajan 2013). On the whole, this literature implies that the solution to data problems, however difficult, is to increase statistical capacity in one way or another. While intuitive, this kind of technical fix does not fully address the politics intertwined with statistical capacity building. That is, it ignores the interests, ideas, and distributional implications behind IO interventions into the statistical systems of low- and middle-income member countries. It is not immediately clear whether such efforts truly prioritize ‘country ownership,’ as IOs often claim. The IMF’s Data for Decisions Fund, for example, is described as “member-focused...” (IMF 2018a, p. 11). Nor is it clear whether these efforts aim to support member countries’ own policy priorities or to promote the priorities of individual IOs. These questions are taken up in Chapter 5.

More broadly, the dissertation also enters into debate with several major theories of international relations that are concerned with the role of knowledge and information in global politics. The findings cannot be explained by functionalist or rationalist approaches, wherein international institutions are assumed to lower transaction costs and reduce information asymmetries (Garrett 1992; Keohane 1984; Krasner 1991). Such literature assumes that states create and participate in international institutions and organizations because the expected shared gains are in their interests (e.g. Kapstein 1989). IPE literature on standard-setting, often related to cross-border trade, has also often assumed a functionalist logic (e.g. Abbott & Snidal, 2001; Mattli & Buthe, 2003). Abbott and Snidal (2001, p. 366), for example, suggest that “International pooling of information and resources may produce superior responses to standards problems” even when national solutions are feasible. These approaches treat information as unproblematic and efficiency-promoting. The dissertation, in contrast, suggests that this is far from the truth, at least where economic statistics are concerned. As such, the dissertation pursues questions that are outside the scope of functionalist and rationalist research.

The dissertation also has implications for the epistemic communities literature. This literature marked an important breakthrough in IR by centering the political agency of specialized experts (Haas 1992, p. 11). This approach has had a strong influence on the dissertation. Yet, although the epistemic communities approach is often understood as a major departure from rationalism, both approaches tend to treat knowledge and information relatively uncritically. Haas (1992) argues that epistemic communities play crucial roles in reducing uncertainty and information asymmetry in policymaking. There is an implication of a pure science or technical expertise that often becomes mixed with “nontechnical issues centering around who is to get what in society and at what cost” (ibid., p. 11). But, in economic statistics, these distributive matters are inseparable from the production of statistics. The distinction between science and politics, I argue, is misleading. The most relevant of this literature has focused on epistemic communities of economists. Such communities are portrayed as influencing political outcomes from the outside rather than engaging in politics directly (e.g. Ikenberry 1992; Verdun 1999). In contrast, the dissertation treats economic statisticians as agents participating directly in political life by shaping the informational infrastructure of global governance.

Multi-level harmonization

Taken together, the four empirical chapters (3-6) investigate the three levels of harmonization introduced above: the creation of statistical standards, their diffusion around the world, and their domestic implementation (see Figure 1).

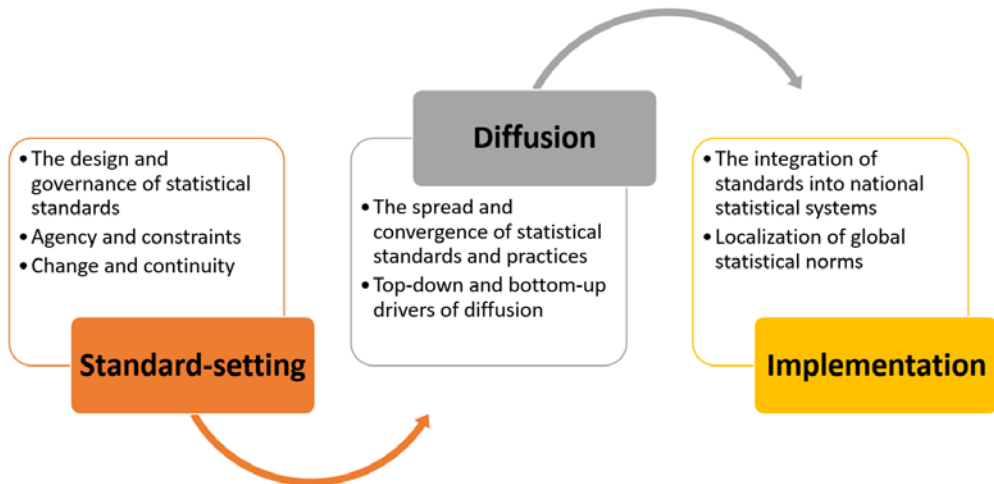


Figure 1: Three levels of global statistical harmonization (figure created by author)

The first level of harmonization that the dissertation investigates is statistical standard-setting. This is the focus of Chapters 3 and 4. Standard-setting entails decisions about the definitions, concepts, and methodologies underlying the measurement of economic statistics. It is a theoretical undertaking and thus led by specialized experts. For example, the former editor of the Systems of National Accounts described the standard-setting and revision process as

a question of looking at how we think the system *ought* to work and looking at the pros and cons – why is it like this, and why isn't it like that, what would the problem be in changing it. And so it's very theoretical, as opposed to when you're out in a country you're saying 'where on earth can I get figures for this or that'. (Interview 2)

Statisticians both inside and outside of IOs continuously develop the methodologies that contribute to standard-setting. However, the official creation and revision of standards is typically delegated expert groups composed of staff of several IOs. Some examples are the Intersecretariat Working Group on National Accounts (ISWGNA), the IMF Committee on Balance of Payments Statistics, and the International Conference of Labour Statisticians.

The second level of harmonization is diffusion. IPE and IR literature on the diffusion of policies, norms, ideas, and standards posits several pathways for the spread of these objects (Börzel and Risse 2012; Zimmermann 2016). They can be roughly split up into top-down and bottom-up explanations. Top-down diffusion can be actively driven by external actors or more passively through functional pressures. Bottom-up diffusion, in contrast, is driven by domestic factors. In the diffusion of statistical standards and practices we can observe both top-down and bottom-up dynamics. This issue is addressed in Chapters 5 and 6.

The third level of harmonization is implementation. Implementation is about how, and to what degree, statistical standards and practices are actually integrated in domestic policymaking and bureaucratic institutions. International standards sometimes clash with domestic or local economic conditions, such as the size of a country's informal sector, the functioning of its financial system, or cultural differences that do not chime with standardized 'Western' notions of work or production. These kinds of factors affect how (and to what degree) standards and practices are actually implemented in domestic government and policymaking. Implementation is also addressed in Chapters 5 and 6.

Empirical focus

The empirical chapters (Chapters 3-6) investigate the politics of harmonizing economic statistics. The concepts 'statistics' and 'data' are sometimes used interchangeably. As Round (2014, p. 1) explains, "The raw data collected via sample or surveys are usually processed or summarised into statistics. These may be simple summary statistics ... or they may be more complex measures (such as price indices, national accounts aggregates, etc)". The dissertation focuses on economic statistics – that is, the measures such as GDP into which raw data are aggregated. Yet the raw data collected by national statistical systems are also relevant in harmonization efforts, and thus also discussed in the empirical chapters.

GDP receives the most attention in the dissertation, especially in Chapters 3 and 6. Other chapters address GDP alongside statistics of trade, foreign direct investment, debt, poverty, unemployment and others. The System of National Accounts is the focus of much of the dissertation. The SNA is an important international statistical standard for many reasons. It is the basis for GDP and other aggregate indicators, including Gross National Income, Gross National Disposable Income, Gross Fixed Capital Formation, and many others. It is also integrated with other important standards such as the IMF's Balance of Payments Manual and Government Finance Statistics Manual (UNSD 2013).

At the margins, the distinction is blurred between economic statistics and other types, such as financial, social, environmental, or demographic statistics. Unemployment statistics, for example, are sometimes labeled as economic and sometimes as social statistics. Recent trends encourage a blurring of categories. For instance, the UN System of Environmental-Economic Accounting (SEEA) blends economic and environmental data, and the Human Development Index (HDI) blends economic and socio-demographic dimensions. Nonetheless, economic statistics – such as those on economic growth, trade, debt, unemployment, and inflation – continue to dominate many areas of policy and research, and IOs and national statistical offices typically make a distinction between economic and non-economic statistics. Moreover, the fact that statistical categories are clear-cut on the surface, but blurry in practice, only increases the importance of academic scrutiny.

The key actors in global harmonization efforts are economic statisticians (particularly IO staff), national statistical offices, and, to a lesser degree, domestic policymakers. When IOs engage with member countries on statistical matters, typically the staff or directors of IO statistics departments communicate with staff or directors of national statistical systems. In relatively rare cases when statistics become more politically salient, IO executives and central government can be directly involved.

The chapters focus on the United Nations Statistics Division (UNSD), the World Bank, and the IMF as the most relevant IOs attempting to influence macroeconomic measurement practices in low- and middle-income countries. To be sure, these are not the only actors involved in statistical governance. Individual states provide bilateral aid and lending, contribute to capacity building trust funds, and make up the membership of IOs. Other notable actors include Eurostat (the statistical agency of the European Commission), the Organisation for Economic Cooperation and Development (OECD), the International Labour Organization (ILO), and regional development banks. The dissertation takes such actors into consideration. The OECD and Eurostat, for example, are part of the inter-secretariat working group that is investigated in Chapter 3. For the most part, however, the wider assemblage of other IOs and non-state actors that comprise the complete governance architecture (Biermann et al., 2009) are outside the scope of the dissertation. The three IOs at the center of the research, including their internal departments and staff, are the most active in truly global efforts to harmonize economic statistics. They also have the broadest reach, given that nearly all countries worldwide are members.

At the national level, national statistical offices (NSOs) are responsible for the production of official statistics, often alongside other government ministries or agencies such as central banks and ministries of finance, agriculture, and so on. Collectively these domestic statistical producers make up a national statistical system, an “ensemble of statistical organisations and units within a country that jointly collect, process and disseminate official statistics on behalf of national government” (OECD 2004). While NSOs are not the primary focus of the research, they are nonetheless important actors in statistical harmonization, and are addressed in chapters 3, 5, and 6.

Research approach and data

This section gives an overview of the research approach. Each of the empirical chapters follows its own research design, and more detailed discussions of methodology and concepts are included in chapters 3-6. This section proceeds with an overview of the qualitative approach of the dissertation, a summary of each chapter’s research design, interview methods and document analysis.

The dissertation is a qualitative study of the attempted harmonization of economic statistics around the world. It aims to uncover, in rich empirical detail, why we quantify the (global) economy the way we do, and why statistics are governed the way they are. This ambition entails identifying causal processes, drivers of historical change and continuity, the influence of discourse on policy outcomes, and power relations in a largely technocratic field. These are research aims best suited to in-depth qualitative analysis. There is also a large descriptive ambition, given that the gaps in existing IPE literature are still wide. The description is guided by theory and contributes to the explanatory aims. In a constitutive approach to explaining global politics, “understanding the constitution of things does essential work in explaining how those things behave and what causes outcomes” (Barnett and Finnemore 1999, p. 701).

Chapter 3 is a qualitative historical analysis of the System of National Accounts revision process. The findings are based on interviews, archival research, and document analysis. The primary contribution of Chapter 4 is a theoretical one, but the theory of the statistical trilemma is also demonstrated empirically. The empirical analysis takes the form of theoretical redescription of seven case studies, each focused on a different type of economic statistics. Chapter 5 follows a comparative design to explain variation in three functions of global statistical governance. It focuses on the UNSD, the World Bank, and the IMF. The cases that are compared are not the IOs, but the governance functions, namely standard-setting, capacity building, and compliance monitoring. It is based on document analysis and interviews. Chapter 6 is a historical qualitative analysis that applies process tracing to identify causal relations that led to the adoption of the SNA in China. It is based on document analysis and interviews (conducted by Joan van Heijster).

As a whole, the dissertation is supported empirically by semi-structured expert interviews, archival research, official document analysis, participant-observation, and extensive literature reviews. The research process began with an exploratory literature review and document analysis. Despite the availability of social science literature on the topic, there has been relatively little written about the actual practices of statistical governance outside of technical literature aimed at economic statisticians. Through an initial review of secondary and primary literature, I identified some of the most relevant institutions, actors and policies.

The second step was to conduct several exploratory interviews with retired economic statisticians. These early interviews included former staff and directors of UNSD, the World Bank Data Development Group, and the IMF Statistics Department, all of whom had experience both in methodology and capacity building. These interviews yielded important information reaching back several decades, and pointed my research in the direction of key standards and policies to explore in more detail. Several of the interviewees put me in contact with other statisticians currently working in the field.

From that point, I continued with document analysis and interviews guided by the information and theoretical expectations that emerged from earlier interviews. In total, I conducted 27 interviews. Several of the interviews were conducted with multiple interviewees at the same time. In total there were 35 interviewees. A partially anonymized list of interviewees, along with numbered references, is included in the appendix.

Interviews were conducted between 2017 and 2019. The interviewees were selected for their expertise on specific topics such as national accounts and statistical capacity building. At UNSD in New York in 2019, I interviewed ten staff members, including the Chief of the Economic Statistics Branch. I also interviewed two retired UNSD statisticians in 2017 in Amsterdam, including a former Acting Director. In the UK in 2017, I interviewed three statistical capacity building consultants at Oxford Policy Management. At the UN Economic and Social Commission for Asia and the Pacific (ESCAP) in Bangkok in 2018, I interviewed three staff of the ESCAP Statistics Division including the Acting Director. At the World Bank in Washington, DC in 2019, I interviewed three staff of the Data Development Group. I also interviewed a former World Bank statistician in 2017 in the UK. In Bangkok in 2018, I interviewed four staff and directors at the Statistical Office of Thailand. In Vientiane in 2018, I interviewed one high-ranking statistician at the Lao Statistics Bureau. In Accra in 2019, I interviewed three staff of Ghana Statistical Services.

The research also draws upon archival documents, some of which were available online, including reports of the UN Statistical Commission sessions (ranging from the 1950s to 2019). I also scanned documents at the physical UN archives in New York. The rest of the document analysis included, among other sources: the text of the System of National Accounts and supporting guidelines and handbooks; reports by the Independent Evaluation Group of the World Bank and the Independent Evaluation Office of the IMF; and IO documents on statistical capacity building programs. Further details about data sources and research design are included in the following chapters.

Outline of the thesis

The dissertation proceeds with a more detailed discussion of existing literature and relevant theoretical approaches in Chapter 2. Chapters 3 through 6 are the empirical chapters, followed by a conclusion in Chapter 7. The empirical chapters are all based on academic articles that are either published, under peer review, or will be submitted for publication in the near future.

The first empirical chapter, Chapter 3, describes and explains the global governance of GDP measurement. It focuses on standard-setting, arguably the most contentious and insulated aspect of statistical harmonization. The chapter closely investigates one of

the most controversial aspects of GDP measurement, namely the exclusion of unpaid household services (or ‘housework’) from the production boundary of the SNA. Therefore, it addresses the central research question by asking why internationally harmonized statistical standards are designed the way they are. The chapter starts from the observation that the SNA production boundary has in fact expanded over time to include various activities including financial intermediation services (Christophers 2013), the products of subsistence farming, and goods and services from the informal sector, but continues to exclude unpaid household services. The explanatory aim of the chapter is to explain why standard-setters have not included unpaid household services when there are compelling reasons to do so. More broadly, Chapter 3 describes the governance of standard-setting in the SNA and demonstrates the power of experts to determine how economies are measured. The chapter is published as a journal article in *New Political Economy* and has been modified only slightly to conform with the style of the dissertation.

Chapter 4 is co-authored with Prof. Daniel Mügge. The chapter starts from the observation that, while statistics have become increasingly central to global economic governance, there is a gap between the aspirations for statistics and what they can actually deliver. As the other chapters show, these gaps can come from biases and blind spots in methodologies, or from data gaps due to low statistical capacity or to non-compliance with standards. These shortcomings are found in all types of indicators and have no simple solutions. Thus, the chapter asks whether there are universal limits to the harmonization of economic statistics. The argument is that standard-setting is constrained by a trilemma that results from contradictory demands of stakeholders, including policymakers, analysts, and citizens. Statistics are expected not only to be harmonized, but also guided by standards that are detailed and prescriptive (for the sake of transparency and comparability), and also to be suitable to local socioeconomic realities on the ground. We argue that it is only possible to satisfy two of these criteria at the same time. This is demonstrated through empirical examples explaining six types of statistics: national accounts, poverty, unemployment, trade, debt, and foreign direct investment. The chapter is based on an article that is currently under review.

Chapter 5 addresses cooperation between IOs in global statistical harmonization. It addresses the central research question beyond standard-setting, thus moving beyond the scope of the preceding chapters. While standard-setting is sometimes viewed as synonymous with harmonization, actual harmonization efforts have additional layers. The chapter focuses on the efforts of, and interactions between, the UNSD, the World Bank and the IMF. The descriptive aim is to determine the extent of inter-organizational cooperation in the governance of economic statistics. The explanatory aim is to determine why the degree of cooperation varies between three main governance functions: standard-setting, statistical capacity building, and efforts to monitor compliance with standards. To explain cooperation in statistical issues, I investigate the internal bureaucratic politics

of each IO as well as the epistemic community of economic statisticians that transcends individual IOs. This analysis is guided by what I refer to as the bounded autonomy of the statistical epistemic community. I introduce the concept of bounded autonomy in the chapter. The chapter is based on an article that is currently under review.

In Chapter 6, which is co-authored with Joan van Heijster (van Heijster is first author), we explain the diffusion of the SNA (and thus also GDP) to China. Diffusion is a crucial step in the harmonization of economic statistics. Harmonization is only meaningful if standards spread globally and are adopted and implemented at the national level. GDP was not always produced worldwide, despite being formally harmonized in the SNA since the 1950s. For decades, the former Soviet Union, China, and some other ‘socialist’ countries used a different indicator of aggregate production called Net Material Product (NMP). As GDP is based on the SNA, NMP is based on the Material Product System (MPS). China was the last major country to abandon the MPS and adopt the SNA in 1993. This is surprising, considering how much attention China’s GDP figures receive now, and how much influence they have on China’s domestic policymaking (van Heijster 2020). We argue that GDP was not imposed on China in a top-down fashion. It was a bottom-up process driven by domestic actors and shaped by domestic debates and competing economic ideas. This is an important case not only because of China’s political-economic importance, but also because of the clash between the economic ideas of the SNA and the material and ideational conditions in China.

Finally, Chapter 7 is the conclusion. In that chapter I summarize the findings and discuss the academic relevance, policy implications, and suggestions for future research. The conclusion also discusses the overarching themes and insights that arise when all four empirical chapters are taken together. It returns to timely questions including the ongoing backlash against expertise, and the limitations and potentials of quantified global governance. It also raises important forward-looking questions to which the dissertation can provide some clues, including about alternative governance arrangements for international statistics, the measurement of the SDGs and future development agendas, and the growing power of big data and algorithms.

2

Theorizing the politics of statistical harmonization

What are the drivers of international harmonization of economic statistics? To answer this central question, the dissertation builds upon several academic debates in the disciplines of IPE, IR, and economic sociology. This chapter summarizes the most relevant theories and concepts from these literatures, as well as the theoretical approaches taken in the empirical chapters. The aim is to present an overview of the main theoretical issues that run throughout the dissertation. As the empirical chapters are each based on stand-alone articles, more detailed theoretical discussions are included there.

Seeing and governing the world through numbers

Quantifying the economy allows us to ‘see’ it and make it legible (Scott 1998). As Espeland and Stevens (2009, p. 415) argue, seeing something allows us to govern it. Nowadays, seeing the world through the lens of statistics is taken for granted. We can easily find out about a country’s population, unemployment, literacy rate, inflation, and so on. We can compare any two countries or regions and rank them by the size of their economies or by levels of inequality, democracy, transparency or poverty.

This was not always possible. In the early 1800s, Adolphe Quetelet transformed social science by introducing the concept of statistical averages into the mainstream (Porter 1986, pp. 51-52). Quetelet showed that “statistical laws may be true when applied to groups even though they are false in relation to any particular individual” (ibid.). Today, this phenomenon – which Porter (ibid.) calls “the obliteration of the particular by the general” – is omnipresent.

Just as important as the ability to see the world through numbers is the way that numbers shape perceptions and policies. The process by which social phenomena are translated into statistics is unavoidably political, and masks contestation over the underlying economic theories and ideas. In the words of Fukada-Parr et al. (2014, p. 106), “translating social phenomena into measurable outcomes involves a transformation that reifies intangible phenomena, simplifies complex concepts, and abstracts social change from local contexts.”

As Chapter 3 demonstrates, many of the ideas underpinning statistics like GDP can be traced back to the first efforts at creating comparable national income accounts (roughly the 1930s through the early 1950s). Ward (2004a, p. 300) argues that the System of National Accounts, which was first published in 1953, “assumes there is a standard underlying economic model that serves all countries equally”. As Speich (2011, p. 27) describes it, the dominant worldview among ‘Western’ economists in that time period was that

All distinct entities of economic activity were assumed to work according to the same universal principles. Their malfunctions could be analysed in an engineering perspective that generated policy advice on how to gear up productivity.

It is not immediately clear whether this mechanistic worldview has gone unchallenged throughout the roughly seventy years of attempted statistical harmonization. The following chapters suggest that the assumptions of economists in the mid-twentieth century have indeed exerted strong path-dependent effects, and many of the same ideas are still common today. However, in addition to this ideational explanation, the empirical chapters also pay attention to structural features of statistics themselves (Chapter 4), inter-organizational interactions (Chapter 5), and domestic agency (Chapter 6).

Ontology and epistemology of economic statistics

Existing literature has argued that statistics play a large role in the social construction of the economy as something separate from the rest of social life (Allan 2018; Mitchell 2002, 2005). Breslau (2003, p. 380) aptly writes that “No one has ever seen the economy or touched it except through statistical reports and the conceptual armature of macroeconomics.” This statement raises important questions about the ontology and epistemology of both the economy and economic statistics. How real is the economy? How real are statistics? And to what degree can the economy be observed through statistics?

The ontology of ‘the economy’ that underpins the dissertation is consistent with Polanyi’s (2001) argument about the embeddedness of markets. Polanyi argued that “There was nothing natural about *laissez-faire*; free markets could never have come into being merely by allowing things to take their course” (ibid., p. 145). Polanyi also insisted that “No society could, naturally, live for any length of time unless it possessed an economy of some sort; but previously to our time no economy has ever existed that, even in principle, was controlled by markets” (ibid., p. 45). In other words, some kind of economic sphere has always been part of human society, but the free market economy is artificial. It is perpetuated by the ideologies that conceive of it as separate from society and the natural world.

My view throughout the dissertation is that ‘the economy’ exists objectively, but the (incomplete) disembedding of the economy from society is the result of social and political construction. In Breslau’s (2003, p. 381) words, “the economy is both a brute reality and a construction”. This construction is a historical process that involved “developments as temporally remote from one another as the eighteenth-century attempts at national income statistics in Europe and the twentieth-century innovation of mathematical functions describing the relationships between macroeconomic variables” (ibid.). This view on the reality of the economy is quite different from a purely constructivist or relativist view that the economy does not exist separate from discourse or quantification.

What about the ontology of statistics? Desrosières (2001) argues that various attitudes about the reality of statistics are “closely associated with situational constraints specific to particular phases of the statistician’s technical, administrative, or managerial work” (ibid., p. 340). For example, national accountants tend to take a pragmatic view of statistics (ibid.). They are reluctant to make estimates, but, because of the immediate policy applications of their work, believe that “Even low-quality estimates ... are preferable to no estimates at all” (ibid., p. 345).

The position of this dissertation is that economic statistics are socially constructed. As such, they cannot objectively capture economic reality, as a purely empiricist view would suppose. Statistics reify and perpetuate the artificial separation between economy and society. This is most clearly demonstrated in chapter 3, which shows how statisticians have grappled with non-market activities in the System of National Accounts. Statisticians recognize that unpaid work, such as cooking, cleaning, and childcare, is ‘productive’ and essential to many peoples’ livelihoods. But because it cannot be exchanged – and thus does not enter into market dynamics – standard-setters have determined it to be ‘outside of the economy’.

This ontology and epistemology of statistics and the economy is broadly consistent with critical realism (on a realist approach to social science, see e.g. Dean, Joseph, Roberts, and Wight 2006).¹ As a realist philosophy of science, critical realism strongly accepts the existence of a real world independent from human cognition (Wight 2007). However, our attempts to observe reality is subjective and always involves interpretation (ibid.). Critical realism also proposes a stratification of reality into three levels – the empirical, the actual, and the real (Fletcher 2017, p. 183). The ‘empirical’ level comprises observable events mediated through human interpretation, while events at the ‘actual’ level occur whether or not we experience them, and the ‘real’ is the level of causal mechanisms (ibid.). Statistics exist at the empirical level. They are useful – but subjective and imperfect – tools with which we can attempt to describe ‘actual’ phenomena in the social world.

In sum, the ontological and epistemological views that guide this dissertation are as follows: (1) the economy exists objectively, but it cannot be fully separated from the wider social and natural world, (2) economic statistics are socially constructed by experts, and (3) statistics cannot objectively describe socio-economic reality. From this perspective, it is not possible for economic statistics to be entirely value-free.

The political nature of economic statistics

The statistics that appear in official databases, or in publications such as the World Bank’s *World Development Report*, are often assumed to be politically neutral. But scholars and policymakers have increasingly recognized that economic statistics are inherently political rather than objective “mirrors of the economy” (Herrera 2010). The indicator that has received the most criticism in recent years is GDP (Fioramonti 2013; Masood 2016; Philipsen 2015; Pilling 2018). For example, as Chapter 3 discusses in detail, GDP has been

1 I do not claim to consistently apply a critical realist ontology throughout the dissertation. However, critical realism has indeed shaped my views on statistics and how to study them, which is why it is briefly discussed here.

criticized for excluding unpaid work. This exclusion disproportionately renders women's work invisible and limits the usefulness of national accounts statistics for gender-aware policy (Hoskyns and Rai 2007; Waring 1999). At the same time, since the 1990s GDP has captured financial intermediation services (Christophers 2013). Does this mean that spending a day preparing meals and caring for sick family members is unproductive and outside of the economy, but financial intermediation is productive? According to the SNA, that is indeed the case.

The problems with applying 'Western' national accounting concepts worldwide were already understood before the first international standard for national income accounting was created in the late 1940s. As Morgan (2009) describes, when economist Phyllis Deane conducted field research in Central Africa and Jamaica in the early 1940s, Deane concluded that

in such economies as those she was trying to measure, national income accounts were not such as to enable the investigator to see sharp lines and clear elements, but rather 'a few large shapes in a thick fog'. (ibid., p. 22)

It is not only GDP that has these kinds of problems. Other literature has pointed out biases and shortcomings in statistics on government debt (De Vlieger and Mügge 2020; Hjertholm 2003), unemployment (Alenda-Demoutiez and Mügge 2019; Baxandall 2004), and trade statistics (Linsi and Mügge 2019), among others. As Chapter 4 argues, reductionism is a universal feature of economic statistics. As necessary as they are, there are limitations to what statistics can deliver.

Historical drivers of statistical harmonization

The emergence of globally harmonized official statistics is a relatively recent phenomenon. Its origins are in two historical developments. The first is the period of European colonialism, when colonial authorities collected a limited range of statistics – primarily agricultural and demographic statistics – for taxation and administration (Khan, Wales, and Stuart 2015, p. 3). Data collection in these areas was often continued by post-independence governments in African countries and other former colonies (ibid.).

The second is the formation of international organizations such as the IMF, World Bank, and United Nations in the aftermath of World War II. Harmonized economic statistics were crucial to the missions of these newly formed organizations, above all monetary stability and post-war reconstruction (Ward 2004b). Only 46 countries were UN members when the system was founded, most of which were industrialized countries. As political independence and decolonization accelerated and many postcolonial states joined the

UN, these countries were incorporated into international statistical standards such as the UN System of National Accounts (SNA) and the IMF Balance of Payments Manual (BPM) (ibid.).

According to Ward (2004b, p. 7), in the early decades of the UN and the Bretton Woods Institutions “an emerging consensus soon began to drive the development debate”. In particular, full employment was a central goal of all industrial countries and written into the mandates of the Bretton Woods Institutions and the UN (ibid.). “For the developed industrial countries, this objective was viewed as synonymous with poverty reduction, and it accounts for the statistical preoccupation with GNP, growth, and the national accounts” (ibid.). The SNA was especially important in the early post-war period of statistical harmonization. One reason is that national income is the basis for determining countries’ dues to the UN and other IOs.

Much of the intellectual groundwork of modern national accounting was established in the period following World War I, and these practices started to spread globally after World War II (Kendrick 1970; Speich 2011). In the postwar period, the rise of policies aimed at economic growth and the embrace of statistical tools to measure it went hand in hand (Schmelzer 2016). The diffusion of these tools was enabled by international organizations setting rules and providing technical assistance to harmonize statistical practices, first in the industrialized countries and then globally (ibid., pp. 23-24). An interviewee at the UN Statistics Division agrees that

Economic statistics was very central in the whole development philosophy. It was about creating efficiency, and GDP, investments, exports and imports that that would drive development. So economic statistics was very central to that process. (Interview 3)

Beginning in the early 1970s, IOs began to incorporate more social statistics into their through new household surveys like the United Nations Household Survey Capability Program (UNHSCP) (Khan et al. 2015, p. 4). At the World Bank, the first set of indicators to complement GDP were introduced in 1978 under Bank president Robert McNamara (Fioramonti 2013, p. 96). The introduction of the Living Standards Measurement Survey in 1983 by the Bank was another major step in statistics toward a development agenda focused on poverty reduction rather than growth alone (Khan et al. 2015, p. 4).

The global demand for data increased markedly from the 1990s onward. Among other things, the IMF launched the General Data Dissemination System (GDDS, described in Chapter 5); the Millennium Development Goals began in 2000, with major implications for statistical systems; and the World Bank and IMF introduced the data-intensive Poverty Reduction Strategy Papers (PRSPs) in the early 2000s as part of the Heavily Indebted Poor Countries initiative.

Overview of relevant theoretical approaches and causal mechanisms

Studying IOs from the inside and the politics of expertise and ideas

The global governance of statistics involves several functions, including developing methodologies and setting international standards; providing capacity building and technical assistance; monitoring countries' compliance with international standards and evaluating data quality. IPE literature has studied IO behavior extensively but has mostly overlooked the statistical units of IOs as consequential actors (with some exceptions, e.g. Clegg 2010; Harper 1998; Samuel 2014). Yet IPE theories of IO behavior give important insights that help explain statistical governance. IO behavior literature can be clustered into studies that focus on intra-organizational influences (generally constructivist and sociological institutionalist) and those that focus on extra-organizational influences (often explained with principal-agent theory).

For the most part, the empirical chapters take an intra-organizational approach. They investigate the inner workings of IOs – their organizational cultures, the roles of statistical departments, and the agency of statistical experts. Intra-organizational analysis focuses on the bureaucratic culture of IOs, the influence of ideas and norms on behavior, and the agency of IO staff and departments (Ban 2015; Broad 2016; Barnett and Finnemore 2004; Kentikelenis and Seabrooke 2017; Momani 2007; Reinold 2017; Vetterlein 2014). Many scholars emphasize knowledge and perceived expertise as a source of authority for IOs. For instance, in the 1990s under president James Wolfensohn, the World Bank rebranded itself as the world's 'Knowledge Bank' (Enns 2015; Mehta 2001). However, as Enns (2015, p. 68) argues, rather than signaling a departure from the Bank's long-standing preoccupation with liberalization and economic growth, the knowledge bank discourse "presented a narrow and reductionist notion of knowledge, characterizing knowledge as a form of capital to be leveraged for economic growth". The cognitive authority of IOs is also evident in the increasingly quantified nature of their work. The generation of benchmarks and indicators widely perceived as authoritative and legitimate, even when they are not, is an important source of informal power for IOs (Best 2012; Broome, Homolar, and Kranke 2018; Freistein 2016; Seabrooke 2012; Sending and Lie 2015). Until recently, this literature has surprisingly paid little attention economic indicators such as those of debt, national income, and trade, which are harmonized and published by IOs as well as crucial to the way they see the countries in which they operate.

Others have emphasized the agency of staff within particular IO departments, such as the area departments of the IMF (Chwieroth 2013) and the research units of the World Bank (Broad 2006; Enns 2015). However, the statistical departments of IOs have not been explicitly considered as actors in their own right. Broome and Seabrooke's (2012) concept of analytic institutions helps make sense of IO statistics departments. Analytic institutions

are “...the specialist units, departments, committees, adjudicatory bodies and others housed by or linked to IOs that develop the cognitive framework for understanding and solving policy problems” (ibid., p. 3). As “institutions endowed with analytical capacities for a programmatic purpose” (ibid.), statistics departments are influenced by professional norms shared by statisticians, but also work within the organizational constraints of the IOs in which they are situated.² The dissertation inquires how influential these IO departments and their staff members actually are in statistical harmonization efforts.

Extra-organizational analysis of IOs is often associated with principal-agent (P-A) theory. Unlike liberal IR theories, and in common with constructivism, P-A approaches accept that IOs can act independently from member states (Abbott and Snidal, 1998; Copelovitch 2010; Elsig 2012; Hawkins, Lake, Nielson, and Tierney 2006). The theory can help explain why IOs such as the World Bank sometimes act autonomously and at other times respond to demands of member states (Nielson and Tierney 2003). Although (powerful) member states are typically understood as principals, IOs can also be understood as both principals (of member countries, especially low-income ones) and agents (of powerful member states) simultaneously (Tamm & Snidal, 2014).

Despite being ontologically distinct, P-A and constructivist theories of IO behavior are not necessarily incompatible (Weaver 2007). They can help shed light on the external and internal dynamics of IO behavior, respectively (ibid.). For example, Gutner (2010) and Clegg (2010) both address aspects of IOs’ engagement with global development agendas and indicators, acknowledging P-A dynamics as well as internal factors. Gutner (2010) argues that the poor outcome of the IMF’s role in poverty reduction is an example of IO pathology resulting from the IMF’s dual role as principal and agent (ibid., p. 269). As the IO “least capable of embracing any bold new initiatives for poverty reduction”, the IMF was ineffective in working toward the poverty reduction goals delegated to it (ibid., p. 268). It is plausible that similar dynamics enter into statistical harmonization. Chapter 5 in particular asks whether the organizational culture and individual priorities of IOs influence how they cooperate with other organizations.

Principal-agent theory and other extra-organizational theories would assume that external factors shape the design of statistical standards. Member states do indeed delegate standard-setting to IOs. Thus, from a P-A perspective, IOs may be considered agents pursuing interests that potentially diverge from those of their principals. However, the evidence presented in Chapter 3 suggests that the direct involvement of member states effectively stops there. The reason is that the development of harmonized statistical methodologies is viewed as a highly technical task, and also a non-political one. This

2 However, see chapter 5 on the limitations of the concept of analytic institutions when applied to global statistical governance.

allows standard-setters to operate as members of a transnational epistemic community, as chapter 5 argues. Therefore, an intra-organizational approach is better suited to this topic.

The theoretical focus on ideas and expertise is especially important in the empirical chapters that address standard-setting, particularly Chapter 3. The SNA has undergone many changes over time, including the inclusion of the informal sector and subsistence farming. In other areas, notably the continued exclusion of ‘housework’, change has not materialized despite much contestation (Waring 1999; Wood 1997). I aim to explain why household services continue to be excluded from GDP despite changes in overlapping topics such as subsistence production. Following a constructivist institutionalist approach (Hay 2006, 2016; Schmidt 2008, 2010), I explain this continuity as the outcome of ideational path-dependency. This explanation is consistent with what Hall and Taylor (1996) call a cultural approach in historical institutionalism, which scrutinizes taken-for-granted ideas and norms. In the SNA, these include ideas about which kinds of activities count as economically productive and which do not, as well shared professional norms about the quality and reliability of official statistics.

Explaining the diffusion of statistical standards and practices

How is it that statistical standards like the System of National Accounts spread around the world? How have they traveled from the highly industrialized countries in which they originated to places that differ, sometimes starkly, in terms of labor markets, financial systems, informality, and so on? Such questions tie into the IPE and IR literature about diffusion. In general, diffusion literature investigates the “process through which ideas, normative standards, or ... policies and institutions spread across time and space” (Börzel and Risse 2012, p. 5).

Functionalist approaches suggest that international standards spread due to their role in reducing information asymmetry between states, and thereby transaction costs (e.g. Abbott & Snidal, 2001; Mattli & Buthe, 2003). Such a dynamic could plausibly explain the early spread of indicators such as Gross National Product among industrialized countries (Bos, 2009, pp. 31-48; Kendrick, 1970, pp. 306-311; Vanoli, 2005). But many questions remain as to how they reached the rest of the world.

Speich (2008, pp. 14-21) argues that, from the 1950s onward, the spread of GDP per capita and similar concepts was driven by experts who, thanks to these indicators, could “travel easily from one developmental case study to another. The performance of the Mexican economy could be used as a benchmark for Nigeria and the East African Community seemed comparable to Indonesia” (ibid., p. 21). Schmelzer (2016) argues that IOs like the OECD had a more hands-on role in promoting the spread of national accounting standards through setting rules and providing technical assistance.

Several authors suggest that powerful states and international organizations – particularly the United States and the United Kingdom, as well as the Bretton Woods Institutions – have historically been the drivers of GDP diffusion (Fioramonti 2013, pp. 40-43; Philipsen 2015, pp. 131-135; Schmelzer 2016, pp. 23-24). These arguments are coupled with a critical view of GDP's dominance with an emphasis on its biases and misuses in policymaking. What they suggest is that the global spread of GDP has taken the form of coercive diffusion (Lai et al. 2017).

Socialized diffusion is a more subtle process. Existing literature expects to find socialized diffusion “when actors attempt to solve problems or policy challenges in an environment that is rooted in uncertainty and bounded rationality” (Lai et. al 2017, pp. 961-962). Socialization can in some cases be a direct mechanism, as in transnational socialization promoted by norm entrepreneurs or other transnational actors (Zimmermann 2016, p. 101). Emulation is an example of an indirect socialization mechanism. In emulation, “Actors need to solve a problem or to overcome a crisis and look around for ‘best practices’ and institutional solutions...” (Börzel & Risse, 2012, p. 5).

Others argue that domestic factors drive the spread of statistical standards. Herrera (2010) explains Russia's adoption of the SNA as a norm-centered process driven by national statisticians. In the case of South African unemployment statistics, Alenda-Demoutiez & Mügge (2019) argue that, despite their poor fit to the domestic labor market, ILO standards were not imposed but rather embraced by South African statisticians looking to guard against accusations of politically motivated bias.

The empirical chapters suggest that the diffusion of statistical practices cannot be explained by just one kind of mechanism. Chapter 5 addresses several ways that IOs attempt to promote the spread of statistical norms and practices. They do so through a combination of formal and informal mechanisms. Whereas the focus of that chapter is on active, top-down harmonization efforts, chapter 6 focuses on bottom-up mechanisms. That chapter provides a nuanced account of the diffusion of GDP to China. While that process was driven above all by domestic variables, international socialization provided the necessary conditions for diffusion to occur.

Localization, translation, and practical constraints on implementation of statistical practices

Even once standards are created and promoted around the world, there are still questions about how, and the extent to which, they are actually implemented at the domestic level. While there is some overlap with the level of diffusion, the focus of this sub-section is on the actual statistical practices within countries. There are two main issues at stake. First, what are the main practical constraints at the domestic level that stand in the way of complying with international standards? And second, what are the means by which countries adapt international standards to domestic preferences, ideas, and socioeconomic realities?

The most important practical constraint for developing countries is low statistical capacity. Political economy literature on statistical capacity highlights the scope, causes, and consequences of missing and unreliable data (Dargent, Lotta, Mejía, and Moncada 2018; Devarajan 2013; Jerven 2013; Taylor 2016). A notable example is Jerven's (2013) *Poor Numbers*, which demonstrates how low statistical capacity results in incomplete and often unreliable national accounts statistics in sub-Saharan African countries. World Bank economist Shantayanan Devarajan (2013) has described the situation of statistical capacity in sub-Saharan Africa as a "statistical tragedy". Devarajan (*ibid.*, p. S11) argues that "in presenting GDP per capita for many African countries, we cannot be sure of either the numerator or the denominator". Within countries, low statistical capacity hinders the abilities of policymakers to make well-informed choices and of analysts to make well-reasoned assessments (*ibid.*). Globally, it damages cross-national comparability, foreign policy (including trade and development policy), and limits the usefulness of statistics in international diplomacy. Chapter 5 discusses and builds upon the statistical capacity literature.

However, statistical capacity is not the only reason for countries to diverge from international standards and 'best practices'. If countries cannot or will not comply with standards, the standards fall short of fulfilling their stated purpose (even if they still have a symbolic function). Governments also might have reason to deviate from standards, or even to 'manipulate' official statistics (Wallace, 2016; cf. Aragão and Linsi 2020). Whatever the reason, incomplete compliance means incomplete harmonization. To make sense of this, the literature on localization and translation is relevant. IR literature on localization emphasizes the ways that international norms are frequently contested or modified (Acharya 2004; Eimer et al. 2016; Heilmann and Schulte-Kulmann 2011). In Acharya's (2004, p. 245) words, localization is the "active construction ... of foreign ideas by local actors, which results in the former developing significant congruence with local beliefs and practices". Countries on the receiving end are thus viewed as agents rather than passive recipients. Ban (2016, p. 18) suggests a similar argument with the concept of translation: "rather than 'copy and paste' ideas developed in foreign 'labs,' receivers tend to actively filter and even reshape these ideas before 'adoption'".

In Chapter 6, we find that even though diffusion of GDP to China was a bottom-up process, GDP was at first only partially implemented domestically. The ability of Chinese policymakers and statisticians to localize or translate the SNA was limited. This was a result of unique challenges posed by importing a foreign national accounting system to a domestic setting in which an alternative system already existed, and in which national ideology clashed with the economic ideas underpinning the SNA.

3

Hidden in plain sight: Unpaid household services and the politics of GDP measurement³

Abstract

Gross domestic product (GDP) is one of the world's most influential and widely cited economic indicators. However, outside of the industrialized, market-based context in which the indicator was first designed, GDP measurement suffers from a number of biases and blind spots. This chapter zooms in on one of these: the exclusion of unpaid household services from the production boundary of the System of National Accounts, the international standard underpinning GDP methodology. While GDP has expanded over time to include activities as diverse as financial services and the informal sector, the treatment of unpaid household services has remained unchanged. Why is this? I find that staff in the statistical departments of international organizations such as the United Nations, International Monetary Fund and World Bank have a tremendous degree of agency in the governance of GDP. While these statisticians are aware of and engage with criticisms, they reject the inclusion of unpaid household services based on shared professional norms and economic ideas.

³ This chapter is an adapted version of a published journal article: DeRock, D. (2021). Hidden in Plain Sight: Unpaid Household Services and the Politics of GDP Measurement. *New Political Economy*, 26(1), 20-35. The chapter also builds upon the following online publication: DeRock, D. (2019). Unpaid Work and the Governance of GDP Measurement. *E-International Relations*. <https://www.e-ir.info/2019/01/11/unpaid-work-and-the-governance-of-gdp-measurement/>

Introduction

Few numbers are as ubiquitous in political and economic analysis as gross domestic product (GDP). This powerful indicator is enlisted to rank and compare national economies, it influences lending and investment decisions, and is often taken as a proxy for well-being (Stiglitz et al. 2010). Because they are produced by governments and based on internationally harmonized guidelines, GDP and other economic indicators appear to be objective and unbiased. But statistical concepts originally designed for industrialized market economies do not travel seamlessly to other kinds of socioeconomic settings – for example, low-income areas with high levels of subsistence and informal activity.

The result is that certain economic activities are captured while others are rendered invisible in GDP figures (Morgan 2009; Mügge 2020a). One topic in particular has been a thorn in the side of economic statistics for decades: the measurement of unpaid household services (or unpaid services, for conciseness).⁴ These services largely overlap with care work and domestic labor and are disproportionately performed by women. While GDP has expanded over time to include other forms of unpaid work such as subsistence agriculture, the exclusion of unpaid household services has only become more concrete in international statistical standards. This exclusion has persisted alongside considerable contestation and deliberation among economists, statisticians and feminist scholars. What explains the persistent exclusion of unpaid household services from GDP?

Two competing perspectives have dominated debates about this exclusion. One is rooted in feminist – and, to some extent, postcolonial – critiques of the microeconomic theories underlying GDP methodology. The other is the pragmatic and seemingly depoliticized perspective advanced by statisticians themselves. Feminist scholars have demonstrated that the exclusion of unpaid household services from official statistics introduces a major bias into economic analysis and policy (Benería 1992; Waring 1999, 2003). The implicit argument is that this oversight is a result of institutionalized gender bias. Economic statisticians, on the other hand, contend that this kind of work is difficult to quantify and falls outside of the market dynamics that GDP figures should capture. Thus, including these activities would compromise the reliability and cross-national comparability of GDP. The international political economy (IPE) literature has paid little attention to these

4 I use the term ‘unpaid household services’ throughout to refer a range of activities (specified below in the subsection ‘The Household Sector in the 1993 SNA Revision’) performed by households for their own use. While other terms with clear political connotations are available, these alternatives are either too broad or too restrictive. For example, ‘unpaid work’ and ‘unpaid labor’ are unsuitable in this context because several forms of unpaid work are included in GDP. ‘Domestic labor’ and ‘care work’ both have the advantage of emphasizing the gendered character of the work, but arguably leave out some of the activities in question, such as transportation and minor home repairs. The term ‘unpaid work’ is used occasionally in the chapter in reference to a broader category of activities including, for example, subsistence agriculture.

debates (cf. Hoskyns and Rai 2007). Yet IPE approaches are well suited to address two key characteristics of GDP that have largely been ignored: the global governance of GDP measurement and the agency of the experts who shape it.

This chapter demonstrates that economic statisticians, specifically those working as staff in the statistical departments of international organizations such as the International Monetary Fund (IMF) and the United Nations (UN), have a high degree of agency to determine the concepts and methodologies underlying GDP. Contrary to what would be anticipated by interest-based approaches, the preferences of member states, and of any other external actors, have little impact on the governance of GDP. Rather, the authority that statisticians derive from their positions as specialized experts leaves them largely immune from formal constraints. Rather than ignoring the problem, statisticians are well aware of criticisms but reject the inclusion of unpaid services on the basis of shared ideas about the limits of ‘the economy’ and the policy applications of official statistics.

These conclusions are based on an extensive literature review, official reports and minutes of meetings, and in-depth interviews. 29 interviews (some with multiple participants, for a total of 36 interviewees) were conducted between 2017 and 2019 with current and former staff and directors of statistical agencies including the IMF, World Bank, United Nations Statistics Division, UN ESCAP, private development consultants, and national statistical offices in Laos, Thailand, and Ghana. Document analysis focuses on the two most relevant international bodies, namely the United Nations Statistical Commission and the Intersecretariat Working Group on National Accounts. The latter is composed of experts from the IMF, World Bank, United Nations, OECD and the European Commission.

There are two main contributions. First, the chapter adds historical and institutional context to ongoing debates about the limitations of GDP (Fioramonti 2013; Philipsen 2015; Pilling 2018; Masood 2016). On the whole, scholars have taken aim at GDP without recognizing how and by whom it is governed. By taking the criticisms seriously but taking a step back from the debates, I show why there is a gap between aspirations (of critics and advocates alike) and actual outcomes. This approach takes up Mügge’s (2016, p. 422) call to study indicators as “powerful, institutionalized ideas”. Second, the findings contribute to the wider IPE literature, particularly constructivist perspectives on IO behavior (e.g. Barnett and Finnemore 2004; Broad 2006; Enns 2015; Vetterlein 2014) and expert-centered theoretical approaches (e.g. Dersnah 2019; Kunz et al. 2019; Leander and Waever; 2018; Seabrooke and Tsingou 2016; Sending 2015).

This chapter proceeds with a review of literature on the politics of GDP, unpaid work and plausible theoretical approaches. The third section describes the global governance of GDP measurement and explains why it is largely insulated from formal constraints. The main body of the chapter is an empirical analysis of the GDP revision process. This section first

demonstrates the resistance of economic statisticians to the inclusion of unpaid household services in the GDP production boundary, and then closely analyzes expert deliberation in the revision process leading up to the 1993 System of National Accounts. This revision process, which spanned approximately ten years, was a decisive period for the treatment of unpaid services. A fifth section concludes and discusses the limits and possibilities of fundamental changes to GDP measurement in the future.

GDP through the looking-glass: debates and theoretical approaches

Contested Perspectives on Unpaid Work and GDP

The measurement of GDP has come under increased academic scrutiny in recent years. Several authors have argued that GDP is a misguided and potentially harmful benchmark for policymaking (Fioramonti 2013; Hoekstra 2019; Philipson 2015), while others have argued that it is simply too narrow of a measurement from which to draw any conclusions about the well-being of societies (Coyle 2014; Stiglitz et al. 2010). Yet others have echoed these criticisms and shed light on the historical rise of GDP from a little-known statistic to, arguably, the world's most powerful number (Lepenies 2016; Masood 2016; Schmelzer 2016). The criticisms span a wide range of issues, from environmental depletion to the impact of free digital services on well-being.

Another cluster of literature focuses on one crucial aspect of GDP measurement, namely the production boundary – a conceptual line drawn between economic and non-economic activity. As Coyle (2017, p. 7) describes it, the production boundary “distinguishes paid-for activities in the market economy from unpaid activities, which are considered outside the productive sector”. In GDP, “what is defined as economic activity is, literally, anything deemed to sit inside a designated ‘production boundary’” (Christophers 2011, p. 115). With some notable exceptions such as the inclusion of financial services (Christophers 2011, 2013), the production boundary has been one of the most consistent features of GDP methodology (Bos 2009, p. 40). This is certainly true for unpaid household services, which have been excluded for as long as GDP has been in existence. This continuity should not be mistaken for a lack of controversy. On the contrary, the measurement of unpaid work has long been a contentious issue in debates about national accounting, both among economic statisticians as well as in academia and social movements.

The category of unpaid household services corresponds to ‘own-account services’ in the terminology of the SNA. We might think of many of these services alternatively as housework, care work, or domestic labor. This includes activities such as childcare, cleaning, cooking and care for the sick and elderly. Hoskyns and Rai (2007, p. 297) maintain that “[w]ithout unpaid services and their depletion being measured and valued, predictions

are likely to be faulty, models inaccurate and development policies flawed”. Folbre (2014, p. 1130) attributes the lack of effort to measure unpaid services to biases in mainstream economic theory, which in turn “shaped the assumptions embedded in national censuses and income accounts”. Since unpaid household services are disproportionately carried out by women (International Labour Organization 2016), failing to measure them introduces a gender bias into economic data and analysis (Elson 2005; Folbre 2014; Miranda 2011; Waring 1999). The problem is particularly acute in developing countries, where the overall amount of time spent on unpaid services is higher (International Labour Organization 2016, p. 20).

According to the International Labour Organization (2018, p. 43), based on time-use data from various years for 64 countries, women spend an estimated three times longer than men per day in unpaid care work. The amount of time spent by women on unpaid care work varies from a maximum of 5 h and 45 min (Iraq) to a low of 2 h and 48 min (Taiwan) with a median of 4 h and 29 min (Austria and Germany) (ibid.). Typically, as countries industrialize, a large part of household production shifts to the market (Miranda 2011, p. 6). This shift from non-market to market “... translates into a rise in income as measured by income and production aggregates and gives a false impression of an improvement in living standards” (ibid.). A classic example of this phenomenon, which has been variously attributed to several late economists, is that marrying one’s (ostensibly female) cook or housekeeper would lead to a reduction in GDP (Lequiller and Blades 2014, p. 121). This is the case “even if, as a wife, her household activities might not have changed or might even have increased” (Benería 1992, p. 1548).

Moreover, the categories of globally harmonized official statistics – such as formal versus informal, or productive versus unproductive – often make little sense in local contexts that do not resemble developed market economies. Waring (2003, p. 36) demonstrates this point with an example:

The woman goes to collect water. She uses some to wash dishes from the family evening meal (unproductive work) and the pots in which she previously cooked a little food for sale (informal work). Next, she goes to the nearby grove to collect bark for dye for materials to be woven for sale (informal work), which she mixes with half a bucket of water (informal work). She also collects some roots and leaves to make a herbal medicine for her child (inactivity). ... She will also collect some dry wood to build the fire to boil the water to make both the medicine and the dye (active and inactive labour). All this time she will carry the baby on her back (inactive work).

These concerns are nothing new. In fact, they were matters of fundamental concern in early debates over national accounting principles. Reid (1934, p. v), more than eighty years ago, warned that a singular focus on “that part of our economic system which is

organized on a price basis” had blinded economists to productive work of the household, “our most important economic institution”. Pigou (1920, p. 11), in contrast, argued that national accounts should only include those things that can “be brought directly or indirectly into relation with the measuring rod of money”. In the postwar period during which national income accounting rose to prominence, some economists argued that it would be impossible to meaningfully compare the economies of industrialized and non-industrialized countries (Dominguez 1947; Frankel 1953).

One of the most prominent issues in these debates was a distinction between the so-called “money economy” (Ady 1962, p. 52), which can be relatively easily captured in statistics, and more elusive non-monetary or non-market activity (Frankel 1953, pp. 165–6; Kuznets 1949; Rao 1953, pp. 179–87; Samuels 1962, p. 170). This “countability bias” (Mügge 2020a, pp. 9–12) still persists in that goods and services with monetary values (or for which values can easily be imputed) are more readily quantified. Kuznets (1949, p. 206) – widely considered the founding intellectual of GDP – insisted in 1949 that applying the statistical conventions designed for industrialized countries to non-industrialized countries would lead to unacceptable distortions (*ibid.*, p. 211). He argued that “... if national income is to be merely a measure of goods exchanged for money, an estimate had better not be attempted for pre-industrial countries at all” (*ibid.*).

Although the status of unpaid household services in GDP has remained unchanged, the issue has not been ignored entirely. The Social and Gender Statistics section of the UN Statistics Division, for example, has taken a leading role in designing and implementing time-use surveys. Between 2005 and 2015, “75 countries collected time-use statistics through a time-use survey or have included a time-use module in a multipurpose household survey” (United Nations 2015, p. 88). The increased recognition of unpaid household services and “time poverty” (Bardasi and Wodon 2010) is due in large part to decades of research, theorizing and advocacy by feminist economists in response to dominant microeconomic theories about the household and labor markets (e.g. Folbre 2009, 2014; Goetz 1997). Importantly, there is by no means a consensus among feminist scholars that inclusion in GDP is the most desirable way forward for measuring unpaid household services. Esquivel (2011), for example, argues that an overemphasis on the GDP production boundary may be in fact be hindering progress on time-use surveys and other forms of data collection for gender-sensitive policy.

Theorizing Change and Continuity in GDP Measurement

The exclusion of unpaid household services is important to understand in its own right. As a case study, it also adds to a broader understanding of the origins and governance of global statistical standards. The most relevant actors are international organizations (IOs) such as the World Bank, IMF, and UN, as well as expert groups comprised of staff of these IOs. Two strands of literature in particular illuminate how GDP methodology has been

shaped by these actors over time. The first is a path-dependency approach that emphasizes institutional change or continuity in statistical standards. The other is a constructivist approach that highlights the role of expertise in global governance.

Considering the question from a path-dependency angle highlights the potential for change or continuity in global statistical standards. Hall and Taylor (1996, pp. 939–40) distinguish two broad tendencies within historical institutionalism: a calculus approach and a cultural approach. The calculus approach assumes that actors adhere to institutions because deviation would lead to worse outcomes than adherence (*ibid.*). A cultural approach emphasizes the taken-for-granted nature of some institutions, which allows them to avoid scrutiny (*ibid.*). In other words, “Institutions are resistant to redesign ultimately because they structure the very choices about reform that an individual is likely to make” (*ibid.*). This is consistent with Hay’s (2006, p. 65) description of ideational path-dependency, in which “it is not just institutions, but the very ideas on which they are predicated and which inform their design and development, that exert constraints on political autonomy”. The focus on shared ideas and norms of statisticians moves the analysis from the relatively rigid assumptions of historical institutionalism (e.g. Mahoney 2000) to a more agent-centered constructivist institutionalism (Hay 2006, 2016; see also Schmidt 2008, 2010 on discursive institutionalism).

Constructivist and expert-centered approaches direct attention to questions of ideas, agency and global governance. Several constructivist studies of IOs (Barnett and Finnemore 2004; Broad 2006; Enns 2015; Vetterlein 2014) demonstrate that IOs are not unitary actors with fixed interests. Rather, individual staff and departments shape IO behavior (Ban 2015; Kentikelenis and Seabrooke 2017; Momani 2007; Reinold 2017). Statisticians have received little attention as agents within IOs. The statistical departments of IOs are best understood as “analytic institutions” (Broome and Seabrooke 2012). These are “the specialist units, departments, committees, adjudicatory bodies and others housed by or linked to IOs that develop the cognitive framework for understanding and solving policy problems” (*ibid.*, p. 3). Analytic institutions differ from epistemic communities (Haas 1992) in at least one important respect: “they are not free-floating or autonomous . . . , but institutions endowed with analytical capacities for a programmatic purpose” (*ibid.*, p. 4).

Experts can gain leverage, and even moral authority, from their specialized knowledge and experience (Ban and Patenaude 2018; Davis, Kingsbury and Merry 2012; Tsingou 2015; Seabrooke and Tsingou 2016; Seabrooke and Wigan 2016). Expert knowledge endows IOs with authority over issue areas and allows them to dominate the framing of issues, such as the World Bank on hunger (Sridhar 2007) and poverty (Vetterlein 2012). Enlisting the help of external experts can also enable mission creep into new areas of governance (Littoz-Monnet 2017). Expertise and knowledge are sources of power largely because they are perceived as technical, and thus non-political (Haas 1992). But this perception

obscures the deeply political nature of expertise in governance (Dersnah 2019; Kunz et al. 2019; Leander and Waever 2018; Sending 2015). As Desrosières (2000) pointed out, the profession of official statistics is “at one and the same time, scientific – directed at the production of knowledge – and social – directed at the production of a common language as a foundation for debate on social issues” (ibid., p. 173). The choices made by experts have distributive consequences and are inspired by normative orientations, even if these are not made explicit.

Knowledge is crucial in the constitution of objects in international politics, including ‘the economy’ itself (Allan 2018). The emergence of the economy as a distinct object, separate from the social and natural world, owes a great deal to the emergence of statistical indicators (Breslau 2003; Mitchell 2002, 2005). In Breslau’s (2003, p. 380) words, “No one has ever seen the economy or touched it except through statistical reports and the conceptual armature of macroeconomics”. Polanyi (2016, p. 400) proposed a similar ontology of economic statistics in the 1922 article ‘Socialist Accounting’ (*Sozialistische Rechnungslegung*):

History in fact directly points to the inverse relationship of dependence between accounting and economic theory: accounting is historically not a practical application of economic theory; on the contrary, economic theory developed historically through the interpretation, analysis, and systematizing of accounting concepts.

The ambition of statisticians to remain objective in their work does not imply ignorance of the social implications. Official statistics are valuable public goods and there is nothing inherently malicious about them. Ideally, at the domestic level, national statistical systems are independent from central governments in order to prevent manipulation. In this light, the goal of objectivity is laudable. With that said, the potential danger of depoliticization is that it can mask uneven power relations. By framing problems and solutions as politically neutral, technocratic actors camouflage the antagonisms and structural inequalities inherent in development issues (Petiteville 2018; Rajão and Duarte 2018; Telleria 2017). Ideas originating from the Global South, including influential ideas related to human development and sustainable development, are deeply influenced by the local contexts and origins of the actors who advocate for them (Acharya 2016). Governance arrangements that remain insulated from ideas originating from a Global South context might ignore local particularities such as structures of work, care and production that differ from those in highly developed countries.

The theoretical approach followed in the analysis below draws from the constructivist institutionalism and expertise literatures as a basis for explaining non-change in the production boundary. Expertise endows statisticians with a high level of autonomy. Yet, the agency of economic statisticians is not distributed equally – those employed by or

associated with IOs such as the World Bank and IMF, or with experience working in European and North American bureaucracies, have the most influence. This claim is supported in the following section. The unique governance structure has allowed for a relatively undisturbed path-dependency. In the absence of strong constraints and explicit demands from member states or IO executives, the most important explanatory factor is the consistency over time of ideas about the measurement of unpaid household services.

The Global Governors of GDP

Much of the intellectual groundwork of modern national accounting was laid in the period following World War I (Kendrick 1970; Studenski 1958). Gross National Product (GNP), the predecessor to GDP, first emerged in a small number of industrialized countries in the 1930s and 1940s and attracted the attention of policymakers in part through its role in economic planning during the Second World War (Kendrick 1970). Since then, GDP has become a global institution. Although GDP grabs the most attention, it is only one of many indicators derived from the System of National Accounts (SNA). The SNA is an internationally harmonized “set of recommendations on how to compile measures of economic activity in accordance with strict accounting conventions based on economic principles” (ISWGNA 2008, p. 1). In other words, it is a framework for measuring the total economic activity of a country.

The SNA was first published in 1953, followed by revisions in 1968, 1993, and 2008, with a new revision currently in progress. The length and detail of the SNA has grown substantially over time. Since the disappearance of alternative national accounting systems in post-communist states, it is now the only internationally accepted standard (Herrera 2010, p. 18). The development and revision of the SNA is now carried out by an intersecretariat working group composed of statisticians from five international organizations. Prior to the 1980s, it was carried out by one of these, the United Nations Statistical Office (UNSO, later renamed the United Nations Statistics Division, UNSD).

The UNSO (now UNSD) was formed at the Nuclear Session of the UN Statistical Commission in 1946 (Ward 2004b, pp. 37–8). It emerged out of the League of Nations, and was created in large part to establish harmonized economic statistics in support of Marshall Plan reconstruction (*ibid.*, pp. 43–9). One of the earliest projects of the newly formed Statistical Commission – carried out chiefly by the UNSO – was to draft the first version of the SNA (*ibid.*). At that time, there were only 46 UN member states, most of which were industrialized countries (*ibid.*, p. 6). Over the following decades, the number of member states increased along with the wave of independence and decolonization in the 1960s. It became more difficult to balance different policy aims in the international statistical system (*ibid.*). Nonetheless, GDP quickly gained a solid foothold in development

policy (Speich 2008, pp. 14–21). According to Ward (2004b, p. 7), in the early decades of the UN and the Bretton Woods Institutions, “an emerging consensus soon began to drive the development debate”. The concept of full employment was a central goal of all industrial countries and was written into the mandates of the UN, IMF and World Bank. “For the developed industrial countries, this objective was viewed as synonymous with poverty reduction, and it accounts for the statistical preoccupation with GNP, growth, and the national accounts” (ibid.).

Since the early 1980s, responsibility for the SNA has been shared between five international agencies in the Intersecretariat Working Group on National Accounts (ISWGNA). The participating agencies are the UNSD, the IMF, the World Bank, the Organisation for Economic Cooperation and Development (OECD) and Eurostat. The ISWGNA was established in part because the costs were becoming too high for the UN alone (Interview 3). The ISWGNA and its members are overseen by the United Nations Statistical Commission, “the apex entity of the international community of official statistics” (ibid.). Despite the importance of collaboration between the five IOs, the SNA is still often seen as a UN initiative. This is due to the proximity of UNSD to the Statistical Commission and the fact that the SNA was long referred to as the UN System of National Accounts (ibid.).

The ISWGNA is responsible, among other things, for bureaucratic tasks such as planning meetings and deciding who will attend. The meetings, which include a rotating group of country experts in addition to ISWGNA members, are referred to as Expert Group meetings. However, the country experts present at Expert Group meetings have less influence over the content of the SNA than the permanent members for several reasons. For one thing, many of the background documents that are considered during meetings are written by members of the ISWGNA. These members have more time to write these documents compared to country experts, who are typically in charge of national accounts in their own countries (Interview 2). Another reason for this imbalance has to do with language and training.

There’s a bit of a problem, in that efforts are made to make sure it’s regionally diverse. And so you’re trying to include people from Asia, from Africa, from Latin America. That can sometimes be problematic on two counts. The first thing is whether someone would have the same depth of knowledge as some of the others. And there can be a bit of a problem about language. All of this is done in English. And the people who go to meetings in OECD and Eurostat, whether they’re English mother tongue or not, are used to working in that sort of environment. If you have somebody from [another region], they are not quite as comfortable working in English as others. ... So, for both of those reasons, it tends to be the developed countries in the ISWGNA that tend to dominate the discussions. Not exclusively, but to some extent that happens. (ibid.)

The ease of communication for statisticians comfortable with the working language, and the privilege given to a specific body of knowledge – namely, the national accounting practices that originated in the US and UK and evolved in close connection to the UN system – contribute to an uneven distribution of influence. Expert Group meetings do indeed include regionally diverse country experts, and statisticians from developing countries are by no means excluded from these meetings, nor from the international statistical system more broadly. Yet, the permanent members of the ISWGNA and those with experience in European and North American bureaucracies and international organizations do have more agency in the SNA revision process. According to a former ISWGNA member, “It is largely, not exclusively, but largely up to the UN, the World Bank and the IMF to speak up for developing countries, to the extent they don’t speak up for themselves” (Interview 2).

There are very few formal constraints on the ISWGNA. The most straightforward constraint is the mandate of the Statistical Commission, which was not enforced until after the 1993 SNA. After the 1993 revision, which took longer and resulted in much more substantial change than initially planned (Ward 2004a), the Statistical Commission increased its oversight. Now the ISWGNA submits a list of priority issues to the Commission prior to the start of the revision. Once agreed upon, the ISWGNA is mandated to deal only with these issues (Interview 4). In the SNA revision process, ISWGNA members work collectively toward producing an updated manual, not as representatives of the missions of their respective organizations. According to a former ISWGNA member, any instances of conflict stemmed from personal convictions rather than pressure to act on behalf of international organizations. For instance, in both the 1993 and 2008 revisions, a representative of Eurostat – the organization with the most funding of the five – pressured the other members into accepting a change in the SNA by threatening to not approve the final version (Interview 2). In this example,

It’s not that there’s somebody in Brussels leaning on them to say something. It’s basically they want to rule the statistical world – ‘we think it’s good, therefore it’s good for everybody else’. (ibid.)

Nor do the statistical departments of the constituent IOs face any prohibitive external constraints from member states or other actors. According to a former acting director of UNSD, “Apart from having to comply with all the millions of UN rules, we were pretty independent in our work. I’ve never noticed any political pressure of any kind. I mean they were just not interfering in contents of handbooks or publications apart from the regular editing process of course” (Interview 5). In this environment of expertise-based autonomy, the ISWGNA has a high degree of latitude to make – or at the very least submit to the Statistical Commission for approval – changes to GDP methodology as its members see fit. The following section traces the choices they made in the 1993 revision with respect to unpaid services.

Ideational Path Dependency in the SNA Production Boundary

Between publications of the SNA, the ISWGNA takes up the complex job of revising the international standards. Out of the three official revisions, the revision process of the 1993 SNA was the period during which unpaid household services was most prominently on the agenda. The topic was not discussed at all in the 2008 revision, and the relevant sections of the 1993 and 2008 SNA manuals are nearly identical. During the revision, which began in 1982 (Vanoli 2005, p. 104), the production boundary was discussed several times in both Expert Group meetings and within the UN Statistical Commission. Ultimately, changes were made to the production boundary. Yet, the changes that were made had the effect of reinforcing the exclusion of unpaid household services while including other activities. The 1993 SNA made these exclusions explicit for the first time, formally cementing this historical idea into the international standards.

This section highlights the choices made with regard to unpaid services and the justifications for these choices given by the ISWGNA. Throughout the process, the problem is clearly recognized by statisticians but changes are rejected. The first part of the section demonstrates the reluctance of national accountants to include unpaid household services in the production boundary. The second part is an analysis of official reports from the 1993 revision process with a focus on the arguments made for the exclusion. These arguments, which ultimately rely upon historical precedent, reflect an undisturbed part-dependency made possible by the autonomy of the ISWGNA (and the UNSO before that).

Outside the Market, Out of Mind: Expert Views on Unpaid Household Services

The 1993 SNA states that the reluctance of the ISWGNA to include unpaid household services in the SNA production boundary is explained by a combination of factors: the isolation of these activities from markets, the difficulty of estimating monetary values, and “adverse effects ... on the usefulness of the accounts for policy purposes and the analysis of markets and market disequilibria” (ISWGNA 1993, p. 149). A former ISWGNA member echoed these concerns about valuation, comparability and isolation from markets:

It is an issue that is very topical at the moment. But it’s actually been there for decades, lurking around. On the whole and by and large, most national accountants say, we recognize that unpaid housework is really important, but my lord it’s difficult to put a value on it. And if we put a value on it, and we added it into GDP, how would you know whether you’re doing it consistently over time or making comparisons across countries? It’s okay to do it, but could you do it a little bit apart from the main national accounts? (Interview 2)

Not all economic statisticians and data users agree on whether or not GDP should be expanded to include unpaid services. Data users who are interested in national accounts

data for administrative purposes tend to prefer a narrower production boundary for the sake of maximizing reliability and comparability. Those who use the data for economic analysis tend to prefer a more inclusive production boundary.

If you're really fixated by administrative purposes, especially on a cross-country basis, you might prefer to leave out the informal part, so that you can more strictly compare one country to another. But if you're interested in a time series then, in the sense of doing economic policy analysis, you might well say, 'well I'd sooner have a bad estimate of something than no estimate'. So, this trade-off between the two is quite problematic. And I think that is fundamental of where we're at at the moment. (Interview 2)

A national accounting expert at the Economic Statistics Branch of UNSD made similar remarks. According to the interviewee, debates about potential revisions to GDP measurement often come down to the question of 'what's the purpose?'

My view, and it may be a bit of a conservative view or a narrow view, is that there's one key reason why nation states invest in something like the national accounts. And to me that's primarily because they care about employment and they care about taxation. And the national accounts allows them to model and forecast and, you know, look at the relationships that lead to both of those. So, volume growth in GDP is strongly tied to employment outcomes. Current price GDP, probably tied to taxation.... The informal economy, the household sector, they're important to understand for other reasons, but you're not going to be designing your monetary or fiscal policy to impact on those, and in fact there's going to be very little government policy that is directly targeted at changing those. (Interview 1)

This divide was also apparent at a 2015 conference in Paris, hosted by the OECD and the International Association for Research in Income and Wealth, called "W(h)ither the SNA?". While most participants supported an expansion of GDP, a "significant minority of people" emphasized the difficulty of implementing such changes and the increased demands it would place on national accountants (Interview 2). The latter were those concerned with the policy applications of the data.

If you're the Ministry of Finance, for example, it doesn't matter what you're doing with housework. You can't tax housework, and so, and so there was quite a tension there. But a lot of, mostly the people who were defending the status quo were people who were concerned with administrative uses. And the people who wanted the massive expansion were the ones who wanted to do analysis. That's a bit simplified, but not much. (ibid.)

These practical arguments – which include isolation from markets, data collection and valuation challenges, reliability and comparability, and policy applications – reflect

professional norms that place high value on the reliability and comparability of official statistics. Yet, they also coincide with state interests in employment, fiscal and monetary policies. Because household services and care work are not imagined as productive work, they are not considered relevant for these policy areas. In contrast, feminist economists have convincingly demonstrated that gender gaps – including in unpaid work – have far-reaching effects on macroeconomic outcomes (Seguino 2019). Furthermore, whereas attempts have been made to incorporate other sectors and activities that are hard to measure, no such effort was made for unpaid household services. The informal sector, for example, is difficult to measure directly and current data rely heavily on estimates (International Labour Organization 2013a, p. 244), yet it has been inside the SNA production boundary since 1993. And imputations are applied in other areas of the SNA, notably financial services and owner-occupied housing. In contrast, recent innovations toward measuring and valuing unpaid household services (such as time-use surveys) have primarily gone on outside of national accounting and on a comparatively limited scale.

The Household Sector in the 1993 SNA Revision

The report of the 1981 Statistical Commission session recognizes the need to reconsider the household sector, as the following passage indicates:

For most developed countries, [imputations for non-market activity] are of relatively minor significance in present estimates of the gross domestic product (GDP). For developing countries, however, they may be much more important. For both developed and developing countries, furthermore, there are demands for new kinds of imputations beyond those presently included in the gross domestic product. (UNSC 1981, p. 12)

But a cautious attitude prevailed, as several delegates expressed “a strong resistance to losing sight of the transactions-oriented base [of the SNA], not only because its data are likely to be relatively much firmer but also because market transactions are often the vehicle for government actions” (UNSC 1981, pp. 12–13).

The SNA states that the biggest problem in determining the activities included in the production account is deciding how to treat “... activities that produce goods or services that could have been supplied to others on the market but are actually retained by their producers for their own use” (ISWGNA 2008, p. 6). One of these gray areas is own-account production, which includes activities such as subsistence farming. Another is own-account *services*, a category that includes “the preparation of meals, care and training of children, cleaning, repairs, etc.” (ibid.). The 1993 SNA confirmed the inclusion of own-account goods and added some activities (including water collection and repairs to

buildings) inside the production boundary on these grounds (Harrison 2005, p. 150).⁵ Services, in contrast, were explicitly excluded.

Two main discursive justifications for this exclusion emerge from reports of the ISWGNA and the UN Statistical Commission during the 1993 revision process. The first justification can be labelled the ‘market criterion’. The second is a distinction between non-market goods and non-market services – a distinction that is in many respects arbitrary but leaves no ambiguity about the status of these services. These two ‘lines of defense’ are both applied – sometimes quite explicitly – in expert deliberation during the revision process leading to the 1993 SNA.

The first line of defense is the market criterion. The market criterion is equivalent to what is often called the ‘third party criterion’. The third party criterion is derived from Margaret Reid’s definition of household production as consisting of unpaid activities that could conceivably be delegated to a paid worker or replaced by market goods (Reid 1934, p. 11). Along similar lines, Benham (1953, p. 173) reasoned that if we “...can find another economy, with markets, where consumption patterns are very similar, why not price the goods and services at the prices ruling in the latter?”. This position had become a professional consensus among economic statisticians by the early 1970s (Sakuma 2013, p. 5F56). To the third party criterion, Wood (1997) adds an additional “first world criterion”. Wood argues that a nonmarket activity is only considered productive if it is bought and sold in developed market economies. The market criterion in this analysis comprises both the third party and first world criteria.⁶

The market criterion is evident in the report of the 1981 Statistical Commission. The report acknowledges that “[t]he distinction between what is considered to be subsistence output and what is not is essentially an arbitrary one. It reflects mainly the traditional limits of marketed output in developed countries” (UNSC 1981, p. 14). The market criterion is also implied in the SNA’s definition of production:

5 From ISWGNA Expert Group meeting on the Household Sector, September 1987, Florence, Italy. The document *The Background to the 1993 Revision of the System of National Accounts*, edited by Anne Harrison (2005), is an annotated collection of all reports from the 13 Expert Group meetings of the ISWGNA between 1986 and 1983.

6 While the third party criterion is accepted among statisticians, the third world criterion is a critique made by Wood (1997) and not acknowledged by statisticians. The third world criterion is useful in this context because it highlights the persistence of neo-colonial modernization theory in development policy.

All goods and services produced as outputs must be such that they can be sold on markets or at least be capable of being provided by one unit to another, with or without charge. The SNA includes within the production boundary all production actually destined for the market, whether for sale or barter. (ISWGNA 1993, p. 5; ISWGNA 2008, p. 6)

The second justification is the distinction between household goods and household services. The production boundary of the 1968 SNA included some primary products for own consumption, such as the goods processed from agricultural or mining products, but excluded services (except for housing repairs by owner-occupiers) (Chadeau 1992, p. 87). In the 1993 SNA, it expanded to include *all* goods produced by households for their own consumption but continued to exclude services, “except for housing services produced by owner-occupiers of dwellings, and storage which is considered as an extension of the goods production process” (ibid.).

In a 1987 Expert Group meeting, the ISWGNA discussed several possible changes to the production boundary. These include the issues of how to value subsistence agricultural goods, how to classify repairs to buildings, how to treat water collection, and the activities of midwives and funerals (Harrison 2005, pp. 150–1). The discussion resulted in a few changes to the production boundary. Two of these in particular – water collection and midwives and funerals (discussed as a single topic) – illustrate the goods-services distinction. Water collection was moved inside the production boundary based on the argument that it “should be treated as the production of a good (that is making the water available where it is needed)” (ibid.). Regarding midwives and funerals, the expert group decided that, as services, neither should be moved within the production boundary.

These choices were based on convention rather than strict criteria. “In general it was not felt possible to have a single succinct definition of the production boundary that would explain why some items were included and some excluded ...” (Harrison 2005, p. 148). To get around this ambiguity, the ISWGNA decided “to give fairly general indications followed by specific lists of examples that would make clear where the boundary should be drawn” (ibid.). Such a list appears in the SNA (ISWGNA 1993, p. 149):

1. The cleaning, decoration and maintenance of the dwelling occupied by the household including small repairs of a kind usually carried out by tenants as well as owners;
2. The cleaning, servicing and repair of household durables or other goods, including vehicles used for household purposes;
3. The preparation and serving of meals;
4. The care, training and instruction of children;
5. The care of sick, infirm or old people;
6. The transportation of members of the household or their goods

These activities were explicitly excluded, and remain so in the most recent version of the SNA. In making these choices, statisticians relied on historical precedent, noting that “the only extensions to the production boundary previously accepted are for the production of goods” (Harrison 2005, p. 150). Services provided within the household, on the other hand, “are always immediately consumed by those producing them and therefore do not add to the pool of goods and services available for redistribution” (ibid.).

As several scholars have pointed out (e.g. Waring 2003; Wood 1997), neither of these lines of reasoning – the market criterion nor the goods-services distinction – are consistently applied. For instance, washing clothes or taking care of children can be (and frequently are) done by paid domestic workers and day care centers (Wood 1997, p. 51). Likewise, there is nothing inherent in services that makes them any less productive than goods. It was not until the 1993 SNA that services were given a strict statistical definition (Broussolle 2015, p. 574; ISWGNA 1993, p. 148). Services were a major topic in the 2008 SNA, given the growing importance of, among others, digital services, financial services and intellectual property. The 2008 SNA introduced several clarifications to the definition of services, which had the paradoxical effect of further blurring the goods-services distinction (Broussolle 2015).⁷ The distinction between non-market goods and services is especially arbitrary in the context of a subsistence household (Waring 2003, p. 36).

Summary

The first part of this section showed that not all statisticians and users of national accounts data agree on the exclusion of unpaid household services from the SNA production boundary. The arguments given by national accountants with direct involvement in standard-setting are grounded in practical concerns such as data collection challenges, international comparability, and policy applications. Yet, the decades-long history of contestation surrounding the issue makes clear that even a practical and technocratic framing is deeply political. And, because nearly all of the excluded activities are performed by women, the technocratic framing covers up deeply gendered ideas regarding what constitutes productive work and what does not. When the exclusion was clarified in the 1993 SNA, it was justified with reference to the choices that had been made in the previous SNA manuals starting in 1953. This ideational path-dependency is only possible because of the hands-off global governance of the SNA.

⁷ Among the changes to the definitions of services in the 2008 SNA is the distinction between change-effecting services and margin-services (Broussolle 2015, pp. 575–6; ISWGNA 2008, pp. 96–7). The 2008 SNA (ISWGNA 2008, pp. 96–7) lists several examples of change-effecting services, including: transportation, cleaning, repairs, healthcare, providing accommodation, improving one’s appearance, education, entertainment, and providing advice and information. Notably, all of these activities are excluded from GDP when no money is exchanged, but are explicitly listed as examples of productive services.

Concluding Remarks

Throughout its history, from the UNSO to the ISWGNA, GDP has been the domain of a small group of economic statisticians. The technical nature of international statistical standards is a source of power for statisticians employed by IOs and those with experience in international bureaucracies originating in the Global North – even if these statisticians do not perceive their work as political. This form of governance is largely insulated both from interests of outside actors and from competing ideas, particularly ideas from the field of feminist economics. This leaves decisions about GDP methodology in the hands of experts with shared norms (about the quality of official statistics) and ideas (about the boundaries of markets).

Although the treatment of unpaid household services in the SNA has been unchanged until now, this does not mean that change is impossible in the future. As indicated in the interview passages and commentary above, statisticians are aware of the criticisms and tend to be sympathetic. Increased public debate about GDP could reduce the insulation of experts from competing ideas. Moreover, innovations in data collection, such as the use of big data, could offer solutions to some of the practical barriers. The current SNA revision began with the 50th session of the UN Statistical Commission in March 2019. The key issues, which make up the mandate for the next manual, are globalization, digitalization, well-being, and sustainability (ISWGNA 2018a, p. 3). Whether or not unpaid household services is deemed to fall under the category of well-being remains to be seen. Considering that the topic was not on the agenda of the previous revision, it is not likely to return in the near future. In the longer term, however, there are no immovable barriers to major changes. It is also conceivable that GDP will become less influential as alternative indicators receive more attention (Fioramonti 2017).

From GDP figures to the Sustainable Development Goals and corruption indices, numbers and rankings shape global politics in important ways. The origins of these numbers and the governance of the international statistical system have largely been neglected. The example of unpaid services in GDP shows that statisticians, due to their expertise, possess a great deal of agency over global standards for economic measurement. While this agency allows for institutional change, continuity often prevails. Professional norms and shared ideas remain important drivers of stability in the way economies are quantified. As Ward (2004a, p. 300) observed, “The adoption of the SNA assumes there is a standard underlying economic model that serves all countries equally”. This assumption has effectively marginalized large amounts of women’s work, as feminist scholars have shown. GDP has far outgrown its role as an indicator of physical output in mid-twentieth century North America and Western Europe. As such, it is often expected to tell us a great deal more about social and economic progress and performance than it is capable of doing. Yet, as long as this figure remains so important in public life, its biases and shortcomings will lead to distortions in the way we see the economy.

4

The statistical trilemma: built-in limitations of international economic statistics⁸

Co-authored with Daniel Mügge

Abstract

Economic statistics are central to global economic governance. They are the informational background to the Sustainable Development Goals, conditional lending by international organizations, and other dimensions of development policy. But there is a growing chasm between aspirations for economic statistics and what they can deliver on the ground. We argue that many such shortcomings are rooted in what we call a trilemma of official statistics, a general limitation that goes beyond methodological deficiencies of individual indicators. Data users demand that economic statistics should (a) use harmonized standards to be comparable, (b) be guided by standards prescriptive enough to guarantee reliability and prevent manipulation, and (c) be suited to local socioeconomic contexts. Yet as we show, statistics can only satisfy two of these conditions at once. Importantly, we can only increase the suitability of statistics to local contexts if we make concessions on either prescriptiveness or harmonization. We examine three cases in detail: national accounts statistics, poverty lines, and unemployment statistics. To strengthen external validity, we also briefly consider inflation, trade, and debt statistics as additional cases. The statistical trilemma clarifies the inevitable trade-offs statisticians face when designing economic measurement standards.

⁸ This chapter is based on an article that is currently under review: DeRock, D. & Mügge, D. (2021). The statistical trilemma: built-in limitations of international economic statistics. Under review. Both authors contributed equally.

Introduction

International organizations (IOs) such as the World Bank, IMF, and OECD have put economic data central to their contemporary approach to global economic governance (Best 2017; Broome, Homolar and Kranke 2018). It is the informational background to the Sustainable Development Goals (SDGs) and conditional lending, as well as to economic policy and socio-economic research more generally. Such data, so the idea, allows us to track national economies over time and to compare them with each other.

This aim has inspired the norms underpinning international economic statistics (cf. Mügge and Linsi 2020). Three goals are particularly important for our analysis: First, countries should subscribe to a shared global statistical standard to make their data comparable, and hence use a harmonized set of standards (*harmonization*). Second, statistical standards should be prescriptive enough to avoid major ambiguities or loopholes, which might invite opportunistic data-tweaking (*prescriptiveness*). Third, statistical standards should be attuned to the socio-economic conditions of a country in question, lest they offer a distorted image of it (*suitability*). All three are meant to promote international economic statistics as useful and authoritative information sources, and the international statistical community has invested much energy and resources to support them.

Yet as we show, far from being complementary, fundamental tensions exist between these desiderata of official statistics. In stylized terms, to achieve any two, the third must be sacrificed—a dynamic we label the statistical trilemma. Combining harmonization and prescriptiveness generates potentially *ill-fitting statistical straight-jackets*, which are out of sync with, for example, local labor market or ownership structures. Misleading data and skewed comparisons follow. To improve the fit, harmonized standards can offer countries leeway to choose between alternative measurement approaches. The result, however, is *mock-comparability*, particularly when governments exploit such leeway to present themselves in a positive light. Suitability to local socio-economic contexts and prescriptiveness can then only be safeguarded if we accept limits to harmonization and hence a modicum of *statistical fragmentation*.

To illustrate the statistical trilemma at work, we discuss three prominent domains of statistics in detail: the gross domestic product (GDP) as the central metric in the System of National Accounts (SNA), poverty indicators, and unemployment statistics. For GDP, we find that standard-setters have prioritized harmonization and that many of GDP's commonly diagnosed defects are rooted in the tension between suitability and prescriptiveness. The situation for unemployment statistics is roughly similar. Yet because differences in local circumstances are more obvious, the lack of data comparability is more conspicuous than for GDP. Poverty statistics, in contrast, are fragmented, because home-grown yardsticks dominate measurement. An International Poverty Line, meant to

promote international comparison, offers only an unsatisfactory attempt to circumvent statistical fragmentation. To evidence the relevance of our trilemma-framework beyond these three cases, we also briefly discuss inflation, trade and debt statistics as shadow cases, in which the trilemma logic surfaces equally. For all of these, we draw on international statistical guidelines and compilation manuals, in combination with secondary data. From the latter, we place existing findings within the framework of our theory.

The argument has broader implications for scholars and policymakers. For the growing literature on the politics of statistics and other numbers in global governance, we provide a framework for understanding the constraints on what statistics can achieve and on the actors that govern them. It complements agent-centric approaches by showing the structural constraints international statistical standard setting confronts. By implication, there are hard limits to the quantification of economic governance around the globe.

Global statistical ambitions and their limits

Statistics carry authoritative knowledge about social and economic life (Broome, Homolar and Kranke 2018). Policymakers, citizens, and investors draw on them to guide their decisions. Choices based on inaccurate information generate poor outcomes (Devajaran 2013). Data biases can systematically disadvantage people, for example along racial or gendered lines (e.g. Benjamin 2019; Waring 1999), when whole sectors are unduly valorized (Mazzucato 2018), or when entire countries' economic contributions are unfairly distorted in the data (Smith 2012).

Nevertheless, the ambition to “quantify the world” and create a border-spanning statistical view of social life has flourished for roughly a century now (Ward 2004b). The labor movement had championed systematic unemployment statistics already in the late 19th century (Desrosières 1998). Even if they were not formally harmonized, international discussions about such measures in the newly-founded International Labour Organization in the 1920s reflected the breadth of this concern (Baxandall 2004). National Income measures, the predecessors to present-day GDP, had initially been developed for isolated national contexts (Studenski 1958). Yet after the Second World War, leading statisticians proselytized for them around the world (Giovannini 2008). Four factors came together. First, colonial metropolises had early on kept economic records of their transactions with dependencies, not least to monitor surplus extraction and its interaction for example with international payment imbalances (cf. Appadurai 1996; Mitchell 2002; Speich 2008). Second, after decolonization, newly independent countries frequently empowered technocratic elites trained at leading British or American universities (cf. Masood 2016 for the case of Pakistan). These technocrats diffused a predilection for fostering economic growth and quantitative macroeconomic management outside the West. Third, UN policy elites themselves championed and promoted a more

quantified, evidenced-based approach to economic development, and advocated it around the world (Ward 2004b). Finally, international organizations, in particular the IMF, were tasked to monitor financial flows, creating a need for frequent data provision from IMF member states (Reichmann 2016). Together, these factors universalized economic statistics as a tool of statecraft and turned the availability of comprehensive and comparable economic statistics into a common-sensical, hardly questioned ambition.

Since then, the role of economic statistics in international affairs has increased further. Initial development thought prioritized economic growth and hence GDP. While GDP has remained central, indicators for living standards and poverty levels gained prominence since the 1970s (Khan et al. 2015). The Human Development Index (HDI) broadened the notion of development to include health and education alongside income. Since then, the dozens of indicators in the Millennium Development Goals (MGDs) and Sustainable Development Goals (SDGs) have heralded a data-saturated “World That Counts”, in the words of the UN-sponsored Data Revolution Group (2014).

As statistics have become more central to development policy, the users of statistics have become more diverse and the demands on them greater. We see this, for example, in the contentious deliberations leading to the selection of the SDG indicators (Kapto 2019). Data users of all stripes have high expectations for what statistics should deliver.

For our purposes, three expectations are particularly important: (1) economic statistics should capture socio-economic realities on the ground, (2) they should be compiled according to unambiguous procedures (so that we actually know what is in them), and (3) they should be comparable across countries. We translate these ambitions into three desiderata of statistical standards: *suitability*, *prescriptiveness*, and *harmonization*.

Let us consider each of these briefly in turn. Suitability means that both the concepts underlying a measure and the actual measurement procedure should be relevant to a country’s social-economic conditions. Traditional unemployment measures, for example, may carry little information in a country where informal employment dominates. Classic debt measures can be seriously misleading when a government’s liabilities are tucked away in derivatives deals, or when one country issues debt in its own currency while another has to borrow US Dollars. Consumer price indices may miss rising costs of living when families’ disposable incomes are primarily squeezed through climbing real estate prices. How important such suitability gaps are differs across countries, depending on the structure of their economy, the strength of the welfare state, gendered and racial fault lines in society, factors such as a potential reliance on migrant labor, and so on.

The same applies to actual measurement approaches. Household surveys may offer little information when people move around frequently or their shelter has no unambiguous

address. Official payroll data will systematically miss informal employment, just as growth figures ignoring the digital economy miss an important part of wealth creation (Bean 2016). Price developments in supermarkets are a poor guide to changing living costs where people grow most of their food on their land.

In consequence, it is essential that statistical standards are suitable to the country they try to reflect. That standard is a pragmatic one: it is inevitable that “complexity and contextual detail is ‘lost in translation’ in the pursuit of quantification and comparability” (Broome and Quirk 2015, p. 827). Suitability means that there are no glaring mismatches between a country and a measurement standard and, easier to spot, that the mismatch does not vary enormously across countries.

Prescriptiveness means that measurement standards should be as unambiguous and transparent as possible. Clear guidelines allow government statisticians to adhere to internationally accepted practices faithfully and buttress public confidence in official statistics. International standards such as the SNA, the Balance of Payments Manual, or the Government Finance Statistics Manual are highly detailed publications. The most recent SNA, for instance, spans more than 600 pages and is supported by over 40 handbooks and compilation guides. Prescriptiveness is meant to avoid random noise in or deliberate manipulation of data, particularly when getting the numbers wrong might obscure critical problems such as chronic payment imbalances. Therefore, the IMF permits little flexibility in the production of financial and macroeconomic statistics, and monitors compliance with standards through Article IV consultations (De Las Casas 2016).

Harmonization, finally, requires that countries subscribe to shared standards such as the SNA, and the concepts, classifications and definitions specified in them (UNSD 2013). The demand for comparable data took off after the Second World War, with increased foreign trade and growing foreign exchange markets (Giovannini 2008, p. 14). Systematic comparisons of country “performance” have become common in the form of benchmarking (Fougner 2008; Broome and Quirk 2015). As economic globalization has progressed and policy has become increasingly quantified, the demand has grown further.

Yet, harmonization is not necessarily rigid. Shared standards often allow flexibility to encourage international take-up despite diverse data sources or compilation methods and varying degrees of quality. Desrosières (2000, p. 175) called this the harmonization of “products” as opposed to harmonization of “methods”. That is, the outputs can be harmonized even if the inputs are not. Ambiguity of this sort is common in international agreements, frequently by design (Best 2012).

We contend that there are substantial trade-offs between suitability, prescriptiveness, and harmonization. When two are maximized, the third will suffer. Statistics that are both

prescriptive and widely suitable cannot be globally harmonized, because prescriptive measurement guidelines will be tailored to domestic, local, or regional contexts. We label the resulting condition *statistical fragmentation* (see Figure 1 below). Statistics that are both prescriptive and harmonized cannot be suitable to all the geographic and demographic contexts they are supposed to measure. Such statistics sacrifice local relevance for comparability, resulting in *ill-fitting statistical straightjackets*. Lastly, statistics that are both harmonized and widely suitable cannot be highly prescriptive. They must be flexible enough to allow countries to adapt concepts to socioeconomic differences, to pick or choose from multiple methodologies, ignore parts of frameworks irrelevant to domestic structures or analytical priorities, and so on. The resulting *mock comparability* means that formal compliance with international standards hides a lack of comparability.

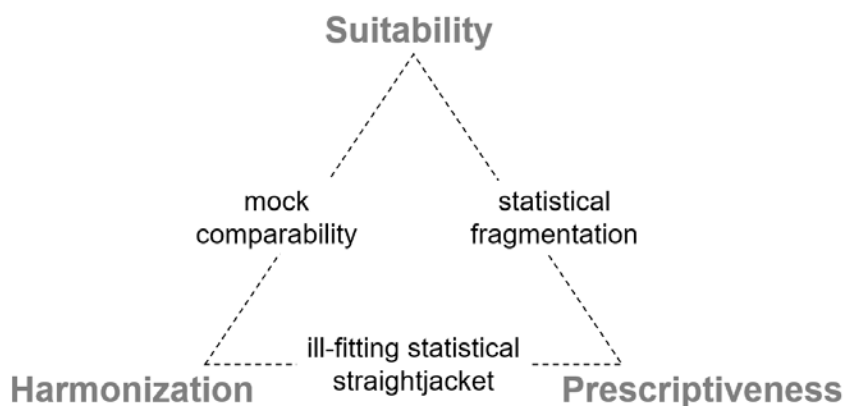


Figure 1: The statistical trilemma and its outcomes (figure created by authors)

Empirical case studies

Our case studies apply the statistical trilemma framework to prominent indicators in macroeconomic and development policy and analysis. For each, we show how the indicator is limited by the statistical trilemma and how standard-setters have responded to these limitations. GDP is globally harmonized and is characterized by trade-offs between suitability and prescriptiveness. While neither of these ambitions is fully sacrificed, GDP can be neither as universally suitable nor as prescriptive as its users demand. The case of unemployment statistics is one of mock comparability, where a general ILO definition covers up a lack of prescriptiveness and also of suitability. Poverty statistics are fragmented, a result of countries prioritizing suitability and prescriptiveness over harmonization. While countries share certain best practices, each country sets its own guidelines and the resulting figures lack comparability. Statistics of trade, debt, and FDI constitute additional examples of ill-fitting statistical straightjackets and mock comparability.

GDP and National Accounts

GDP is one of the world's most widely-used economic indicators, and also the most criticized. Its success stems partly from its promise to capture the total economic output of a given country and for the resulting figures to be comparable worldwide (Speich 2011; Lequiller and Blades 2014, p. 15). With GDP's simplicity and its centrality in policymaking, expectations are high for what it should deliver.

Of the three possible outcomes described by the trilemma, standard-setters are least willing to countenance statistical fragmentation. Users – especially those using GDP for analytical (rather than administrative) purposes – expect GDP to be harmonized. The United Nation Statistical Commission and its members have pursued harmonization since the 1950s (Ward 2004b). Most recently, it has been coordinated through The Implementation Programme for the System of National Accounts 2008. Even when GDP shared the stage with Communist countries' Net Material Product (NMP), the UN led efforts to convert NMP to GDP, and the World Bank and IMF nudged NMP-countries toward adoption of the SNA (Herrera 2010; van Heijster and DeRock 2020).

With fragmentation out of the question, the biggest challenges for standard-setters and the most salient objects of criticism have been shaped by tensions between suitability and prescriptiveness. Prescriptiveness is particularly urgent for users concerned about manipulation or misrepresentation of GDP figures, whether intentional or not (cf. Aragão and Linsi 2020). Reflecting these worries, IOs promote compliance with the SNA and monitor data quality through programs like the IMF's Data Quality Assessment Framework and the UN's National Accounts Questionnaire.

Debates about GDP's suitability are often framed in terms of developed versus developing countries. This is an oversimplification. Accounting for the digital economy, for example, is not strictly a developed country issue, considering the importance of mobile banking in low- and middle-income countries. Nonetheless, GDP's largest suitability deficits have indeed been biased against poorer countries.

The second official publication of the SNA in 1968 forced statisticians to confront the poor fit of the standard to dozens of new UN member states following independence. Ward (2004b, pp. 302-306) summarizes two main shortcomings. First, the SNA was “internally” focused, whereas the main problems facing developing countries were “external”, such as unequal trade relations and the operations of foreign firms within their borders (ibid.). This problem persists. For example, as Smith (2012) argues, GDP undervalues the contributions of low-wage workers in global value chains. Second, and also still relevant today, developing countries' economies were heavily concentrated in nonmarket activities that can only be valued indirectly through imputations (Ward 2004b, pp. 302-306).

As of the 2008 SNA, many of the suitability deficits biased against developing countries have been amended. Although subsistence farming was already within the production boundary of the 1953 SNA (under the category of primary production), other forms of subsistence production, such as water collection, were added in 1993 (Harrison 2005). Also noteworthy is the adoption of the ILO definition of informal sector enterprises in the 1993 SNA (ILO 2013). Most of these changes have entailed either imputations for non-market activities or more detailed guidelines for estimating hard-to-measure activities, as with the informal sector.

For the most part, however, changes to the production boundary in the 1993 SNA and 2008 SNA have increased suitability primarily for rich, highly industrialized countries. Changes to the treatment of financial intermediation services, research and development (R&D), and weapons production have disproportionately increased the GDP of countries that have seen structural shifts toward these industries (Assa and Kvangraven 2021; Bos 2009, p. 40). Financial intermediation services did not contribute to GDP until the 1993 SNA. Since then, their prominence has grown, however, with a 2008 SNA in which “even banks’ own money could now be used to create ... ‘production’, without the pretext of providing an intermediation service” (Assa and Kvangraven 2021, p. 6; see also Christophers 2011).

Military expenditures, too, have shifted further inside the SNA production boundary. The 2008 SNA redefined all government weapons spending as government investment, which is substantially higher for high- and middle-income countries compared to low-income countries (Assa and Kvangraven 2021, p. 7). R&D, too, is concentrated in developed countries. R&D began contributing to GDP only in 2008, whereas before it was considered intermediate consumption rather than a productive activity.

Such expansions of the production boundary are plausible efforts to keep GDP up to date with technological change and to increase suitability to countries that have shifted away from manufacturing. But these changes are puzzling in relation to GDP’s continued exclusion of unpaid household services. The SNA production boundary excludes own-account household services – specifically, “activities undertaken by households that produce services for their own use” – with the exception of services of owner-occupied dwellings such as home repairs (World Bank 2012, p. 59). This exclusion does not necessarily affect developing countries more than developed ones, but it is highly gendered within countries. Worldwide, women spend an average of three times longer than men on unpaid care work, including cooking, cleaning, childcare, and caring for the sick and elderly (ILO 2018; Miranda 2011; Wood 1997).

According to the SNA, this limitation is necessary because, if a product cannot be exchanged, “there can be no division of labour, no specialization of production and no gains from trading” (World Bank 2012, p. 57). When pushed, however, statisticians concede the arbitrariness of this exclusion (see DeRock 2021, p. 20). Services produced and consumed inside households remain outside the production boundary because it is nigh impossible to quantify their market value in a convincing way. Trying to do so would thus open the door to arbitrariness and undermine the comparability of data across countries. Rather than measuring something important imprecisely, it is not measured at all (cf. Mügge and Linsi, 2020).

In essence, this move avoids statistical fragmentation at the expense of suitability. The 2008 SNA does give countries the choice not to quantify R&D if it is a small or nonexistent sector, or if capacity constraints make such data collection impossible. But to open up GDP to services such as unpaid care work would – at least, according to standard-setters – have threatened the comparability of GDP at its core.

Standard-setters also emphasize the “flexibility” of the SNA (ISWGNA 2008, p. 37). The SNA allows countries to adapt the guidelines, to “give greater emphasis ... to one part rather than another” (ISWNGA 2008, p. 37). Flexibility accommodates differences in countries’ policy applications of the data as well as statistical capacity and data availability (ibid.). To be sure, statisticians recognize that, given the wide variation between national economies, certain details of the SNA may have to be locally adapted or even ignored (Harrison 2017; Vanoli 2005, p. 127). For example, some countries, such as Australia and others, measure pension contributions in ways that diverge from the international guidelines (Harrison 2017, p. S214).

The flexibility of the SNA, however, is moderated by demands for prescriptiveness. The SNA does not allow deviation from core concepts such as the production boundary, financial assets, or capital formation – concepts that are built upon the experiences of industrialized countries (Bos 1995). Today, “the basic concepts and definitions of the SNA depend upon economic reasoning and principles which should be universally valid and invariant to the particular economic circumstances in which they are applied” (ISWGNA 2008, p. 1). In sum, GDP is not quite an ill-fitting straightjacket or only superficially comparable. But it is neither as locally suitable, nor as prescriptive, as many users would want (see Figure 2 below). The tension between the three poles of the statistical trilemma means that trying to mend defects in one dimension will always entail concessions in the others.

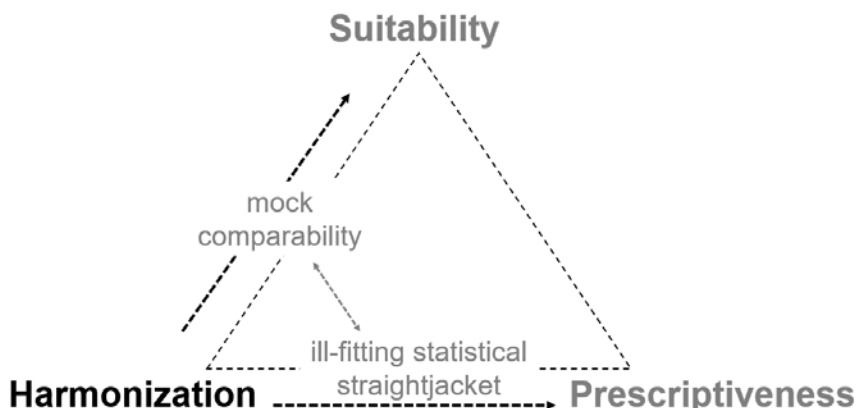


Figure 2: Trade-offs in the harmonization of GDP (arrows indicate unresolved tensions; figure created by authors)

Unemployment

Unemployment indicators aim to capture a crucial dimension of socio-economic life – to what degree large groups of people are structurally unable to find paid employment. While such difficulties are widespread, the specific metrics to quantify them have their roots in what are today heavily industrialized and relatively rich countries in the Global North (Salais, Baverez and Reynaud 1986; Baxandall 2004; Zimmermann 2006). In Baxandall's words, unemployment statistics had originally aimed to capture

[an] able-bodied, prime-age male industrial breadwinner with plant specific skills who [had] been laid off from full-time formal work as the result of a plant closing in a declining industry. (Baxandall 2004, p. 212)

Even though it is not obvious how readily this concept travels to other parts of the world (Topalov 1994), the 1982 ILO standards build on this image when they identify unemployed people through three core criteria: they do not have paid employment, they are available for it, and they are actively looking for it.

These three criteria make intuitive sense. The first is self-evident. The second means that a person must be able to take up a job more or less directly to qualify as unemployed. If she had other structural obligations, for example by being enrolled in an educational program, she might still want a job (eventually), but would not count as unemployed. The third criterion acknowledges that people may be voluntarily jobless. After all, unemployment is only meant to capture a lack of jobs for those people who actually want one.

Once we try to operationalize these criteria and make unemployment measurable, things become much more difficult, however. What is the lower threshold for employment? ILO guidelines suggest that one hour of paid work per week is enough not to count as

unemployed – a remarkably low bar. Also, “having a job” is readily equated with salaried wage labor. But people can earn money with their work in ways that fall below the radar of authorities. Such informal labor for example includes unregistered self-employment or small enterprises, as well as unregistered day-laboring (Schneider and Enste 2002).

The second criterion is equally tough to operationalize in a comparable manner. National authorities typically build employment statistics from survey data, meaning that respondents have to report on their own availability. Some countries offer cheap and easy access to re-training or other educational programs for people without work, such that de facto unemployed people would seem to be unavailable. At the other extreme, people in poor regions in particular may be heavily tied up in subsistence agriculture or care for other family members, without potential recourse to public services. In those situations, people might experience a chronic shortage of jobs, but fail the availability criterion.

This problem ties in directly to the third requirement – the need to look for a job actively. This criterion is the most thorny of all. Operationalizing it requires specifying which activities count as active search and then measuring it. Different proxies are conceivable: actual job applications, visits to job centers, review of job adverts online or in newspapers, and so on. But such criteria are based on an image of highly formalized labor markets, potentially with public institutions in place to aid job search. Job centers or fairs of the kind that may be common in rich, industrialized countries are often absent in poor ones.

Yet structural differences between national labor markets mean that they are hard to compare using one and the same yardstick. For example, the extent of informal work varies widely around the world: recent ILO data see it ranging from more than 90 percent in many sub-Saharan African countries Africa to 10 in many Western European ones (International Labour Office 2018, p. 85ff). Roughly 60 percent of employment worldwide is classified as informal. Globally, the experience of people in rich countries with highly structured labor markets is the exception, not the rule.

To be sure, statistical agencies can devise ways to measure it still, for example through large-scale surveys (cf. Hussmanns 2004). These efforts are costly, however, and Labour Force Surveys are commonly customized on a national basis to accommodate inter-country differences (e.g. Rivera 2008). They do not generate internationally comparable figures, and by that token, they are unable to mend the lack of comparability of unemployment statistics.

More importantly still, however, relatively disconnected areas may lack the infrastructure to allow people even to apply for formal employment, for example because such employment opportunities are far removed from where they live, or because a life in poverty leaves little time and resources to invest into active job search. A uniform criterion can therefore

fail to do justice to the diverse conditions under which people can or cannot seek paid employment if they want it.

Consider the case of Philippine employment statistics (Rivera 2008). In 2005, the Philippines adapted its labor force survey to track the three standard criteria for unemployment: having no job, being available for one, and actively seeking one. In contrast to international best practice, however, it relaxed the last criterion to increase the local suitability of the definition. Henceforth, Philippines authorities would still count someone as unemployed if they had not been actively job-hunting, provided they reported being too tired, believed no work was available, awaited results from previous applications, expected to be rehired or recalled to their former (informal) work soon, or could not search for jobs because of bad weather. On the other hand, people reporting not to be looking for work because of family duties were not counted as unemployed but simply outside the labor force. The appropriate way to define and measure employment conditions is locally specific.

Reflecting such diversity, ILO statistical standards are clearly less prescriptive than in the case of the SNA. Many countries report unemployment rates that claim to be based on the ILO *definition*. Yet countries differ significantly in the measurement tools that they use—business surveys, household surveys, and so on. As illustrated with the Philippine example above, they also interpret abstract concepts (“availability”, “active search”, etc.) in different ways. The result is *mock-comparability*, certainly across continents or between countries with highly different socio-economic structures or wealth levels (see Figure 3 below).

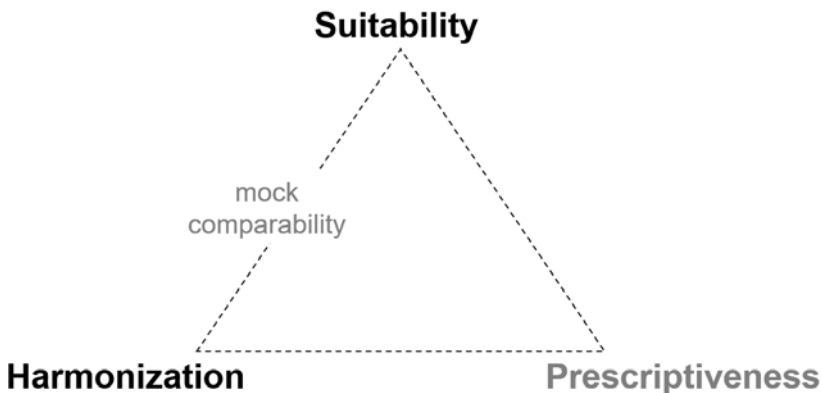


Figure 3: Mock comparability of the ILO definition of unemployment (figure created by authors)

Poverty statistics

We look to poverty statistics to find out whether, on the whole, livelihoods in a given country are improving or worsening. We also expect the data to tell us if one country

or region has a higher rate of poverty than another, and whether the world as a whole is achieving poverty reduction. These are important but deceptively simple questions to ask (Hickel 2016).

Users of poverty statistics have to grapple with the co-existence of national poverty lines and the World Bank's International Poverty Line (IPL). While the IPL has enabled cross-country comparison, it has also obscured national differences. The IPL is not an international standard. Poverty statistics are fragmented, and the IPL is a centralized attempt by the World Bank to convert incomparable statistics into comparable ones. The IPL is instructive, however, because its travails illustrate why harmonization has not been achieved in the first place.

National poverty lines are set individually by countries and are typically based on household surveys. Poverty statistics then combine those with a welfare indicator. At the household level, that is typically measured in terms of current income or consumption expenditures, with income the more common indicator for OECD countries and consumption for developing countries (Ravallion 2010, p. 2). Once statistics on income poverty are obtained, the next step is establishing a poverty line, below which people are said to be poor (*ibid.*). The final step in measuring poverty is calculating the poverty rate – the number or percentage of people living in poverty based on the poverty line.

There is no fixed approach to choosing a national poverty line. Most developing countries follow a 'cost of basic needs' approach (Demombynes and Vu 2015, p. 2). It starts with a food poverty line, based on the cost of a minimum caloric intake for a certain level of health, to which a non-food component is then added (*ibid.*). Both are estimated from survey data for households clustered around the poverty line.

"Basic needs" obviously vary across countries, if only because people have been socialized into different expectations and because of the social stigma attached to be on the lower rungs of any given society. But there are also concrete differences: "In a poorer country, for example, participating in the job market may require only clothing and food, whereas someone in a richer society may also need access to the internet, transportation, and a cell phone" (World Bank 2018, p. 7).

In 1990, based on the observation that the national poverty lines of several of the poorest countries clustered around US\$1 per day, the World Bank adopted the so-called dollar-a-day line as the first IPL (Ravallion, Datt and van de Walle 1991). The IPL has been updated several times and was last raised in 2015 to US\$1.90 per person per day at 2011 purchasing power parities (Ferreira et al. 2016). The IPL is obtained by averaging the 15 lowest poverty lines out of a sample of 74 national poverty lines (Jolliffe and Beer 2016, p. 2).

The World Bank now also reports on two higher poverty lines of US\$3.20 and US\$5.50 per day, to reflect that “the majority of people and most of the world’s poor now live in middle-income countries” (World Bank 2018a, p. 7). In fact, this proliferation of international poverty lines implicitly appreciates that no single poverty line can do justice to all the divergent socio-economic circumstances. The higher ones of course give a less rosy picture compared to the level of ‘extreme poverty’ depicted by the IPL. Almost half the world’s population falls below the \$5.50 line, and a quarter falls below the \$3.20 line (ibid. 2018, p. 7).

The actual situation is one of statistical fragmentation. Jolliffe and Beer (2016) argue that national poverty lines are fundamentally incomparable. The IPL attempts to circumvent this problem by sidestepping the problem of local suitability (see Figure 4), but simply averaging out the lowest national poverty lines results in “a conceptually incoherent average value for the IPL” (ibid., p. 2). The result, consistent with the trilemma heuristic, is a severely ill-fitting global yardstick. As with the other case studies, the constraints imposed by the trilemma do not force the World Bank into making any *specific* choices. But no matter at which level an IPL would be set, it could hardly claim universal relevance.

The ambition for comparable poverty statistics is understandable and appeals to benchmark-minded IOs, NGOs, and their member countries and donors (Clegg 2010). But the IPL obscures more than it reveals, and it arguably underestimates the true extent of global poverty. At the same time, abandoning an IPL altogether would entail giving up on comparable poverty statistics, thus leaving the SDGs and other global development agendas without a fixed target. As with GDP, the simplicity of a single, comparable figure gives staying power to the IPL, even if a great deal of information is lost along the way.



Figure 4: Trade-offs in national poverty statistics and the International Poverty Line (figure created by authors)

Additional shadow cases

To highlight the breadth of the trilemma dynamic, this final section briefly outlines how it reverberates in three additional cases, namely the measurement of public debt, foreign direct investment, and in trade statistics.

Debt

Public debt figures are meant to reveal the fiscal burden governments have assumed. They are meaningful when they can tell us something about public authorities' ability to carry that burden—which is why we typically express debt in relation to GDP as a measure of the “size” of the economy.

While public debt as a concept sounds straightforward, it is riddled with measurement problems (Bloch and Fall 2015; de Vlieger and Mügge 2020). For our purposes, several challenges stand out: considering liabilities in isolation obscures that governments may have substantial assets that, in theory, could be sold to service debt. While less “hard” than regular debt, public pension liabilities can constitute debt-like fiscal burdens (OECD 2013). Debt structures can vary significantly, for example because liabilities may come due soon or only decades from now, and they may be owned by relatively loyal domestic creditors, or by much more fickle international investors. And crucially for global comparisons, public debt varies according to the currency in which it is issued, which may or may not be the domestic one.

The most authoritative set of public debt figures is compiled in the IMF's Government Finance Statistics (GFS) data base according to the IMF's manual (IMF 2014).⁹ Debt figures are consequential for governments because they feed into public borrowing costs (Cottarelli 2017). Within the range of measurement options, public authorities therefore have an incentive to choose that which portrays them in the most favorable light, rather than the one that might seem most appropriate (Aragão and Linsi 2020). Recognizing this, statistics manuals for debt tend to be relatively prescriptive.

Prescriptive, harmonized standards mean that local conditions remain unheeded in headline figures. The actual fiscal burden that any nominal amount of debt constitutes—the ultimate object of interest—can vary enormously depending on the specific conditions of

⁹ The discrepancies between the GFS data and the figures reported in the OECD's National Accounts at Glance Database (Bloch and Fall 2015) already reflect the kinds of measurement ambiguities just mentioned. Both are different again than the “Maastricht debt” data that Eurostat collects to monitor compliance with the EU's Stability and Growth Pact. Because Maastricht debt is only reported for the relatively homogeneous group of Eurozone member states, the trilemma dynamics are less acute there than for yardsticks applied across the world. That said, even in the EU, with its exceedingly tight measurement standards, governments exploit what little room for maneuver they have to their own advantage (Savage 2005).

any country. Countries in the Global South typically borrow from international investors, in foreign currency, and relatively short term. From a fiscal sustainability perspective, a debt burden of, say, 60 percent of GDP may be much more of a concern than the more than 200 percent of GDP that the Japanese government owes other domestic creditors. Standardized debt measures therefore lead us to compare apples and oranges. Without further contextual information, they create ill-fitting straightjackets – especially once we leave the OECD world and compare across the globe.

Foreign direct investment

FDI statistics exemplify how multiple options for data compilers can undermine international comparability. Again the underlying concept seems simple enough: FDI statistics should capture investments from people or companies from one country in another. The “direct” element signals that FDI is unmediated; through the investment, the investor becomes actively involved in the business in question. This criterion separates FDI from portfolio investment (PFI), for example through the acquisition of a small number of shares.

In practice, FDI is tracked through financial flows. Any particular flow, however, does not advertise who the ultimate originator or target is, and whether it is meant as a passive or active investment (Damgaard and Elkjaer 2014; Kerner 2014). Source and destination of a flow can be hard to track because investors may have structural incentives to deceive authorities on these counts (Damgaard and Elkjaer 2017). Investments are frequently channeled through multiple countries and holding companies to arbitrage tax rules (Finér and Ylönen 2017). The uneven distribution of financial conduits around the world systematically distorts our image of which countries invest, and where (Fichtner 2017). In consequence, IMF researchers have lamented that up to 40 percent of globally registered FDI may in fact be phantom investments – FDI that exists on paper but in fact is something else, for example domestic investment through offshore vehicles (Damgaard, Elkjaer and Johannesen 2019).

Equally difficult to pin down is the difference between FDI and PFI. Measurement manuals before the 1990s had acknowledged the importance of contextual information and emphasized the importance of national discretion (Linsi 2019, p. 53ff). Later BPM versions, in contrast, have included a highly prescriptive but equally arbitrary “10 percent rule”, according to which investments over that threshold would count as direct investment, and everything else would not.

Where new rules for measuring FDI *flows* substituted prescriptiveness for suitability, rules for FDI *stocks* have eschewed such uniform standards. Measuring the value of FDI as it has accumulated over the years in a country, such inward FDI stocks can be seen as an indication of the degree to which “foreigners” have acquired stakes in any

given economy. Measuring them requires adding up investments made over many years and specifying their present value. Different options exist, but all have clear drawbacks: historic acquisition prices are unlikely to reflect present values of investments, even with inflation adjustments. Company data may be skewed for reasons similar to other FDI data. Market values for corporate stakes are only available for companies that are publicly traded. The value of stakes in untraded companies is anybody's guess. And over long periods of time, foreign subsidiaries of a company may reinvest profits made abroad or funds borrowed there (Kerner 2014). Both would constitute FDI, but neither would be a cross-border financial flow.

Because the best measurement approach depends on case-specific circumstances, the current BPM offers no less than seven BPM-compliant valuation techniques for FDI stocks (Damgaard and Elkjaer 2014). Using the Danish example, Damgaard and Elkjaer show that estimates for inward FDI stock vary between 22 and 156 percent of Danish GDP, depending on the technique used.

FDI statistics are thus a good example of how adherence to international standards is bought at the expense of multiple measurement options that ultimately generate mock-comparability. Indeed, given the commercial and political stakes attached to measuring and reporting FDI flows and stocks, we can expect data reporters to be opportunistic in the figures they produce. Data is neither really comparable internationally, nor are standards prescriptive enough for us to know at least on a country by country basis what exactly is in the figures.

Trade

Trade data, too, is compiled in line with the BPM. But in contrast to FDI, the measurement of stocks is not a central concern, and tracking flows is riddled with relatively fewer problems. The classification of trade flows in the BPM is therefore relatively prescriptive.

Nevertheless, problems abound (UNECE, Eurostat and OECD 2011). Countries measuring one and the same bilateral flow commonly arrive at very different numbers (Linsi and Mügge 2019), meaning that a single set of measurement rules still leads to widely diverging figures. Some of these divergences reflect genuine uncertainty and measurement errors; others, however, stem from the distortions that national idiosyncrasies introduce. Examples include that trade hub countries typically report high import and export figures even though they only function as conduits. Countries in which components are assembled into final products—say, cars or consumer electronics—look as if all the value of their exports was locally produced (“German” cars or “Chinese” smartphones) even though they, too, partially pass on value added created elsewhere. National statistical authorities would need to make locally specific adjustments to get figures into line. Harmonized standards generate discordant numbers.

Things become more complicated once we move from merchandise trade to trade in services. Imagine a New York branch of a Dutch bank arranging a transaction in Japanese Yen between a Russian and an Italian party, both of whom use offshore vehicles to execute the transaction. Which country is exporting a service where, in this example? Any rule set trying to capture this transaction will be found wanting—squeezing such a deal into the conceptual corset of international trade does violence to its transnational character. In parallel to the FDI example above, whatever universal rule we establish may end up substituting technical precision for common sense. Importantly for our argument, whichever dimensions of the deal we would privilege to define the two trading countries, we would end up inflating trade figures for one kind of country at the expense of another. “Heavily trading countries” might be those with the actual headquarters, or transacting branches, the currency in question, or the legal but fictitious offshore incorporation of either service provider or client. Prescriptive standards clash with the diverse trading profiles of countries around the world.

Conclusion

Statistics have become increasingly important in international economic governance in the pursuit of transparency, evidence-based policy, and a clear sense of economic developments in the global economy. At the same time, limitations and biases of these statistics have come to light in recent scholarship. As a result, economic statisticians are now recognized as agents in global politics rather than simply as technocratic experts (Herrera 2010; Schmelzer 2016). The methodologies they design—either domestically in national statistical offices, or internationally in the creation of standards—shape how we see and interpret the world.

We have argued there are inherent limits to what economic statistics can achieve—what level of detail they can provide, how many stakeholder demands they can satisfy, and so on. International economic statistics cannot simultaneously be prescriptive, harmonized, and locally suitable. This argument builds upon critiques of economic statistics, such as the reductionism of poverty lines and unemployment figures or the biases and blind spots of GDP. Without question, these political features of economic measurement have real and potentially harmful impacts as they shape how policymakers see the world and act upon it. However, as the empirical examples above aim to demonstrate, underlying structural factors limit the degree to which these problems can actually be tackled. The deficiencies of international economic statistics are therefore not the result of professional incompetence or indifference, but of endemic constraints global economic quantification confronts.

In making this argument, we do not claim that problematic economic statistics should get a free pass. What we contest, rather, is that there are obvious solutions available to

statistical standard-setters. Making economic statistics “better” is not simply a matter of tweaking existing methodologies, or even abandoning the ones we have and replacing them. There are built-in limitations to what international economic statistics can achieve. For users of economic statistics, this argument suggests more than just being cautious about official statistics. Instead, the statistical trilemma allows users to pinpoint which criterion is sacrificed in the inevitable trade-offs, or which criteria have been watered down due to tensions with others. Social scientists and policymakers in particular should be aware of these limitations.

Our argument speaks to IPE scholars in two regards. First, aside from economics, IPE is chief among disciplines that makes use of the kinds of data we discuss. Attention to built-in limitations can help improve the quality of quantitative and mixed methods research in a way that is not captured by traditional criteria of reliability and validity. It can also help qualitative researchers who use descriptive statistics to place them in a more accurate social context. Second, the theory provides a missing framework for the growing literature on the political economy of statistics and other kinds of numbers. It complements arguments about the shortcomings of statistics, and the consequences thereof, by offering a structural explanation for the emergence of these shortcomings. The pessimistic conclusion is that some shortcomings, on at least one corner of the trilemma, are inevitable. Future research might build on this argument by asking why, within the built-in limitations, standard-setters have opted for the trade-offs that they have.

5

Bounded autonomy in the global governance of economic statistics¹⁰

Abstract

International organizations (IOs) such as the UN, the World Bank and the IMF are at the forefront of efforts to govern statistical practices worldwide. The degree to which these IOs (fail to) cooperate with each other impacts how economies are made legible to policymakers and investors and how IOs themselves see the world. The first aim of the article is to determine the extent of cooperation in three functions: statistical standard-setting, capacity building, and compliance monitoring. The second is to explain variation in the degree of cooperation. I find that cooperation is driven by two factors: expertise-based autonomy of IO staff, and the degree to which statistical governance is motivated by broader IO priorities (such as surveillance and lending). Cooperation is highest in functions requiring highly specialized knowledge, as long as such cooperation does not interfere with IOs' core priorities and mandates. I support this argument empirically with qualitative data from 27 semi-structured expert interviews combined with extensive document analysis. The article engages with literature about inter-organizational dynamics in global governance. The findings also have important implications for relations between IOs and the national statistical systems of member states, especially developing countries.

10 This chapter is based on an article that is currently under review: DeRock, D. (2021). Bounded autonomy in the global governance of economic statistics. Under review. The chapter also builds upon the following working paper: DeRock, D. (2020). Leaving no data behind? International organizations and the conflicted global governance of monitoring the Sustainable Development Goals. Fickle Formulas Working Paper 01-2020.

Introduction

International organizations (IOs) are at the forefront of efforts to harmonize and govern economic statistical practices around the world. These efforts matter in part because they provide the informational foundation for IOs and thus shape how they see the countries and regions in which they operate. They matter too because the quality and measurement of economic statistics can have distributional impacts, including misguided policies resulting from missing data (Devarajan 2013) and global inequalities perpetuated by statistics biased toward the Global North (Smith 2012). Global statistical governance is presumably geared toward mitigating such problems.

Three IOs in particular lead the way in global statistical governance: the United Nations Statistics Division (UNSD), the World Bank, and the International Monetary Fund (IMF). They set and revise international statistical standards such as the System of National Accounts (SNA) and the Balance of Payments Manual (BPM), provide capacity building and technical assistance, and promote member countries' compliance with standards (Alenda-Demoutiez and Mügge 2019; DeRock 2021; Mosley 2003; Samuel 2014). There is substantial functional overlap in their governance efforts. That is, these IOs specialize in similar types of economic statistics and focus their efforts largely on the same countries and regions. Yet we know little about the degree to which IOs cooperate in statistical governance, what drives this cooperation, or the implications. This article aims to address that gap, asking two related questions. First, to what extent do IOs cooperate in global statistical governance? And second, what explains variation in the degree of cooperation across different governance functions?

Existing literature on the political economy of statistics has focused on shortcomings of economic measurement (e.g. Hoskyns and Rai 2007; Linsi and Mügge 2019; Yarrow 2020), statistical capacity problems (e.g. Devarajan 2013; Jerven 2013), and the functions of numbers in global politics (e.g. Davis, Kingsbury and Merry 2012; Fukuda-Parr, Yamin and Greenstein 2014). This article brings these topics together in a three-pronged typology of global statistical governance: standard-setting, capacity building, and compliance monitoring. The focus is at the level of IOs, rather than the downstream effects of their governance efforts in particular countries. This departs from studies that have focused on the shortcomings of statistics or their potential biases, but complements them by looking into the actual practices of statistical governance. Meanwhile, literature on cooperation and division of labor between IOs often focuses on more high-profile examples, such as the Poverty Reduction Strategy Papers (PRSP) (Momani and Hibben 2015). As a highly technocratic domain, global statistical governance tends to remain behind the scenes of those more salient aspects of IOs' activities.

I argue that the degree of inter-organizational cooperation is driven by what I call the bounded autonomy of the statistical epistemic community. This argument starts from the observation that most cooperation on statistics issues is initiated by IO staff rather than through formal joint efforts. The concept of bounded autonomy captures the varying degrees of autonomy possessed by IO staff in different sub-specialties of statistical governance. Highly technocratic functions, such as standard-setting, grant IO staff a high degree of autonomy and enable them to cooperate in a synergistic manner across organizational boundaries. That is, as long as such cooperation does not interfere with IOs' broader priorities. In contrast, cooperation is lower in functions that are closely bound up with the core priorities or mandates of IOs, such as macroeconomic surveillance or lending. Put differently, statistical governance is a form of expertise-driven governance in the shadow of IO mandates.

I support this argument empirically with qualitative data from 27 semi-structured interviews (some with multiple participants, for a total of 35 interviewees) at IOs and national statistical offices, combined with analysis of relevant documents such as reports of the IMF's Independent Evaluation Office and the World Bank's Independent Evaluation Group. Interviews were conducted between 2017 and 2019. At each of the organizations and statistical offices, interviewees included a mix of high-ranking officials (deputy directors, a retired director, sections chiefs) and staff selected for their expertise in specific areas such as national accounts or capacity building. At the IO level, I conducted interviews at UNSD, the UN Economic and Social Commission for Asia and the Pacific (ESCAP), the World Bank, and the IMF. At the national level, I conducted interviews at the Statistical Office of Thailand, Lao Statistics Bureau, and Ghana Statistical Services. The argument is also supported by interviews with statistical capacity building experts at the international development consulting firm Oxford Policy Management and participant-observation at the 2019 United Nations Statistical Commission meetings.

The evidence shows that statistical governance is indeed driven by experts, as long as the consensus of these actors does not conflict with the mandates or priorities of individual IOs. This argument has implications for a wide range of policy domains and academic debates. The article builds on literature about the politics of statistics and indicators (e.g. Fioramonti 2013; Linsi and Mügge 2019), but focuses on the actual global governance of statistics rather than their pathologies and functions. It also builds on literature about inter-organizational and inter-institutional dynamics in global governance (Gehring and Faude 2014; Kranke 2020; Momani and Hibben 2015). This is a unique case in which the policy domain is at once highly technocratic but also, in some instances, crucial to the data-driven governance of IOs. Moreover, as demand for data grows (for the SDGs, among other reasons), the pressures on national statistical systems are increasing. It thus matters how the UNSD, World Bank, and IMF – as the actors with the most in-depth and sustained relations with national statistical systems – coordinate their roles amongst each

other. The article proceeds with a literature discussion and the theoretical approach. The third section is the empirical analysis, followed by a concluding section.

The politics of statistics and the bounded autonomy of the statistical epistemic community

As a combined result of member state demands and bureaucratic culture, economic governance by IOs has become increasingly quantified in recent decades (Broome and Seabrooke 2007; Clegg 2010; Enns 2015; Harper 1998; Samuel 2014). Statistics are central to the “economistic” way in which IOs make sense of poverty, economic growth, and other issue areas (Mügge 2020a; Vetterlein 2012). They enable governance at a distance and make national economies legible in ways that reflect the mandates and organizational cultures of the IOs promoting them (Espeland and Stevens 2009; Hansen and Porter 2012).

Statistics can serve a straightforward information-providing function for IOs, but also a symbolic function. Namely, the creation of ‘authoritative’ indicators and statistical standards is an important source of informal power (Best 2012; Broome, Homolar and Kranke 2018; Freistein 2016; Sending and Lie 2015). More broadly, quantitative indicators and benchmarks shape policy goals and outcomes across virtually all global policy domains (Davis, Kingsbury and Merry 2012; Fukuda-Parr, Yamin and Greenstein 2014; Kelley and Simmons 2015). While numbers can be enabling, they can also distort and mislead, as in the case of the World Bank’s Doing Business Indicator (Broome, Homolar and Kranke 2018).

The measurement of statistical indicators is based on international standards such as the System of National Accounts (SNA) and the Balance of Payments Manual (BPM) (van Heijster and DeRock 2021; Linsi and Mügge 2019). In Mosley’s (2003, p. 335) words, “any global standard or institution, even one of a technical nature, also has distributional implications”. The technocratic nature of standard-setting can mask these distributional aspects and cement uneven power relations (on data and statistical standards, see e.g. Christophers 2011; Hoskyns and Rai 2007; more broadly, Jeffs 2008; Perry and Nölke 2006; Quaglia 2017; Quark 2012). For example, as Smith (2012) argues, the SNA undervalues contributions of low-wage workers in global value chains. Recent literature has paid attention specifically to the measurement of economic statistics. It argues that economic indicators such as GDP (Coyle 2014; Fioramonti 2013; Philipsen 2015; Pilling 2018), balance of payments statistics (Mügge and Linsi 2020), and unemployment statistics (Alenda-Demoutiez and Mügge 2019) at best fail to adequately capture what they claim to, and at worst are inherently biased.

IOs do not only create these standards, but also actively promote their uptake domestically. They do so in two ways. One is through a combination of monitoring and enforcement of

member countries' compliance with statistical standards. For example, the IMF checks data quality during Article IV consultations, and the UN monitors SNA implementation through an annual questionnaire (De Las Casas 2016). The other way IOs promote the uptake of standards is through statistical capacity building, given that capacity is a prerequisite for data provision. In reality, national statistical systems in developing countries often cannot produce the full range of statistics demanded of them, let alone the statistics desired by domestic policymakers and other data users (Devarajan 2013; Jerven 2013). Low statistical capacity is due mainly to budget constraints, but also to frequent staff turnover, access to training, and other organizational factors (Dargent, Lotta, Mejía and Moncada 2018; Round 2014; Taylor 2016). Meanwhile, demand for statistics by outside stakeholders continues to grow, spurred by initiatives such as the Sustainable Development Goals (SDG) indicator framework (Kapto 2019; Thérien and Pouliot 2019) and the IMF Data Quality Initiatives (Mosley 2003; Lombardi and Woods 2008; Reichmann 2016). Both the support and the demands that IOs place on member countries, especially low-income ones, have important consequences for domestic statistical practices.

The bounded autonomy of the statistical epistemic community

In sum, statistical governance matters. Statistical standards, statistical capacity, and data quality all impact the ways that domestic economies and the global economy are made legible, and thus shape policy interventions and investment. But we still know little about how economic statistics are actually governed. This article focuses on three of the most influential IOs in statistical governance, and builds upon literature on inter-organizational and -institutional cooperation, fragmentation, and division of labor (e.g. Biermann et al. 2009; Busuioc 2015; Gehring and Faude 2014; Jonsson 1986; Kranke 2020; Momani and Hibben 2015).

Momani and Hibben (2015) find that while the quality of cooperation is often low, the formal requirements for collaboration have increased since the launch of joint initiatives such as the PRSPs and the Financial Sector Assessment Programs (FSAP). Kranke (2020) argues that, in practice, IMF-World Bank cooperation has actually lessened in these same initiatives since the 2008 crisis. While, in the past, crises tended to increase cooperation between the two organizations, the trend has now been reversed (*ibid.*). Taken together, what these arguments suggest is that actual cooperation may be informal. That is, it might depend upon initiatives taken by staff to reach out to staff of other IOs, regardless of formal decrees.

However, it is not clear under which conditions informal cooperation of this sort is possible. I expect that informal cooperation is possible when staff have high levels of autonomy relative to the IOs that employ them. In any given function, the degree of autonomy is shaped mainly by two factors: 1) the degree of technical expertise required, and 2) the importance of that function to broader IO priorities. An example of a broader

priority is macroeconomic and financial surveillance by the IMF. These two independent variables negatively co-vary: where I observe higher levels of technical expertise, there is also lower connection to IO priorities. In other words, the *higher* the required technical expertise, the higher the autonomy. This is consistent with arguments about specialized knowledge in global governance (Haas 1992). At the same time, the *lower* the proximity to IO priorities, the higher the autonomy.

The first variable (technical expertise) is at the level of the governance function itself. For example, a high degree of technical expertise is a characteristic of the standard-setting function. The second variable (individual IO priorities) is at the intra-organizational level. For the latter, cooperation can be negatively impacted even if the proximity to IO priorities is high for only one of the organizations. This would apply *a fortiori* to analyses of two IOs, but less so for macro-level studies of, say, regime complexes, institutional complexes or governance architectures (Biermann et al. 2009; Gehring and Faude 2014).

This argument addresses both the internal organizational characteristics of IOs and the characteristics of the statistical epistemic community. In this way, the concept of bounded autonomy is influenced partly by Broome and Seabrooke's (2012) concept of analytic institutions. Analytic institutions are the "specialist units, departments, committees, adjudicatory bodies and others housed by or linked to IOs that develop the cognitive framework for understanding and solving policy problems" (ibid., p. 3). As Broome and Seabrooke (ibid., pp. 3-4) define them, analytic institutions, in contrast to "free-floating" epistemic communities, are expected to employ their analytic functions in the service of an IO's policy priorities. Thus, they are also expected to be subject to organizational constraints as well as the constraints (including principal-agent dynamics) on the IO as a whole. In the case of statistical governance, however, IO staff indeed belong to an epistemic community of statistical experts, and sometimes operate freely in that community. They fill this role simultaneously with their formal roles as staff in IO statistics departments. On this point, therefore, the concept of bounded autonomy departs from that of analytic institutions.

The statistical epistemic community

The epistemic community based around economic statistics is comprised of professional economists and economic statisticians. It is a community based on specialized expertise and shared norms. Mügge & Linsi (2020, pp. 2-3) identify four norms that shape the design of international standards for macroeconomic statistics. These norms are

comparability (the desire to compare statistics across countries), *continuity* (the ambition to build time-series datasets), *certitude* (the predilection for reliably quantifiable data), and *coherence* (the aspiration to integrate separate statistical domains into one overarching representation of "the economy"). (ibid.)

DeRock (2021) also argues that GDP measurement is guided by shared norms among national accounting experts, above all the norm of pursuing internationally comparable statistics through harmonization efforts.

The members of this transnational community are primarily employed by IOs or by national statistical offices. The community also extends somewhat to academia. Economic statisticians, including IO staff, publish in journals such as *Review of Income and Wealth* and meet at conferences such as those of the International Statistical Institute. While the epistemic community is based around expertise and norms, it also overlaps with the United Nations Statistical Commission (UNSC, not to be confused with the Security Council). UNSC was created in 1947 and has its mandate from the UN Secretariat. It is a subsidiary body of the UN Economic and Social Council (ECOSOC). UNSC is composed of a rotating selection of 24 member countries, elected every four years by ECOSOC. The UNSC convenes annually for a week of meetings at the UN headquarters in New York, attended by hundreds of statisticians from UN member states and IOs, not only representatives of the currently elected member states. There, UNSC discusses and makes decisions on a wide range of agenda items, such as launching new partnerships for regional capacity development, forming new working groups or task forces, beginning the revision of a statistical standard, and so on. Statisticians also meet in many small sessions focused on specific topics, such as the SNA, trade statistics, or the SDGs, and also in informal settings.

Statistical functions within international organizations

While IO staff indeed belong to this epistemic community, they are not always free from organizational constraints. At the end of the day, they are employees of IOs, and are thus tasked with furthering the priorities of those IOs. This section briefly sketches the statistics-relevant aspects of each IO.

UNSD is a division of the UN Department of Economic and Social Affairs (DESA), a department of the UN Secretariat. The mandate of UNSD is to “facilitate the coordination of international statistical activities and support the functioning of the United Nations Statistical Commission as the apex entity of the global statistical system” (UNSD 2020a, p. 3). There are four main priorities of UNSD: data (collection, processing, dissemination), methodology, capacity building, and coordination (ibid., p. 3). It is divided into five Branches, the most relevant of which is the Economic Statistics Branch (UNSD 2020b). UNSD is unique in several ways. First, it is not a distinct international organization in the proper sense, but rather a division of DESA. Nonetheless, it functions like an IO when it participates in inter-agency working groups alongside the World Bank, the IMF, the OECD and others. UNSD is more than simply a statistical unit of DESA or of the larger UN system, which is too complex to be compared to the World Bank or IMF (which, after all, are formally part of the UN system as well). Second, UNSD is unique in that its mandate is focused on statistics, unlike other IOs where statistics support other priorities.

At the World Bank, the unit responsible for the coordination of statistical issues and the maintenance of databases is the Development Data Group (DDG). According to a World Bank statistician, “DDG is a de facto statistics department rather than a real statistics department” (Interview 6). That is, there is no formal statistics unit equivalent to that of the IMF’s STA (see below). This has begun to change over time as DDG has evolved in the direction of a more formal statistics unit (ibid.). DDG is part of the Development Economics Vice Presidency, the research and data branch of the World Bank. Within DDG there are smaller units, such as the Data Production and Methods Unit that specializes in household survey methodology.

The IMF Statistics Department, also called STA, is one of nine functional and special services departments, and reports to the office of the Managing Director. It plays a central role in all three main categories of IMF activities, namely surveillance, lending, and technical assistance. Each of these functions relies heavily on regular and timely macroeconomic and financial sector data. STA has an advisory role within the IMF, such that staff of other departments consult with the statistics department on country data or methodological queries (Interview 7). “The most important mandate of STA” is to train member countries to provide IMF area departments with the data they need for analysis (Interview 8).

The UNSD, the World Bank, and the IMF have different priorities and engage with member countries in different ways. The following interview response sums up the differences in priorities:

There’s always a bit of a tension, because you know it’s never going to be perfect. All the different organizations do have different priorities. Here at the UN it’s very much about involving all countries. Everybody’s got one vote at the UN, you know, ‘leaving no one behind’. The real drive at the IMF is that you’ve got to maintain the stability of the world financial system, so you’ve got a slightly different focus. The World Bank is more about development goals. You’ve got a lot of overlap but it’s not 100 percent overlap. (Interview 1)

Those priorities, combined with the technical nature of statistical governance functions, shape the degree of staff autonomy and inter-organizational cooperation in each governance function. These dynamics are investigated in the next section.

Inter-organizational cooperation in global statistical governance

To what extent to the UNSD, the World Bank, and the IMF cooperate in global statistical governance? What drives the degree of cooperation? This section addresses these questions

empirically. The focus is on the three main types of statistical governance: standard-setting, capacity building, and compliance monitoring and enforcement. Standard-setting refers to the creation of international statistical standards and the methodological development that feeds into this. Capacity building refers to training, technical assistance, organizational support, and lending for national statistical systems. Compliance monitoring and enforcement refers to efforts to promote or enforce compliance with statistical standards and with the 'in-house' standards and requirements of IOs. Each of the following sub-sections follows the same structure. First, there is a brief description of the function and of the characteristics of the two independent variables (technical specialization and proximity to IO priorities). That is followed by discussion of how these characteristics have impacted governance practices, and of the wider implications.

High cooperation in statistical standard-setting

Statistical standard-setting involves the creation and revision of international guidelines such as the SNA and the BPM. Cooperation is high in statistical standard-setting because, more than other functions, standard-setting is highly technical and not directly linked to broader IO priorities. Neither IO executives nor member states closely manage the revision of standards or the methodological development that precedes it (Interview 5; see also DeRock 2021). Standard-setting is a theoretical undertaking, as the former editor of the SNA described in an interview:

It is a question of looking at how we think the system *ought* to work and looking at the pros and cons – why is it like this, and why isn't it like that, what would the problem be in changing it. And so it's very theoretical, as opposed to when you're out in a country, you're saying 'where on earth can I get figures for this or that'. (Interview 2)

The theoretical and technocratic nature of standard-setting has led to its delegation to inter-agency expert groups. As a result, IO staff engaged in this process function more as members of an epistemic community than as IO staff. They face few constraints, and the IOs' priorities – the creation or revision of international standards – are aligned. The staff of each IO bring expertise to certain aspects of the standards, but work toward a common goal (Interview 2). What is more, standard-setting does not conflict with any of the broader priorities of IOs. While harmonized statistics are required for lending, surveillance, and analysis, the precise methodologies underpinning those statistics are left to experts.

Each IO specializes in different types of economic statistics, but there is a great deal of overlap. As a whole, UNSD works on a wide range of statistics, including economic, environmental, geospatial, social and demographic statistics. The Economic Statistics branch specializes in national accounts, trade statistics, and business statistics, and also, to a lesser extent, on price statistics and others (Interview 3). The World Bank is

less specialized in methodology than the IMF and UNSD, but the Bank does assign staff to relevant interagency working groups and task forces for macroeconomic statistics (Interview 9). Earlier in the World Bank's history, DDG was focused primarily on macroeconomic statistics and a few major tasks, including the production of flagship World Bank publications such as the World Development Report, the compilation of debt statistics and other headline indicators, and managing databases. Since roughly 2000, DDG has increasingly specialized in poverty statistics and household surveys (Interview 6). At the IMF, methodological work is clustered in specialized divisions within STA, such as the Real Sector Division and the Balance of Payments Division. The IMF has a number of flagship statistical standards, most notably the Balance of Payments Manual (BPM), which first was published in 1948 (IMF 2016, p. 2). In addition to the BPM, the IMF also publishes the GFSM, along with several methodological guides and manuals such as the *Monetary and Financial Statistics Manual and Compilation Guide* and the *Quarterly National Accounts* manual.

These differences in emphasis do not impact the degree of cooperation, however. When statisticians come together in expert groups such as the ISWGNA, or in conferences like the UNSC sessions, they operate as members of an epistemic community. The Intersecretariat Working Group on National Accounts (ISWGNA), for example, is responsible for revising the SNA. It is made up primarily representatives of UNSD, the World Bank, the IMF, Eurostat and the OECD. Similarly, IO staff cooperate in the Intersecretariat Working Group on Price Statistics and the IMF Committee on Balance of Payments Statistics. Although such expert groups typically include representatives of member countries and other actors (such as regional development banks), often on a rotating basis, they are primarily composed of staff of IO statistical units. Although this governance arrangement is exclusionary – in that it privileges the knowledge of a small number of like-minded experts (see DeRock 2021) – it also enables a high degree of synergistic cooperation.

Standard-setting is not only highly cooperative within the revisions of particular standards like the SNA and BPM. There is also synergy between these standards. This ambition for consistency between standards is laid out in the UNSD publication *Guidelines on Integrated Economic Statistics* (UNSD 2013). The concepts of the BPM and the SNA have been aligned since 1993, and more closely since the 2009 BPM6 update, which was integrated with the 2008 SNA (IMF 2016, pp. 2-3). The GFSM is also conceptually compatible with the SNA and the BPM (ibid., p. 3).

The result of this autonomy and cooperation is that experts have a great deal of control of standards. This matters because the shared ideas and norms of statisticians are not neutral. For example, as Linsi & Mügge (2019) argue, certain concepts underpinning the Balance of Payments Manual – such as distinct national economies – do not match the reality of the globalized economy. Nonetheless, standard-setters have continued to

push for harmonization, resulting in inaccurate data on capital and trade flows (ibid.). And, in the SNA, shared norms and ideas regarding economic production boundary, environmental ‘externalities’, and other issues have shaped GDP measurement (Assa and Kvangraven 2021; Hamilton 2003; Waring 1999).

Moderate cooperation in statistical capacity building

Capacity building involves direct (such as training) and indirect (such as lending and aid provision) support for national statistical systems. In most countries, a national statistical office (NSO) is the main organ of a national statistical system, responsible for collecting data (either on their own or from other government ministries), compiling it according to standards, and reporting it to central government and to IOs. Cooperation in statistical capacity building has important implications. According a survey by PARIS21 (2019, p. 16), “more than half of all African NSOs perceive that capacity [programs] did not involve sufficient consultation between national and international stakeholders; worldwide, one third of NSOs consider that [programs] are not meeting their needs.” For one thing, donors are uncoordinated, often placing different demands on a single country at the same time for various types of statistics (ibid., p. 17). Moreover, “areas such as environmental statistics are not getting enough support, while the majority of recent funding has been allocated to economic and demographic statistics” (ibid.). As a UNSD statistician elaborated,

It’s that constant tension. Various organizations with funding come in and say, you know, we’ve got money, we want you to do a study of trade or environment or whatever the issue is, gender. And of course the countries aren’t going to say no to somebody with funding. (Interview 1)

Statistical capacity building does require specialized knowledge about statistical practices and methodology. Typically, the staff conducting capacity building are also involved in methodology, whether simultaneously or at different points in their careers. However, capacity building involves the practical application of that knowledge. In relative terms, it is moderately technical. Capacity building is also moderately linked to broader IO priorities. While statistical capacity aims partly at improving domestic development outcomes, it also affects IOs’ own capacity for evidence-based governance. If a member country struggles to produce indicators like GDP or price indices – or cannot do so in a timely or reliable way – IOs may have to make estimates or collect the missing data on their own. Thus, capacity building occupies an intermediate position between expert autonomy and organizational priorities.

There is some overlap in the types of statistics prioritized by the three IOs, particularly national accounts and price statistics (Interview 3; Interview 9). There is also geographic overlap. Because statistical capacity is lowest in developing countries, IOs focus their efforts there. For example, the IMF allocates nearly half of its statistical capacity building

to sub-Saharan Africa (IMF 2019a). At the same time, there is substantial variation in the approaches to capacity building. This variation reflects differences in the priorities and constraints of each IO. In short, UNSD prioritizes regional training, the World Bank prioritizes long-term capacity building, and the IMF prioritizes targeted capacity building to fill data gaps.

UNSD staff are rarely involved in country-specific capacity building missions due to inadequate funding (Interview 3). But UNSD has not requested this funding, according to the Branch Chief, because "...these are tasks that lie with other agencies" (ibid.). Instead, UNSD focuses on "sharing knowledge" at the regional level and between regions (Interview 1). As a UNSD statistician explained,

The broad idea is that we set the agenda and the program and the strategy, and the implementation happens at the regional commissions. It's not always that clean, but broadly that's the idea. (ibid.)

The UN regional commissions are crucial to this approach. Most on-the-ground capacity building is done through the five regional commissions, each of which has its own statistics division that cooperates closely with UNSD (Interview 1). In the Asia Pacific region, for example, the UN Economic and Social Commission for Asia and the Pacific (ESCAP) Statistics Division holds regional training seminars on methodological topics, such as specific aspects of the SNA or informal sector statistics (Interview 10). As an ESCAP statistician describes:

We don't want to run a regional workshop where we teach somebody what is an index number, for example, or what is CPI, what is GDP. We want a country to have a cadre of experts that can deliver this basic training. And when it comes to us, we can focus on something that is more advanced and requires highly specialized skills. (ibid.)

The World Bank takes a more long-term approach. World Bank statistical capacity building emphasizes long-term projects and the creation of large global partnerships and lending vehicles to fund statistical capacity building, such as the Marrakech Action Plan for Statistics (MAPS), established in 2004, and the Busan Action Plan for Statistics, its successor since 2015 (World Bank 2004; Ngo 2015).

There are four main statistical capacity building programs of the World Bank. The most relevant are the Trust Fund for Statistical Capacity Building (TFSCB) and the Statistics for

Results Facility Catalytic Fund (SRF-CF).¹¹ These are both multi-donor trust funds that provide grants for long-term (typically two or three years) statistical capacity building projects in low-income countries, sometimes in partnership with regional organizations (World Bank 2011, p. xv; World Bank 2019a, p. 4).

For the World Bank, statistical capacity building is understood as complementary to its broader prioritization of ‘development’ and poverty alleviation. This is reflected in the stated purpose of the SRF-CF, which is to

increase the capacity of developing countries in formulating policies and decision-making for development through sustained improvement in the production, availability and use of quality statistics. (ibid.)

Many TFSCB and SRF-CF projects have focused on supporting the creation a National Strategy for the Development of Statistics (NSDS), a roadmap and budget for a statistical system (World Bank 2019b, p. 4). The NSDS “has been the cornerstone of the World Bank’s statistical capacity building, and most projects have used the NSDS as their operative backbone” (World Bank 2017, p. 25). In practice, this means that the World Bank aims at long-term capacity as a tool to further its broader priorities, namely development through economic growth.

For the IMF, capacity building and technical assistance is closely integrated with the Fund’s surveillance and lending operations (IMF 2015; Reichmann 2015). “Surveillance is the bread and butter” of the IMF, and statistics are crucial in this respect (Interview 7). STA “focuses [capacity building] on areas that are considered most important to reduce risks and vulnerabilities to economic stability and, therefore, improve conditions for sustainable economic growth” (IMF 2018a, p. 10). The overarching strategy of STA is established by the IMF Executive Board. The most recent update to STA’s mandate was in 2018 and clarified in a report called *Overarching Strategy on Data and Statistics at the Fund in the Digital Age* (IMF 2018b). Regarding capacity building, the Strategy clarifies that the priority is to “address data weaknesses by integrating surveillance priorities with capacity development” (ibid., p. 10). The links between statistical capacity and surveillance priorities is also reflected in the Data for Decisions (D4D) Fund, an IMF statistical capacity building trust fund set up in 2018. D4D is

designed to be agile, integrated, and member-focused to help better align [capacity development] financing with member country needs and IMF surveillance priorities

¹¹ Of the remaining two World Bank statistical capacity programs, one, ECASTAT, is a regional program limited to Eastern Europe and the Commonwealth of Independent States. The other, STATCAP, is no longer active.

focusing in particular on addressing data gaps in low and low-middle income countries and fragile states. (IMF 2018a, p. 11)

This approach again reflects the moderate salience of capacity building for IMF priorities. While capacity building focuses on data gaps relevant to surveillance, it is also concerned with “member country needs” (ibid.).

There have been recent efforts to increase inter-organizational cooperation in statistical capacity building. These efforts result from a shared recognition that demand for statistical capacity by member countries is increasing. For example, cooperation has been formalized by the creation of the ISWGNA Task Force on Statistical Capacity Building (ISWGNA 2018b). According to a UNSD statistician involved in the creation of the Task Force, “It at least gives us the opportunity to meet monthly and discuss what we’re each doing and what the priorities are” (Interview 1). The Task Force has compiled an overview of all the various “tools” in existence for capacity building, such as IT systems or compilation guidance notes (ibid.). The stated aim is to create a “knowledge hub” where NSOs, especially those with the lowest capacity, can see which tools are available from each IO (ibid.). This way, countries can venue-shop. There is also informal cooperation and communication. The World Bank, for example, “tries to take into account what others are doing in capacity building” (Interview 6). An IMF statistician also described that the IMF and World Bank “have a relationship around capacity building” (Interview 7).

In summary, cooperation in statistical capacity building is moderate because both the technical specialization and the proximity to IO priorities are moderate. Capacity building must be carried out by specialized experts, but their aims are practical ones. And, while there is some functional overlap in IOs’ capacity building efforts, their approaches differ in accordance with their broader priorities. These factors result in a type of cooperation that is more superficial than in the case of capacity building, where experts work together in insulated working groups. In capacity building, we do not observe experts actively working together on common projects, but rather maintaining communication and coordinating their efforts to increase efficiency.

Low cooperation in statistical compliance monitoring

This final sub-section addresses IOs’ efforts to monitor or enforce compliance with statistical standards and data quality guidelines. Compliance monitoring serves a different function for each IO. For the World Bank and UNSD it is a complementary tool to promote statistical capacity and harmonization, respectively. For the IMF, it is directly linked to the surveillance imperative, and thus tightly intertwined with broader priorities. Compliance monitoring is also relatively less technical than standard-setting and capacity building. In some cases, it involves IO staff manually evaluating data quality or interrogating countries’ compilation methods. In many cases, it relies on countries

self-reporting these issues through detailed questionnaires. Although these efforts require specialized knowledge, it is a relatively lower degree than the other functions.

UNSD's approach to compliance monitoring prioritizes the SNA. The primary way that UNSD checks compliance with the SNA is through the UN National Accounts Questionnaire (UN-NAQ), a self-reported questionnaire sent to member countries annually (Interview with UNSD Statistician 1). The aim of the UN-NAQ is twofold: one, to collect national accounts statistics for UNSD's own publications, including *National Accounts Statistics: Main Aggregates and Detailed Tables* and *National Accounts Statistics: Analysis of Main Aggregates*; and second, to monitor the scope of implementation and compliance with the 2008 SNA (Interview 3). In the UN-NAQ, the category of 'conceptual compliance' is evaluated against a minimum required data set (MRDS), "a benchmark in the form of a set of national accounts that all countries are expected to compile" (ISWGNA 2011, p. 4).

The aim of the UN-NAQ is monitoring, rather than enforcement, of compliance with the SNA. There have been four revisions of the SNA, the most recent of which was published in 2008. The UN-NAQ shows that several countries are still on the 1968 version, and many are on the 1993 version in terms of compliance. A 2018 report by the SNA Advisory Expert Group found that, ten years since the release of the 2008 SNA, only 79 countries had implemented it (ISWGNA 2018a, p. 1) According to a UNSD statistician,

We don't get overly concerned about 1993 versus 2008; the changes weren't that big. But what it is, to us, is an indication that [countries] haven't really reviewed their accounts for quite some time. And so it might be based on old patterns in their economy. (Interview 1)

The World Bank is the least involved in compliance monitoring of the three IOs (Interview 9). To the extent that the World Bank is involved in encouraging member countries to comply with standards, it is primarily through informal mechanisms. The primary example is the Statistical Capacity Indicator (SCI). The SCI is the only comprehensive indicator for statistical capacity. Compliance with international standards, including the System of National Accounts and the Balance of Payments Manual, contributes to a country's score. In principle, the SCI aims to increase capacity (measured partly in terms of compliance) through mechanisms of transparency and peer pressure (Interview 9). "It can be good to have a comparison. People would see that other countries and also people within the Bank were paying attention, and they would do something about it" (ibid.). To a limited extent, the World Bank can also place pressure on national statistical systems through conditional lending. In addition to specialized vehicles like the TFSCB, the World Bank also provides loans for statistics through regular lending (in the form of IDA credits or IBRD loans). Regular lending can take the form of stand-alone statistical capacity

projects, or statistics can be part of, say, a large-scale public sector reform project (ibid.). If it is the former (statistics only), there is no need to add conditions. But in some such projects, a World Bank team does in fact add conditions – for example, a requirement to prepare statistics legislation (ibid.). Additionally, World Bank capacity building programs frequently include evaluating data quality (World Bank 2019a),

The IMF takes the most active role in compliance monitoring and enforcement. It employs a combination of binding and non-binding measures that exert pressure to report statistics in accordance with international standards. As described in a 2016 report from the IMF’s Independent Evaluation Office, the first non-binding step in ensuring data quality and compliance is technical assistance and capacity building, during which the IMF “promot[es] the implementation of quality standards” (De Las Casas 2016, p. 14). The second step is checking and refining the data received from countries (ibid.). And the third is conducting and publishing evaluations, including the Data Quality Assessment Framework (DQAF), which “evaluates the quality of statistical systems – their governance, processes, and products” (ibid., p. 15). The DQAF is also the basis for assessment of data quality in the annual Article IV consultations (ibid.).

In addition to quality, the IMF also monitors data dissemination. The IMF’s data dissemination standards were created in 1996 in response to perceived data gaps leading up to the 1994 Mexican financial crisis (Reichmann 2016). To the original Special Data Dissemination Standard (SDDS), the IMF has since added a lower and a higher tier: the Enhanced General Data Dissemination System (e-GDDS) and the SDDS Plus, respectively. The tiers are based on countries’ levels of participation in international capital markets, where the SDDS Plus is the most stringent and intended for countries with “systemically important financial systems” (IMF 2019b, p. 2). These standards monitor the periodicity (for example, monthly, quarterly, or annual) and timeliness of countries’ reporting of a set of macroeconomic and financial indicators.¹² Participation in these standards is voluntary, but there are clear incentives to participate. The incentives are mainly reputational – participation sends signals of transparency to foreign investors and other governments (Interview 8). As of 2019, 95 percent of IMF member countries had subscribed to one of the levels, over 58 percent of which subscribe to the e-GDDS (IMF 2020).

The data needs of the IMF are firmly tied to macroeconomic and financial surveillance. Typically a few key indicators – including GDP, debt and international reserves – get the most attention in country surveillance, according to an interviewee (Interview 8).

12 The e-GDDS requires dissemination of 14 categories of macroeconomic and financial data: national accounts (GDP), CPI, central government operations, central government gross debt, depository corporations survey, central bank survey, interest rates, stock market, balance of payments, external debt, official reserve assets, merchandise trade, international investment position, and exchange rates. The only non-economic or -financial category is population (IMF n.d., pp. 16-17).

Another interviewee discussed the centrality of surveillance in the IMF's support for national statistics:

When countries ask for advice, we take a look at what the needs are. For example, a country might have a large informal sector but a small financial sector. We actually have a big influence over what countries prioritize, and we also know a lot about the countries. We know what information they need to provide in order to carry out good surveillance. (Interview 7)

Binding data provision is written into the IMF Articles of Agreement (see De Las Casa 2016, for a comprehensive history). The “fundamental need for information was transformed into specific obligations for members” in the original Articles of Agreement (De Las Casas 2016, p. 1). The legal framework specifies a minimum set of data required from member countries and the punitive steps the IMF can follow “in the relatively infrequent cases of misreporting” (ibid.). The first disciplinary measure is censure, and if a country does not solve its data problems it can be declared ineligible to borrow from the IMF (Interview 8).¹³

Article VIII, Section 5 of the Articles of Agreement empowers the IMF to “require members to furnish it with such information as it deems necessary for its activities...” (quoted in De Las Casas 2016, p. 2). A recent example of breach of obligation under Article VIII, Section 5 is the case of Argentina's GDP and Consumer Price Index (CPI) in 2011. First, in 2011, “Argentina was found in breach of its obligations under Article VIII, Section 5 due to its inaccurate provision of CPI and GDP data” (IMF 2016). In 2012, the Executive Board assigned Argentina “remedial measures” to resolve the problems (IMF 2012). By 2013 the IMF ruled that progress had not been sufficient, and as a result issued a “declaration of censure” (IMF 2013). The censure included a deadline for adopting the remedial measures, which “[aimed] at aligning these indicators with the international statistical understandings and guidelines that ensure accurate measurement” (ibid.). The censure was removed in 2016. These binding measures substantially strengthen the IMF's leverage over national statistical systems. However, as De Las Casas (2016, p. 10) notes, “the information members share with the IMF de facto, on a voluntary basis, vastly exceeds the minimums required by the legal framework”.

The IMF's approach to compliance monitoring and enforcement is more far-reaching than that of the UNSD and World Bank because it is closely linked to the IMF's surveillance imperative. Thus, it is formally ingrained in the IMF's legal framework. The UNSD and

13 Censure consists of a formal statement of disapproval from the IMF Executive Board with a list of remedial measures that the country under censure is required to follow in a given time period.

World Bank also have incentives to monitor compliance, but their priorities in that regard are less salient. As a result, IMF statisticians do not operate as members of an epistemic community in this function, but in the service of IMF priorities. Nor is there any significant cooperation between the UNSD and the World Bank, because their priorities do not overlap in compliance monitoring. For the UNSD, compliance monitoring is linked to its original mandate from the UN Statistical Commission to promote international statistical harmonization. This connection is more salient than the World Bank's rather limited efforts in compliance monitoring to promote statistical capacity.

Summary of the findings

Inter-organizational cooperation in global statistical governance varies in relation to the degree of technical specialization and the proximity to IO priorities (see Table 1). Statistical standard-setting is the most technical of the three functions. IOs delegate methodological development to statistical staff, and further delegate standard-setting to inter-agency expert groups. This delegation does not impact IOs' broader priorities. As a result, statisticians cooperate closely as members of an epistemic community. Capacity building also requires specialized expertise, but in an applied form. It is moderately linked to broader IO priorities. For the IMF, for example, capacity building is targeted at statistics required for surveillance. Since the individual priorities of IOs diverge in at least some respects, the aims of and approaches to capacity building also differ. This leads to a division of labor between IOs, which they have sought to regulate through cooperative efforts such as the Task Force on Statistical Capacity Building. Overall, in capacity building there is a moderate degree of cooperation wherein statisticians have some leeway in their design and practice of capacity building programs, but within certain organizational limits. In compliance monitoring, overall cooperation is low. Technical specialization is lower than in other functions, and the proximity to IO priorities differs for each organization. It is particularly important to the IMF's surveillance priorities, which leads the Fund to pursue its own relatively coercive approach to compliance monitoring.

	Standard-setting	Capacity building	Compliance monitoring
Technical specialization	<i>High</i>	<i>Moderate</i>	<i>Low</i>
Proximity to IO priorities	<i>Low</i>	<i>Moderate</i>	<i>High (IMF)</i> <i>Moderate (UNSD)</i> <i>Low (World Bank)</i>
Expert autonomy and IO cooperation	<i>High</i>	<i>Moderate</i>	<i>Low</i>

Table 1: Degrees of autonomy and cooperation in statistical governance functions (table created by author)

Conclusion

The bounded autonomy of the statistical epistemic community shapes the degree of inter-organizational cooperation in statistical governance. This is a topic that has received little attention from scholars interested in inter-organizational interaction (e.g. Momani and Hibben 2015), nor from those interested in the global politics of (and by) numbers (Fukuda-Parr and Greenstein 2014). This article has engaged with both of these debates. It has identify the extent to which and how IOs actually cooperate on economic statistics three key functions of statistical governance, thereby contributing a macro-level analysis that compliments research about economic and social statistics focused on individual countries (Samuel 2015) or indicators (Coyle 2014). The article has also put forward an explanation of why cooperation varies between different functions of statistical governance, which has implications for research about IO cooperation and the agency of IO staff members.

The central argument is that cooperation is driven by the degree of technical expertise involved in a given function, as well as the degree to which that function is directly relevant to IOs' broader priorities. On the whole, cooperation increases as autonomy of IO staff increases, because statisticians work collectively as part of an epistemic community rather than on behalf of their respective organizations. As the previous section has shown, cooperation is high in statistical standard-setting, moderate in statistical capacity building, and low statistical in compliance monitoring. Because IO statistical units have been created to support broader priorities, such as surveillance or lending, they are often subject to organizational constraints. This is in line with the concept of analytic institutions (Broome and Seabrooke 2012). However, the article has shown that staff of IO statistical units may at times function as part of an epistemic community that transcends organizational constraints.

These relationships between IOs are important to understand in the context of increased demand for statistics and a growing reliance on quantitative indicators in global policy. By focusing on relations between and within the UNSD, the World Bank, and the IMF, the research covers the most relevant actors in global statistical governance. There is more to be explored in further studies. For practical reasons, the article has left out additional actors in statistical governance such as Eurostat, the OECD, and regional development banks. Private consultants, too, are often contracted by IOs for statistical capacity building. However, the UNSD, the World Bank, and the IMF are indeed the only actors with a truly global (rather than regional) remit over official economic statistics, and have thus been the focus of the article.

6

How GDP spread to China: the experimental diffusion of macroeconomic measurement¹⁴

Co-authored with Joan van Heijster

Abstract

Gross Domestic Product (GDP), one of the world's most influential economic indicators, did not become truly global until it was implemented by China. China officially adopted GDP as an indicator of economic performance in 1993 when the country abandoned its Marxist-inspired national accounting system and joined the internationally harmonized System of National Accounts. As such, it was the last major country to begin producing GDP figures according to international standards. Since then, GDP has become deeply ingrained in China's economic governance. Yet, the adoption of GDP was complicated by mismatches between the ideology guiding China's reform process and the economic ideas underpinning international statistical standards. The Chinese government's translation of the standards into the domestic political-economic context lasted nearly a decade. This process was not foisted upon China from the outside, but rather was driven by domestic factors in an experimental fashion. This is best characterized as an atypical case of diffusion and an unsuccessful case of translation. It makes clear that macroeconomic measurement is inherently political, not a set of neutral 'best practices'. The findings also point to the characteristics of the diffusion object as an underexplored but important factor that can undermine domestic attempts to translate or localize global policy ideas.

¹⁴ This chapter is an adapted version of: Van Heijster, J. & DeRock, D.J. (2020). How GDP spread to China: the experimental diffusion of macroeconomic measurement. *Review of International Political Economy*. <https://doi.org/10.1080/09692290.2020.1835690>. Both authors have contributed equally to this article, Van Heijster predominantly focused on the empirical material, DeRock contributed mostly to the theoretical section

Introduction

Gross domestic product (GDP) has taken root worldwide as a powerful tool for economic policy and analysis. Nearly all countries, with very few exceptions, produce GDP figures based on a common set of international standards.¹⁵ Despite the neutral appearance of GDP, it is rooted in neoclassical economics and is poorly suited to capture non-market economic activity (Herrera 2010; Mügge 2020; Smith 2012). Nonetheless, GDP has become thoroughly institutionalized even in countries that differ from the highly industrialized market economies from which it originated.

An explanation of the global spread of GDP is incomplete without understanding how it reached China. China was the last major holdout, having used an alternative indicator of national income until the 1990s. In the time since this late adoption, GDP has taken on enormous political and social significance for governance by the Chinese Communist Party (CCP) (van Heijster 2020; Wallace 2016), notably in the form of the GDP target, as well as for the rest of the world's perception of 'the rise of China' (Ambrosio 2012; Hopewell 2015; Nölke 2015). Yet the adoption of GDP by China was not a foregone conclusion. China adopted GDP even though the country's economic structures still differed decidedly from capitalist ones, the statistical bureaucracy was ill-equipped to produce coherent GDP figures, and Chinese economic power was likely big enough to resist outside pressures for statistical reform. How and why did China adopt GDP despite these unfavorable conditions? We aim to identify the drivers of diffusion in order to explain why the process took the trajectory that it did.

International political economy (IPE) literature has studied the diffusion and adoption of a wide range of norms, ideas, policies, and institutions, but little is known about the global spread of GDP measurement. In line with coercive or top-down accounts of diffusion, GDP is often portrayed as having been imposed around the world by Western countries and international organizations (IOs) (Fioramonti 2013; Philipsen 2015; Schmelzer 2016). Sociological approaches focus on socialization, learning, and emulation as drivers of diffusion (e.g. Chwioroth 2014; Marsh & Sharman 2009). These dynamics shed light on the ways that statisticians and policymakers, through transnational interactions, seek out foreign ideas to solve problems in domestic statistical practices. Neither of these approaches fully explain how GDP measurement was adopted in China.

We find that GDP adoption in China was an experimental – and ultimately unsuccessful – process of translation (Ban 2016) driven by domestic actors. GDP measurement is governed

15 North Korea is the only country that has not formally adopted the System of National Accounts (SNA), which prescribes the international standards for GDP measurement. Cuba follows the SNA 1968, an earlier version that is partially compatible with the most recent standard (Lequiller & Blades, 2014, p. 443).

by the System of National Accounts (SNA), an international standard designed to capture aggregate economic activity in capitalist economies. When China first attempted to calculate GDP, the country already had a national accounting system in place appropriate for centrally planned economies: the Soviet-inspired Material Product System (MPS). On the one hand, it became apparent in the early 1980s that GDP and other concepts from the SNA were becoming relevant in the context of economic reforms. On the other hand, domestic political debates and compromises over opposing blueprints for economic reform set the parameters for statistical reform.

This resulted first in a hybrid system of national accounting, an experiment based on grafting selected ideas from the SNA onto the MPS framework, and later contributed to abandoning the graft and adopting the SNA framework. The translation of GDP into this transforming domestic setting took nearly a decade, while the Communist Party rhetorically justified these statistical reforms as consistent with national ideology, and the complexity of national accounting systems and their foundations in economic theory complicated the smooth adoption of GDP measurement. We support this argument with qualitative data collected from one month of fieldwork in Beijing, analysis of official government documents and from World Bank archives, and interviews with statisticians in China and at the World Bank.¹⁶

China's late adoption of GDP, coupled with the country's size and political-economic influence, make it a crucial case to study. The findings add empirical and theoretical clarity about the global diffusion of GDP and its emergence as 'the world's most powerful number' (Fioramonti 2013; Philipsen 2015). As part of this special issue, the article reflects on how statistical standards spread globally, specifically how they translate into statistical practices on the ground. While Aragão and Linsi (2020) show that there is a great deal of leeway in statistical standards, this case demonstrates that standards can nonetheless impose rigid constraints when they clash with domestic statistical institutions.

Our analysis also critically engages with IPE literature on the diffusion and translation and localization of policies and norms (e.g. Ban 2016; Bell & Feng 2019; Eimer et al. 2016; Lai et al. 2017). This is an atypical case that can be considered as a failed instance of translation. It points to the unique features of an object of diffusion as a currently underexplored but important factor that shapes the diffusion process and its (temporary) outcomes, to such an extent that it can undermine domestic attempts to localize global policy ideas. The article also contributes to the substantial scholarship on China's reform and opening up process by shedding light on an aspect of reform that has received little attention by political scientists.

16 This article includes four interviews with Chinese statisticians and statistical consultants in international organizations (see reference list for more information), which covered a range of topics such as the interaction between World Bank and Chinese statisticians in the 1980s and early 1990s, the drivers and challenges for China's early statistical reform and the adoption of GDP measurement.

The article proceeds with an overview of the functions of national accounting systems and the ideas and norms embedded in international standards. This is followed by a discussion of diffusion theories and the theoretical approach that we apply. The main body of the paper is the empirical section and analysis in which we trace the adoption of GDP measurement in China over time. A final section concludes.

National accounting systems: international standards and economic ideas

National accounting systems provide an overview of the structure and evolution of a national economy (Eurostat 2014, pp. 21–22). They ‘give specific meaning to the economy and provide factual data for users’ (Eurostat 2014, p. 21). Historically, there have been two internationally-accepted national accounting frameworks: the System of National Accounts (SNA) and the Material Product System (MPS). GDP is an indicator of aggregate economic production derived from the SNA (Lequiller & Blades 2014, p. 15).

The SNA was developed in the late 1940s by economic statisticians in the newly formed United Nations Statistical Office and quickly spread around the globe in the postwar period (Kendrick 1970, p. 285; Ward 2004). The global implementation of the SNA has been an ongoing project for roughly seventy years, spearheaded by international organizations such as the United Nations Statistics Division, the International Monetary Fund (IMF), and the World Bank. Although weak statistical capacity in many countries continues to stand in the way of full harmonization (Jerven 2013), the SNA is now the only internationally accepted national accounting framework.

Until the 1990s, the Soviet Union and Eastern Bloc countries, Cuba, and China (as well as a few other Asian countries for shorter periods of time) employed the Material Product System. The MPS originated in the USSR in the 1920s and spread through the Council for Mutual Economic Assistance (CMEA)-countries and other communist countries in the 1960s.¹⁷ It was formalized in 1969 and published as an international standard in 1971 (Herrera 2010, p. 23). The analogous indicator to GDP from the MPS is Net Material Product (NMP). From 1949 until the mid-1980s the Chinese statistical system provided statistics solely on the basis of the MPS framework.

China’s adoption of the SNA (and thus GDP) was an atypical case of diffusion, as we argue below. Diffusion is ‘a process through which ideas, normative standards, or ... policies and institutions spread across time and space’ (Börzel & Risse 2012, p. 5). The literature

17 Member countries were: Bulgaria, Cuba, Czechoslovakia, East Germany, Hungary, Mongolia, Poland, Romania, Soviet Union and Vietnam.

on diffusion investigates the spread of norms (e.g. Acharya 2004; Finnemore & Sikkink, 1998; Price 1998), ideas (e.g. Ban 2016; Sell & Prakash 2004), policies (e.g. Bell & Feng 2019; Röper 2020), and institutions (e.g. Thomassen 2017).

The SNA is neither a norm, idea, or policy. It is a national accounting framework made up of ‘a set of concepts, definitions, classifications and accounting rules that comprise the internationally agreed standard for measuring such items as gross domestic product...’ (ISWGNA 2008, p. 1). As a ‘collection of practices and rules’ (Finnemore & Sikkink 1998, p. 891), the SNA and MPS can be considered institutions, albeit complex ones. And, as institutions, they are embedded with norms and ideas.

National accounts tackle four fundamental questions, as Herrera (2010, p. 4) summarizes: 1) What counts as productive economic activity? 2) How should activity be generally categorized and aggregated? 3) How should activity be defined and measured? And 4) how should or how might the necessary data be collected and disseminated?¹⁸ The internationally ‘agreed-upon’ (although contested) answers to these questions are formalized in the SNA. In this sense, the SNA is embedded with norms, defined as ‘standard[s] of appropriate behavior for actors with a given identity’ (Finnemore & Sikkink 1998, p. 891). In addition, the notion that ‘the SNA was appropriate for capitalist economies and the MPS for centrally planned economies’ (Herrera 2010, p. 3) is itself a norm.

There are also economic ideas embedded in the SNA and the MPS, both of which are grounded in economic theory. These concrete economic ideas, cause friction between the SNA and the MPS. National accounts indicators such as GDP are ‘transformations of primary data with the aid of statistical techniques and conceptual conventions’ (Bos 1995, p. 4). The concepts defined in the SNA – such as financial assets, the production boundary, or capital formation, to name just a few – are particularly influenced by economic theory. For example, ‘The definition of changes in prices and volumes include references to various index number formulae, e.g. Paasche, Laspeyres, Fischer and Tornquist’ (Bos 1995, p. 7).

The ideas embedded in the SNA clash with those of the MPS in three ways in particular. First, the MPS framework only considers material production as economic activity and thereby excludes a large part of the economy, particularly the service sector (Árvay 1994, p. 225). Second, the MPS uses administered prices instead of market prices to estimate the value of economic activity (Árvay 1994, p. 225; Jefferies 2015, pp. 14–17; World Bank 1992a, pp. 6–7, 104). Valuation in the SNA is based on the concept of current exchange value (Bos, 1995, p. 18). In Bos’s (1995 p. 4) words, ‘Valuation is at the heart of both economic theory and national accounting’. Third, MPS data collection methods report physical output numbers

18 Lequiller and Blades (2014) provide a comprehensive description of how the System of National Accounts (SNA) addresses these questions.

instead of financial and income flows, prioritizing information about the production side of the economy (Herrera 2010, p. 27; Holz 2004, p. 385). These differences between the SNA and MPS reflect variation in economic theory and ideology. There are also practical differences between the two systems. For one, the SNA makes use of a wide range of data sources, including sampling data, while the MPS primarily gathers data through total enumeration.

As a result, indicators derived from the SNA and MPS are not directly comparable. And attempting to measure GDP – an SNA indicator – within the MPS framework is technically challenging. Therefore, properly measuring GDP within an MPS framework entails far more than converting existing indicators or collecting new data. It requires an overhaul of the national statistical system, which in China was closely linked to the larger central planning apparatus (Herrera 2010, p. 20; Xu 2009, p. 447).

Theorizing the diffusion of the system of national accounts and GDP measurement

Despite the unique challenges of replacing one national accounting system with another, as described above, previous literature on diffusion offers several suggestions as to how this happened and what drove the process. The literature proposes a few plausible mechanisms of diffusion, namely coercion, socialization, localization and translation.

Coercive diffusion occurs when host societies are confronted with external pressure to conform to an idea or practice (Lai et al. 2017, p. 961). Coercion is a direct diffusion mechanism, meaning ‘An agent of diffusion actively promotes certain policies or institutional models in her interactions with a receiving actor or group of actors’ (Börzel & Risse 2012, p. 5). Literature on the political history of GDP (e.g. Fioramonti 2013, pp. 40–43; Philipsen 2015, pp. 131–135) and economic growth as a policy goal (Schmelzer 2016) tends to portray the diffusion of GDP measurement as a coercive process. Fioramonti (2013, pp. 42–43), for example, suggests that GDP ‘colonized the very lexicon of global governance’ and that ‘the GDP mantra was imposed on poorer nations’ by IOs and powerful states.

Socialization involves learning from, mimicking, or emulating other actors or practices. Socialized diffusion takes place ‘when actors attempt to solve problems or policy challenges in an environment that is rooted in uncertainty and bounded rationality’ (Lai et al. 2017, pp. 961–962). Domestic actors use emulation to ‘solve a problem or to overcome a crisis and look around for ‘best practices’ and institutional solutions...’ (Börzel & Risse 2012, p. 5). Emulation can be based on an instrumental rationality, or it can follow a (normative) logic of appropriateness (Börzel & Risse 2012, p. 9; Marsh & Sharman 2009, pp. 271–272). Chwioroth (2014) suggests that states might emulate policies because they are ‘fashionable’ among other states with similar characteristics.

Localization and translation emphasize the ways that domestic actors contest foreign norms, ideas, or policies and modify them in accordance with local contexts (Acharya 2004; Dafe, 2020; Eimer et al. 2016; Van Kersbergen & Verbeek 2007). Acharya (2004) describes localization as ‘...the active construction ... of foreign ideas by local actors, which results in the former developing significant congruence with local beliefs and practices’ (Acharya 2004, p. 245). Domestic actors are agents rather than passive recipients (Lai et al. 2017, p. 963).

Grafting is a particular strategy sometimes employed in localization wherein a new norm is framed in a way that resonates with an already existing and accepted local norm (Acharya 2004, p. 244; Price 1998). The process of grafting ‘creates composite products in which alternative ideas (or norms or policies or practices) comingle’ (Lai et al. 2017, p. 963). Grafting is often portrayed as a way to successfully introduce norms that might otherwise be rejected. But, as Lai et al. (2017 p. 963) argue, the product of a graft might be unstable and contain ‘[in]compatible component parts’ that cause it to fall apart after some time.

In translation, ‘rather than “copy and paste” ideas developed in foreign “labs,” receivers tend to actively filter and even reshape these ideas before “adoption”’ (Ban 2016, p. 18). One of the main limits on faithfully replicating foreign ideas, according to Ban (2016 p. 19), is the degree of local translators’ knowledge of those ideas. Actors who have access to transnational networks will be better acquainted with foreign ideas than those who do not, and the latter will rely more on competing local ideas (Ban 2016, p. 22). In the case of German pension policies, Röper (2020) argues that foreign ideas had little effect on domestic preferences but were used symbolically to advance an agenda already preferred by local agents.

In the mechanisms discussed above, localization and translation emphasize domestic factors in explaining diffusion outcomes. Socialization involves both external and domestic drivers, while coercion is a top-down process. Studies of China’s economic transition tend to emphasize domestic rather than external mechanisms, ‘with a range of studies finding China to be a difficult arena for external policy diffusion influences’ (Bell & Feng 2019, p. 1). However, processes of domestic-driven diffusion can lead to a wide range of outcomes. As Zimmermann (2016) summarizes, the outcomes of diffusion are often presented in existing research either as dichotomous (between rejection or full adoption) or in a continuum between these extremes (Zimmermann 2016, pp. 102–103). The steps on this scale are: rhetorical adoption; legal adoption; implementation in policy or practice; and individual internalization of the norm or idea (Zimmermann 2016). These steps are depicted in Figure 1, excluding individual internalization for parsimony.

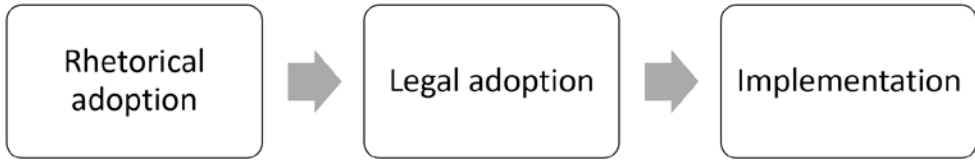


Figure 1: Typical model of diffusion (figure based on Zimmermann, 2016)

This ‘typical model’ assumes that foreign ideas or norms are first embraced superficially in the discourse of local actors; then formally adopted, but not necessarily put into practice; and finally implemented in practice. Intuitively, the degree of adoption increases at each step.

Our approach: tracing diffusion of national accounting practices in China

This article takes the sequence of diffusion less as an indication of the degree of adoption, but more as a means of tracing why and how actors have attempted to influence the outcome. We combine a political economy perspective with a diffusion lens to understand how actors attempt to translate foreign practices, and where they face obstacles and change direction. This approach allows for the possibility that diffusion outcomes are only temporary, and gives insight into the constant process of (re)negotiating foreign norms and ideas.

We identify a sequence of diffusion that ends in full implementation despite attempts by domestic actors to initially resist this outcome through translation and hybridization. This sequence differs from the model described above in important ways (see Figure 2). First, the Chinese government attempted to measure GDP but did not adopt the SNA (partial indicator implementation). Second, hybridization was attempted by blending two national accounting standards. Finally, only when the SNA was adopted, GDP was fully (albeit still imperfectly) implemented. This process also differs from many previous cases of diffusion (Figure 1) in that it does not begin with rhetorical adoption. Rather, discursive embrace of the SNA by the CCP was ongoing throughout the entire process and adapted to changing circumstances. This discursive component, which we call ‘rhetorical adaptation’, allowed the Chinese government to justify (to a domestic audience) the dilution and eventual abandonment of its hybrid system.

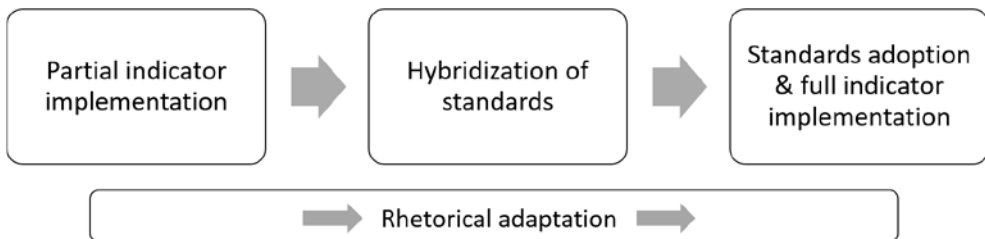


Figure 2: Step-wise model of GDP diffusion to China (figure created by authors)

Each of these steps is driven by different mechanisms (see Table 1). Partial indicator implementation was enabled by socialization, in which Chinese statisticians and policy members reached out to international experts for advice on statistical reform. Hybridization was an experimental process of translating selected ideas from the SNA into the domestic political-economic context, and thus grafting foreign ideas onto the foundation of the MPS. Finally, adoption of the SNA and full implementation of GDP was driven by the CCP’s recognition of growing mismatches, particularly between the MPS and the increasingly prevailing commitment to market-oriented economic ideas and macroeconomic management within the CCP. At each of these steps, rhetorical adaptation took place as the CCP justified statistical reforms to domestic audiences. Statistical reforms were framed as necessary to support economic reforms, in line with changes in the ‘operational ideology’ (defined in the following paragraphs).

Phase of diffusion process	Mechanism
Partial indicator implementation (1985)	Socialization
Hybridization of standards (SNA & MPS) (1986-1992)	Translation & grafting
Standards (SNA) adoption & full indicator implementation (1993)	Translation & pathologies of grafting (mismatches)

Table 1: Phases and mechanisms of GDP diffusion to China (table created by authors)

In China during the period of reform and opening up, local ideas were changing rapidly. To trace (changes in) the prevailing local ideas in China in the early reform period and analyze the extent to which they are in competition with the economic ideas of the SNA, we focus on state-sanctioned operational ideology. Ideology in this case does not refer to an individual belief system. The operational ideology of the CCP was a well-defined set of ideas used as top-down legitimation for the political direction taken by the Central Leadership. Here we follow Gewirtz’s (2020, p. 2, footnote 2) definition: The CCP’s ideology was explicit and official; ideology was ‘the ideas, theory or doctrine endorsed by the leadership, which are an expression of [its] interests that shape its perception of the world’.

This definition allows us to make a distinction between a fundamental (which refers to the core values) and operational (which refers to the practical application) dimension of ideology. This distinction is often made in studies of communist or socialist countries (Chen 1995). The operational ideology is thus also a political tool that enables political actors to maneuver policy processes and create room for policy choices that would otherwise be in tension with ‘fundamental’ ideology (Chen 1995, p. 12). The operational ideology is therefore a reflection of the political debates and power constellations about the appropriate political direction. For the period studied in this case, the operational

ideology mainly reflected power constellations between conservative and reformist camps over their ideas about the direction of economic reform. These camps consisted of high-level politicians, policymakers and economists (Gewirtz 2017, p. 371–375; Naughton 1995: 177). We trace these dynamics in the following section.

The winding road to GDP implementation in China

Partial implementation through international socialization (1978–1985)

GDP was introduced in China through a process of socialization initialized and steered by Chinese policymakers themselves. Starting from China's opening up in 1978, policymakers and statisticians learned about GDP measurement through interactions with foreign economists, academics and international organizations. The State Statistical Bureau (SSB) actively reached out to the UN Statistical Office to learn about alternative statistical practices and methodologies, other than the Material Product System (MPS) on which China's statistical system was based (Ferdinand & Wang 2013, p. 900; World Bank 1983, Annex A: 4.14). Also the World Bank, which China joined in 1980, exchanged knowledge about statistics and GDP measurement with Chinese policymakers and statisticians. The World Bank's 1981 report paid specific attention to the appropriate measurement of GDP statistics. It described in detail how GDP figures could be derived from Net Material Product (NMP) (World Bank 1983, pp. 220–263). World Bank staff also provided information about the use of statistics in economic policymaking and stressed that China lacked quantitative foundations for successful economic planning. They advised Chinese policymakers to reform the statistical system and pay more attention to income statistics and less to direct physical gross output numbers (World Bank 1983, pp. 166, 168–169). Additionally, World Bank reports made projections of China's GDP growth and used (estimated) GDP statistics to present China's economic position relative to developing countries such as Indonesia and India (Naughton 1990, pp. 750–751; World Bank 1983, p. 73; World Bank 1985, p. 21). The figures stressed the underdevelopment of China's service sector compared to other low- and middle income countries (World Bank 1983, p. 73). Through these exchanges, both policymakers and statisticians became familiar with the internationally accepted methodology of GDP and its value for economic policymaking.

Changes in domestic political debates about economic policy-making enabled the socialized diffusion of GDP. From the late 1970s reformists within the Chinese Communist Party (CCP) called for a pragmatic approach to economic policymaking, and Chinese officials and economists became increasingly willing to engage with alternative models to find 'best practices' facilitating China's economic development (Gewirtz 2017, pp. 31, 34). In line with this move away from dogmatic policymaking, reform-minded Chinese politicians initiated the learning process through which diffusion of (foreign) economic ideas and tools, including knowledge about statistics such as GDP, took place. Economists,

policymakers and other experts were sent abroad to conduct fieldwork trips and study the development models of foreign economies, including the US (Gewirtz 2017, pp. 52–56, 62). They increased their interaction with the World Bank and invited experts and foreign officials to conduct missions and exchange knowledge about economic reforms and development (Bottelier 2007, pp. 242–243).

Policymakers took the analyses resulting from these interactions seriously. They widely circulated World Bank reports throughout the bureaucratic apparatus and, based on the findings, started working with the World Bank to implement new planning techniques (Bottelier 2007; Naughton 1990, p. 750). The World Bank's GDP calculations provided insight into the service sector, which had grown as a result of economic reforms in 1978 and 1984 (Brødsgaard & Rutten 2017, p. 87; World Bank 1992a, p. 43). Diffusion of economic ideas and practices through learning thereby contributed to the reformists' goal of accomplishing economic growth and development with new economic policies while opening up to the outside world. Reformists on the highest political level explicitly used GDP to increase their interaction with international organizations and facilitate debates about economic reform. Most notably, Deng Xiaoping formulated China's development goals in terms of GDP. He proposed a GNP target of US\$1,000 per capita in 2000 as China's primary development goal (Deng 1979), and communicated this goal internationally.¹⁹

Despite outside interference in policy debates by international stakeholders, Chinese policymakers kept ownership over the interactions and were able to steer processes of diffusion, only selectively taking up the ideas and policies suiting them best. The World Bank applied a cautious and pragmatic approach to intervening in China's internal policy debates, and accommodated local preferences and ideas about reform (Lim 1993, pp. 9–12, 16; Interview 11). It presented a variety of (practical and ideological) options which were within the Chinese political parameters (Lim 1993 p. 12). The World Bank even invited Eastern European experts, from non-World Bank member countries, to share views on economic reforms with their Chinese counterparts (Lim 1993, pp. 9–11). This non-coercive approach provided Chinese policymakers room to set the pace of economic reform and adopt only those elements they were interested in (Lim 1993, p. 10).

19 In the 1980s and early 1990s China's GDP and Gross National Product (GNP) were virtually identical (World Bank, 1994, p. v). Although in the time period researched in this paper, the World Bank and Chinese politicians generally referred to GNP in the Chinese case, Chinese sector and Input-Output statistics were all in GDP terms. Therefore we consistently refer to GDP instead of GNP, since the statistical difference was very small, and currently GDP is mostly used over GNP. The main difference between the measures is that GDP measures the total production within a country regardless of ownership, whereas GNP only counts the production and income produced by the county's citizens, regardless their place of residence.

As a result, Chinese policymakers were able to selectively respond to international demands for statistical harmonization. First, they negotiated specific official statistical estimates. In 1980, Chinese policymakers disputed the World Bank estimates of Chinese GNP per capita. China argued it was US\$150, while the World Bank estimated a figure of US\$250. They settled on a GNP per capita of US\$180, which became the basis for further calculations (Interview 11). Additionally, until 2002, China refused to take part in the World Bank's International Comparison Program (ICP), an international price survey used to obtain purchasing power parity income (ppp) (Wade 2012, p. 18; World Bank 2018). Even then, it only entered eleven cities into the survey, leading to an overestimation of price levels (Wade 2012, p. 18).

In the case of GDP adoption, the outcome of this non-coercive, socialized diffusion process was the rhetorical adaptation to GDP and partial implementation of the measure in China. In 1985, the State Council acknowledged that, conceptually, GDP could fully reflect 'the scale and level of national economic and social development' (State Council 1985), a comment that in the pre-reform period would have caused ideological debates because it seriously conflicted with the prevailing Marxist conception of the economic production boundary (Jiang 2002, p. 28; Qiu & Song 2010, p. 30). Nevertheless, by mentioning that GDP and service sector statistics could also reflect social development, the State Council rhetorically adapted the concepts to the Chinese context, making it appropriate for the statistical bureau to produce it.

In 1985, Chinese statisticians produced China's first official GDP measurement based on the instructions from the World Bank report. They derived GDP from NMP and made crude estimates of concepts which were missing in this communist measure of national income, but crucial to measure GDP. To produce a GDP figure, the SSB accounted for the service sector, adding 13% of aggregate NMP to the official NMP figure (World Bank 1992a, p. 17). Even though the output suggested an internationally comparable GDP statistic, this first official estimate was still far removed from SNA methodology, let alone constituting the rhetorical adoption, legal adoption or implementation of a new national accounts framework. Instead, the diffusion of SNA to China took off with rhetorical adaptation and partial implementation of GDP.

Translating statistical reform and the hybrid system experiment (1984–1986)

After adopting an ad-hoc version of GDP measurement, the next step in the diffusion process was a debate about the adoption of the SNA resulting in the decision to develop a hybrid statistical system in 1986. The core idea behind the hybrid system was to provide national accounts summary statistics compatible with both the Material Product System (MPS) and the System of National Accounts (SNA) (World Bank 1992a, p. 1). The system would continue producing MPS aggregates, based on Marxist-Leninist economic concepts, while also producing SNA indicators, focusing on GDP in particular (World

Bank 1992a, pp. 102–103). Such a framework for national accounts was uncommon, and with the exception of Hungary (Árvay 1971), unseen in the rest of the world. It was a local solution which translated domestic political settlements into statistical system reform.

The hybrid statistical system was the result of a process of translation. Statistical reforms were adapted to reflect political compromises about the direction of economic reform within the Chinese bureaucracy. Broadly, two competing visions on economic reforms existed: a conservative camp which advocated continuing economic planning and only allowing market forces to enter some (less important) sections of the economy, and a reformist camp which wanted to abolish mandatory economic planning and shift state control to economic tools implying reforms on prices, taxes and interests rates (Naughton 1995, pp. 176–177). Statistical reforms were a fraction of the general economic reforms, and thus not at the forefront of political debates. The implications of reforms for statistics were discussed in the slipstream of other more prominent decisions about price reform, inflation and industrial reform. Statistical system reform however did not neatly mirror actual changes in the economy. Instead, decisions about statistical system reform were taken only when the direction of economic reforms was officially agreed upon between the camps. It would reflect the distribution of power and ideological compromises reached between the groups. A new operational ideology reflected changes in power relations and direction of reforms and therefore functioned as a filter for translating statistical reform to the Chinese context.

A new operational ideology under the slogan ‘socialist planned commodity economy’, endorsed by the CCP Central Committee on 20 October, 1984, shaped the diffusion process in two ways. First, as it gave a new impetus to further economic reforms, it set statistical system reform in motion. Directly after launching the slogan, the State Council established a Leading Group on National Accounts Reform in November 1984 which was tasked with providing recommendations for a new statistical framework (World Bank 1992a, p. 102). It urged the Leading Group to separate theoretical debates and practical work to avoid delays in the reform process (Yue 1989). The diversity of the stakeholders in the group, ranging from economists and financial experts to bureaucrats and statisticians, and attendance of important Chinese authorities such as the State Planning Commission and Ministry of Finance, shows that these discussions were of high political importance (Interview 12; NBS, 1984; World Bank 1992a, p. 102).

Second, the new operational ideology facilitated the translation of national accounting practices. Because the slogan was ambiguously formulated, the degree to which foreign statistical practices could be introduced was left to the Leading Group for interpretation. The operational ideology purposefully contained contradictory elements of planned and market economies. Premier Zhao Ziyang carefully constructed the slogan combining elements of competing economic models – ‘commodity economy’ alongside ‘planned

economy' – to gain acquiescence from both the reformist and conservative camp (Naughton 1995, p. 177). The State Council asked the Leading Group to design a national accounting framework consistent with this operational ideology. The parameters were that the framework should: (1) serve the needs of the government in carrying out economic reform policy (2); be in line with the actual economic situation in China (3); be guided by Marxist doctrine (4); and borrow from strengths of foreign national accounting systems (Yue 1989).

The Leading Group discussed three options, but only the option for a hybrid statistical system was seen as appropriate. The two other options were ruled out because they would only partially address the scope of the new operational ideology or stretch beyond its limits. These options were: continuing with the MPS framework, or a gradual switch from MPS to SNA concluding in the full adoption of the foreign accounting framework. Sticking to an MPS framework corresponded with the idea that the Chinese economy was guided by economic planning only, neglecting market-driven elements such as market prices or free allocation of goods and services. It implied only gathering material output statistics for the purposes of tracking output targets and drafting economic plans. Full adoption of the SNA meant committing to the idea that the economy was driven by the market and economic planning would completely disappear. It implied abandoning material output statistics and developing statistics measuring new concepts as market prices and interest rates. Both options would contradict the political compromise of the operational ideology and fail to satisfy both the reformists and conservatives.

In 1986, after two years of discussion, the Leading Group recommended the hybrid system. The State Council endorsed the recommendation, emphasizing that the hybrid system would serve the needs of the socialist planned commodity economy (World Bank 1992a, pp. 101–103). When justifying the hybrid system as appropriate to the operational ideology, the State Council stated that 'during a period of transition the economy will change, but even in the post-reform evolved state, central planning will coexist with a large market-oriented dimension' (World Bank 1992a, p. 103). SSB statisticians also stressed this point, and explained that the production of MPS aggregates was justifiable given that in a planned commodity economy, economic planning would continue to play a role during economic transition (World Bank 1992a, p. 101).

Furthermore, the SSB and reformers argued that adding elements of the SNA to China's national accounts framework provided information about new market activities and useful indicators for economic management (World Bank 1992a, p. 103). Thus, the hybrid system was a compromise that pleased both the conservative and reformist camps while sticking to the operational ideology. The domestic political conditions facilitated the experimental introduction of limited aspects of the SNA.

Implementing the hybrid system – grafting statistical frameworks (1987–1992)

After the rhetorical adaptation of the SNA by the Leading Group, SSB statisticians started to implement the hybrid system by grafting elements of the SNA into its MPS-dominated framework. The hybrid system aimed to capture SNA concepts including household consumption, business investments, government spending, and GDP. It initially calculated only the annual production-side estimation of GDP. In 1989 it added annual expenditure-side estimation, and in 1992 it introduced the first quarterly GDP estimates (Xu 2009, p. 447; see Table 2). To measure SNA indicators, in 1987 the SSB developed a transition Input-Output (I-O) table from which these concepts could be derived (Qi & Chen 2007, p. 1; World Bank 1992a, p. 2). However, the table was not entirely designed according to MPS conventions, nor to SNA conventions. It distinguished between material and non-material production, reflecting a theoretical distinction based on MPS conventions (Guo, Sonis & Hewings 1999, p. 318; Qi & Chen 2007, p. 1; Xu 2009, p. 450). Furthermore, the table was supposed to be constructed based on meaningful producer prices, but instead confounded implicit subsidies, taxes and prices (World Bank 1992a, pp. 20–21). As a result, the hybrid graft deviated from SNA guidelines, as did GDP measurement. GDP was still mostly derived from NMP data; missing variables were measured through ad-hoc surveys rather than consistently applying the same data collection methods and concepts (World Bank 1992b, p. 7), and some concepts were not measured at all or presented according to MPS conventions. For example, depreciation rates were set centrally by the government and therefore did not correspond to the concept of economic depreciation according to the SNA standards (World Bank 1992a, pp. 18–20).

The outcome of the translation process was a unique graft of two national accounts frameworks in which China essentially adopted a façade of the SNA, not the details (Interview 11). Statisticians used old measures and data collection methods to produce new indicators (Interview 13). The original intention was not so much to fully adopt international standards as to develop a framework that suited domestic economic and political conditions at the time. Contrary to the ‘typical model’ of diffusion (Figure 1) in which legal adoption follows rhetorical adoption and precedes implementation, the hybridization phase did not signal commitment to legal adoption of the SNA. Instead, the hybrid system was rhetorically framed as a domestic solution, only taking up foreign ideas that could be interpreted as consistent with the operational ideology. As a result, the SSB proceeded to the implementation of a limited range of SNA concepts, following an atypical diffusion sequence. While this process was intuitive from an ideological and state planning perspective, it caused significant technical problems in the production of economic statistics. The next section zooms in on the instabilities of the graft and how these contributed to its eventual disappearance.

Phase	Date	Events	National accounting practice & metrics <i>(blank cell indicates no change)</i>
Partial Implementation	1980	China joins World Bank	Net Material Product (NMP) calculated within Material Product System (MPS)
	1981	SSB resumes national income estimations	
	1981 - 1983	World Bank & international advisors train statisticians & policymakers	
	1983	World Bank report published	
	October 1984	CCP Central Committee formally endorses 'socialist planned commodity economy'	
	November 1984	State Council establishes Leading Group on National Accounts reform	
	1985	SSB measures China's first official GDP figure	MPS & NMP + Ad-hoc GDP (derived from MPS statistics)
Hybridization of standards	1986	Leading Group chooses Hybrid System; State Council endorses Hybrid System	
	1987	SSB develops transition Input-Output (I-O) table	MPS & NMP; GDP; Other SNA concepts derived from I-O table: household consumption; business investment; government spending
	1989	SSB measures first GDP annual expenditure-side estimation	MPS & NMP; GDP (annual production-side estimation & annual expenditure-side estimation); Other SNA concepts
	November 1990	World Bank statistical mission China – SSB expresses support for Hybrid System	
	1990 -1991	CMEA-states stop using MPS & switch to SNA	
	January – February 1992	Deng's 'Southern Tour' – reformists regain power	
	August 1992	State Council approves China National Economic Accounting System (trial version)	GDP primary indicator; GDP quarterly production-side estimation; Continuation of NMP from MPS; Additional SNA concepts: flow of funds compilation
	October 1992	14 th Party Congress introduces 'Socialist Market Economy'	
	Standards & full indicator adoption	October 1993	China officially adopts SNA as national accounts framework - SSB stops measuring NMP

Table 2: Timeline of GDP/SNA diffusion process in China (table created by authors)

Adopting the SNA: the influence of increasingly serious measurement mismatches (1987–1993)

The hybrid statistical system was only a temporary institutional solution. Several dynamics caused the hybrid to lose relevance and increased the appeal of the SNA. These dynamics are threefold and each comprise a different type of measurement mismatch, namely: practical, comparability, and ideological. These measurement mismatches provide insight into the dynamic (internal and external) challenges to a graft that influence the composition of the graft or lead to its disappearance. Which of the instabilities to the graft lead to changes in the outcome depends on the drivers of the diffusion process.

Practical mismatch – national accounts frameworks unsuitable for grafting (1987–1992)

The first dynamic that caused instability to the hybrid is the practicality measurement mismatch. The graft, and its competing ideas about the economy within, had been rhetorically justified in light of domestic political debates, but domestic discussions about whether the national accounts frameworks were actually suitable for grafting took place only in the background. National accounts systems are highly technical objects of diffusion. To produce meaningful and interpretable statistics, conceptual categories need to be applied consistently and systematically and data inputs need to suit the categories accordingly. For example, “for each product distinguished in the national accounts total supply (from imports and national production) must equal total use (intermediary and final consumption, export, capital formation and change of stocks)” in both volume and price (Nijmeijer & Hiemstra 2008, p. 5). This systematic approach prohibits mixing concepts from different national accounts frameworks. For instance, when products are valued differently, through market or administered prices, this causes discrepancies between the total supply and use value of different products, causing serious measurement biases.

As a result, when Chinese statisticians started to implement the hybrid system in 1987, practical measurement mismatches caused problems to the graft. The system caused heavily skewed GDP figures for two reasons. First, the data collection methods in the hybrid system were responsible for the most serious shortcomings (World Bank 1992a, p. 16). The data reporting system was almost entirely based on the MPS framework and did not correspond with the scope and concepts from the SNA framework (Holz 2004, p. 395; World Bank 1992a, p. 16). It was unable to cover large parts of the economy, especially the service sector, private sector and rural economic activities (World Bank 1992a, p. 14). Official statistics did not reflect economic output by self-employed citizens and new small- and medium sized and private companies (World Bank 1992a, pp. vi, 14, 32–33). Additionally, there were multiple conceptual definitions for the service sector. Authorities had to convert MPS data to meaningful concepts in the SNA framework, but often interpreted the data differently, causing conceptual inconsistencies within the framework (World Bank 1992a, p. 42).

Second, the valuation of prices according to MPS conventions also caused problems. The MPS values economic activity according to transactions in the form of physical outputs or according to administered prices, not through the principle of value-added according to (market) prices (World Bank 1992a, pp. 7, 12, 14, 104). This caused various evident valuation issues in practice. Many transactions in the service sector were priced too low relative to market prices in general or compared to products in the primary and secondary sector (World Bank 1992a, pp. 13, 47; Xu 1991). As a result, real income flows and inflation were underestimated (World Bank 1992a, pp. 13, 46–47; Xu 1991). Additionally, government subsidies counted as compensation for enterprise losses in the MPS framework, rather than government demand as in the SNA (World Bank 1992a, p. vi). Furthermore, industrial enterprises and other reporting units often reported data in actual government-regulated prices, rather than converting them to market-price values (World Bank 1992a, p. 13; Xu 1991). In sum, price data did not correspond to the actual value of market prices, which led to practical measurement mismatches with measures such as GDP that are based on market prices.

The data collection and valuation issues caused serious biases to GDP statistics. Aggregate GDP statistics were underestimated and did not reflect the true extent of economic activity (World Bank 1992a, p. 13; Wu 2000; Xu 1991). It also distorted the structure of GDP and conflated the values added of different sectors. The value-added to the whole economy of rural services was underestimated and that of industry was overestimated (Wu 2000). GDP growth figures, on the other hand, were overestimated mainly because of valuation problems and the underestimation of inflation (World Bank 1992a, p. 49; Wu 2000).

These problems were cause for concern with Chinese authorities and statisticians (World Bank 1992a, pp. 16, 21). They were aware of the problems and worried about the accuracy of service sector and consumption statistics, the validity of price statistics and whether GDP statistics reflected the economic structure (World Bank 1992a, pp. 16, 21, 23, 54). The Chinese authorities realized that the hybrid system was not capable of producing accurate and reliable GDP figures. Although these concerns were not outspoken issues in debates about economic reform, reformists presumably had more interests in solving GDP's biases by adopting the SNA framework than conservative policymakers. General biases in the data were problematic because they left reformers with less information to evaluate or draft new market-driven reforms. More specifically, due to the biases, 'successful' outcomes of key reforms such as the growth of the service sector were not visible. At the same time, conservatives could use exaggerated GDP growth rates to point to negative consequences of economic reforms such as the 'overheating' of the economy.

However, despite the practical measurement mismatch and incentives for reformists to tackle biases in the system, Chinese policymakers continued to express their commitment to the hybrid system in the early 1990s (NBS 1996, 2017; World Bank 1992a, p. iv; Zhi

1992). Practical measurement mismatches did hamper the usefulness of the system, but this dynamic alone was insufficient to alter the direction of the diffusion process and abandon the hybrid system for the adoption of the SNA.

Comparability mismatch – international abandonment of the MPS (1990–1991)

In addition to the practical measurement mismatch, a comparability mismatch also damaged the relevance of the hybrid system. Within a short time period the MPS disappeared as an alternative to the SNA framework on the international level. In 1990, most former Soviet states rapidly and quite unexpectedly stopped using the MPS framework and switched to the SNA (Herrera 2010, p. 88). They made significant changes to their statistical systems in 1990 and 1991 and fundamentally altered their statistical practices (Herrera 2010). By convincingly embracing the SNA, the former Soviet countries contributed to the hegemonic status of the market-oriented SNA framework (Herrera 2010).

This external event further reduced the international comparability of hybrid system statistics. First, the indicator NMP became less relevant for making comparisons once former Communist countries abandoned it. Second, international organizations shifted their focus fully to the SNA framework, which affected China's statistical cooperation with these organizations. The aim of the World Bank's mission in 1990 was 'assisting China to publish statistics in SNA-consistent categories' (Interview 13). Chinese policymakers and statisticians realized that the rest of the world adopted a different system, and tried to do what was necessary to compare themselves with others (Interview 13; NBS 2003). They adjusted the hybrid system to give the SNA, and particularly GDP, a larger role in the statistical system.

In August 1992 the State Council implemented the 'Chinese System of National Accounts (pilot program)', essentially an extended hybrid system (State Council 1992; Xu 2014). In this framework GDP became the primary indicator (Interview 14). However, China did not yet officially switch to the SNA. It still produced NMP figures and widely used MPS data collection methods. China was still committed to the hybrid system and made small changes within it. The practical and comparability measurement mismatches did not trigger the abandonment of the MPS in favor of the SNA. As with earlier phases, changes on the domestic political level ultimately drove changes in the diffusion outcome.

Ideological mismatch – adopting the SNA (1992–1993)

Although Chinese stakeholders were aware and concerned about the practical and comparability mismatches, these instabilities of the graft did not trigger an official switch to the SNA. It did not happen until 1993 when Chinese political stakeholders settled on a new operational ideology. Until 1992, there was no political space to deviate from the 'socialist planned commodity economy' that justified the choice for the hybrid statistical

system. In 1988, the conservative camp under guidance of Chen Yun won ground compared to the reformist group of Deng Xiaoping. The conservative policymakers opposed further (price) reforms and wanted to consolidate central control to stabilize the economy (Gewirtz 2017, pp. 198, 209). They gained control over key reform institutions in the State Council and downgraded the role of the System Reform Commission (Gewirtz 2017, pp. 202–206). In response to conservatives' criticisms on inflation and overheating of the economy, economic policies moved toward 'rectification' and retrenchment which recentralized planning as opposed to further market reforms (Gewirtz 2017, pp. 214, 228). These domestic political developments strengthened the justification for a hybrid system and the production of planning statistics, despite the incentives provided by the practical measurement mismatch to alter the statistical system.

However, when power shifted back to the economic reformists in 1992, the hybrid system graft came under serious pressure, leading to an ideological measurement mismatch. From 1991 onwards, calls by reformists to give the market a bigger role and return to market-oriented growth policies started to gain ground again in policy debates (Gewirtz 2017, pp. 240–247). High-level politicians also resumed encouraging the adoption of ideas from abroad, capitalist countries in particular (Gewirtz 2017, p. 251).

Moreover, politicians discussed the redefinition of the operational ideology to a 'socialist market economy'. The term corresponded to the idea that (indirect) macro-economic management and new market-oriented reforms were the appropriate direction of economic policymaking (Gewirtz 2017, pp. 245, 251, 254, 258). The Fourteenth Party Congress officially labeled China a 'socialist market economy' in October 1992 and the concept was codified into the Chinese constitution in March 1993. This created an ideological mismatch with the hybrid statistical system that was in place.

The new operational ideology made no reference to centralized planning (Gewirtz 2017, p. 254). As a result, the political mandate for producing traditional planning statistics disappeared. Therefore, it became acceptable to abandon the MPS, a system intrinsically tied to the idea of centralized economic planning. Additionally, the new operational ideology emphasized the importance of the market as the primary form of economic organization and indirect economic management by the state (Gewirtz 2017, pp. 251, 254, 258). Fully switching to a market-oriented national accounting framework was therefore not contradictory to the new ideology, but actually an appropriate translation of it.

In October 1993, the statistical bureau solved the ideological mismatch when it stopped measuring the MPS-based national income indicator NMP. Thereby it officially switched to SNA as its national accounts framework and abandoned the hybrid graft (Xu 2009, p. 447). Several Chinese accounts of this step in the diffusion process justify the choice of adopting SNA in light of the new operational ideology. The former deputy director

of the NBS stresses that China being a socialist market economy led to eliminating NMP measurement and other MPS aggregates and that the operational ideology laid the theoretical foundation for adopting the SNA (Min & Xu 1997; Xu, 2001). Almost 10 years after it first adopted GDP measurement, China adopted the SNA and committed to measuring GDP according to international standards.

Conclusion

China's GDP receives much attention, whether it is to assess economic performance in light of the US-China trade war or COVID-19 outbreak, compare the CCP's performance against its internal GDP targets, or to evaluate the accuracy of Chinese GDP figures in light of suspected data manipulation. The ubiquity with which we now use GDP to debate China's political economy makes it remarkable that this indicator arrived quite late and with serious challenges.

Top-down coercion is an intuitive explanation for GDP's spread to developing countries and non-capitalist economies, especially since GDP was designed with industrialized market economies in mind, and the statisticians responsible for developing and revising the SNA have struggled for decades to increase its relevance for countries that do not fit this mold. However, this is not what happened in the case of China, as our analysis shows. On the contrary, China's adoption of GDP was driven at every step by domestic actors. While IOs and foreign experts were instrumental in introducing new ideas about national accounting to Chinese statisticians, these instances of international socialization were initiated and controlled by domestic actors.

The diffusion process differs in key ways from existing notions of how norms, ideas or institutions spread across the globe. Whereas existing diffusion literature (as summarized by Zimmerman 2016) expects adoption to increase in a more or less linear fashion – from rhetorical embrace to full implementation – the Chinese case was more experimental. It was a nearly ten-year process of translating international national accounting ideas to the changing domestic political-economic context. The earliest attempts to calculate GDP within the MPS framework only signal a very limited adoption of the indicator. Overcoming these problems by switching to the SNA as the 'appropriate' framework was not an option, because national accounting was subject to ideological constraints. Statistical reforms reflected political debates and compromises over opposing blueprints for economic reform. The result was a hybrid system of national accounting, an experiment based on grafting selected ideas from the SNA onto the MPS framework. The graft that resulted was ideologically balanced but technically unstable. Yet even the pathologies of this experiment did not lead to abandonment of the MPS in favor of the SNA. This happened only when the operational ideology – which set the parameters for economic

reform – progressed enough to openly accept and prioritize the market dynamics that characterize the SNA. And, finally, GDP was only fully implemented when the SNA was in place.

Our findings also speak to IPE literature about the translation and localization of transnational ideas and policies (e.g. Ban 2016; Dafe 2020; Eimer et al. 2016; Röper 2020). Modified or hybridized versions of foreign norms and ideas are often assumed to be the outcomes of localization or translation by domestic actors (e.g. Acharya 2004; Eimer et al. 2016). This case, however, points attention to an unsuccessful case of translation – hybridization (through grafting) was only a temporary outcome that eventually led to full implementation. Through engagement with foreign experts statisticians gained access to relevant knowledge, but politicians attempted to ensure that local ideas were not supplanted by foreign ones. Despite this deliberate attempt at translation, the process still ended in convergence. We identify the relative inflexibility of national accounting practices as a barrier to translation in this case. In addition to the importance of the domestic setting for the process of diffusion and translation, the unique features of national accounting standards undermined domestic attempts to localize foreign national accounts frameworks. Norms, ideas, policies, and institutions are sometimes used interchangeably in IPE diffusion literature. Our findings demonstrate that the object of diffusion – its type and features – is a crucial factor that shapes diffusion mechanisms and outcomes.

Whereas ambiguously defined norms and ideas are highly amenable to interpretation and translation, and policies can be implemented differently across legal systems, international statistical standards are more rigid. National accounting standards in particular are rooted in economic theory and closely linked to a country's economic policy institutions. In China, where ideology clashed with the SNA and another national accounting system was already in place, translation was highly constrained. Although SNA methodology is flexible enough to accommodate different types of national economies, it cannot bend as far as accommodating centrally administered prices or a statistical bureaucracy not equipped to collect survey data on economic activity.

This brings us to the broader implications of this paper for the politics of statistics. GDP's spread to China shows how the choice of indicators and how they are measured is inherently political. Nevertheless, the pathologies that arise from these political choices point out that to understand the spread of international statistical standards, insight into the process and challenges of translating standards into statistical practices is crucial. It shows that not only political interests and ideas matter, but potential clashes with existing statistical institutions and practices as well.

7

Conclusion²⁰

Overview

The dissertation has investigated efforts by the United Nations, the International Monetary Fund, the World Bank, and the experts employed by these IOs to harmonize economic statistics internationally. The efforts of these actors influence perceptions and policies in substantial, but often invisible, ways. On the surface, the motivations for pursuing statistical harmonization are straightforward enough. We need a sense of the unemployment rate, trade balance or growth rate of a country to design specific policies. Whether or not the figures get at the right number, some number is indispensable for policy intervention. And, because they come with a stamp of approval from governments or international organizations, official statistics appear to be authoritative and unbiased. But, as the dissertation has shown, this neutral image masks political struggles and distributive consequences. For one thing, macroeconomic statistics are reductive. Certain economic activities and actors are captured in headline indicators while others are rendered invisible. These indicators, in turn, shape how national economies are made legible to policymakers and market actors (Scott 1998; Broome and Seabrooke 2012). Distortions and reductions are particularly acute in the context of countries outside the industrialized core of the global economy (Morgan 2009; Speich 2008).

20 Parts of this chapter build upon the following online publication: DeRock, D. (2018). The Dangerous Depoliticization of Economic Numbers. The Policy Corner. <https://www.policycorner.org/en/2018/02/26/the-dangerous-depoliticization-of-economic-numbers/>

Taken together, the preceding chapters have demonstrated that statistical harmonization efforts are indeed inherently political, both in terms of their design and their downstream consequences. This concluding chapter synthesizes these political dynamics and their broader implications. The chapter includes a summary of the main findings, the academic relevance, the social and policy implications, and suggestions for further research.

Main findings

The aim of the empirical chapters has been to address the central question and sub-questions – and thus research gaps – as much as possible within the scope of the dissertation. At the center of the dissertation has been the following research question: *What are the drivers of international harmonization of economic statistics?* Put differently, why does statistical harmonization take the form that it does? Above all, the research has found that statistical harmonization is primarily led by experts, and that the insulated nature of expertise-based governance allows for shared norms and ideas to influence the form that harmonization efforts take.

However, there are important caveats to this. First, there are built-in limitations to governance by numbers. While economic statistics are undoubtedly important and useful, there are inevitable tradeoffs and limitations. These inherent features of statistics limit what experts can actually achieve, and also leave them vulnerable to critique when statistics fall short of data users' expectations. Second, inter-organizational dynamics and the autonomy of experts vis-à-vis formal organizations influence different types of statistical governance in different ways. In line with an epistemic communities perspective (Haas 1992), specialized expertise allows statistical standard-setting to be shaped by professional norms. But statistical capacity building and compliance monitoring are, to varying degrees, influenced by the priorities of individual IOs. This finding is in line both with constructivist perspectives and principal-agent perspectives that expect IOs to exert autonomy separate from their member states. The third caveat is that experts are not the only powerful agents in statistical harmonization. The case of GDP in China, at least historically, proves to be an exception to this main finding. In that case study, it is not experts but state officials and policymakers with the most agency over the adoption of international statistical standards. Nonetheless, transnational experts (mainly World Bank staff) and domestic government statisticians did still play important roles in transferring knowledge and norms.

The individual empirical chapters also addressed the sub-questions raised in the introduction of the dissertation. The sub-questions were as follows. First, which factors shape the design of international statistical standards? The creation and revision of standards was addressed mainly in chapters 3 (on the SNA production boundary) and 4

(the statistical trilemma), and also in chapter 6 (the adoption of the SNA in China). Second, what explains the spread of international statistical standards and practices around the world, especially to countries where standards have clashed with local economic conditions? The spread (or diffusion) of statistical practices was addressed mainly in chapter 6, and also in chapter 5 (cooperation between IOs in statistical governance). Third, how are international statistical standards and practices implemented domestically within national statistical systems? Implementation was addressed in the outlying case of China in Chapter 6. Chapter 5 also addressed this sub-question, particularly the findings related to statistical capacity building.

Summary of the findings

Chapter 3 has shown that GDP measurement is governed by a relatively insulated group of economic statisticians. This community of experts is led by the staff of IO statistics departments. Because this governance structure is mostly insulated from external constraints, shared professional norms and economic ideas exert strong influence on the design of international standards. These shared norms and ideas have shown remarkable stickiness on the point of unpaid household services. These services – which correspond to what is often called housework or unpaid care work – continue to be excluded from GDP figures, despite this exclusion resulting in gender bias. I found that experts – IO statisticians working in transnational expert groups – are aware of the criticisms and have a great deal of power to change GDP measurement. Yet, the professional norms and economic ideas that have long prevented this particular change continue to be deeply embedded in this particular community of experts. Moreover, the governance structure is such the standard-setting process is highly impervious to the kinds of serious ideational challenges that would make the inclusion of unpaid household services possible.

Chapter 4 identified built-in limitations of international economic statistics. The chapter introduced the concept of the statistical trilemma. The trilemma is based on the observation that data users demand that economic statistics should (a) be harmonized, (b) be guided by standards prescriptive enough to guarantee reliability and prevent manipulation, and (c) be suited to local socioeconomic contexts. Yet as we show, it is only possible to have two of these at once. This concept explains many of the limitations of international economic statistics, and the wide case selection of six indicators (three of them in greater detail) suggests that the theory can be generalized. As chapter 3 argues, standard-setters, such as members of the ISWGNA, do have agency to revise measurement practices in line with certain stakeholders' preferences. But, the built-in limitations mean that in doing so, some other desiderata will be (partially) sacrificed. This finding suggests that various perceived shortcomings of economic statistics are more deeply rooted than assumed by statisticians and data users. The findings for each case study show how the trilemma works in practice. With GDP, we found that harmonization has been prioritized. Many of the debates around GDP and the SNA have arisen out of unresolved tensions between

prescription and reliability. Indeed, the tensions could only be resolved by abandoning one of these criteria in favor of the other. In the case of unemployment statistics, we found that the trilemma has contributed to statistical fragmentation, wherein there is no single harmonized standard. Poverty statistics, too, are fragmented. Alongside fragmented national practices, the International Poverty Line – while not a standard in the strict sense – trades in the suitability of national poverty statistics for comparable but unsuitable ones.

Chapter 5 has investigated cooperation between the UNSD, the World Bank, and the IMF in global statistical governance. The findings contribute detailed description of statistical governance within and between IOs, a policy domain that has received little attention in IPE literature. The chapter also explains variation between IO cooperation in three forms of governance: statistical standard-setting, capacity building, and compliance monitoring. What I found is that statistical staff of IOs are at times constrained by IO mandates and priorities, and at other times act as autonomous ‘free intellectual agents’ in a global epistemic community. Staff of IO statistics departments are influenced by the professional norms and shared ideas of this epistemic community, but there is variation in the degree to which they are subject to the organizational constraints of IOs. When statistical governance is immediately linked to an IO’s priorities, or viewed as such by IO executives, statisticians are more tightly constrained, and the activities of IOs in a given area of statistical governance more clearly reflect differences between the IOs. Empirically, this leads to variation in the three different governance functions: a high degree of cooperation in standard-setting, moderate cooperation in capacity building, and low cooperation in compliance monitoring.

Finally, the findings of chapter 6 describe and explain China’s late adoption of the SNA (and thus GDP). The chapter identifies the causal mechanisms involved in the diffusion and implementation of national accounting practices within the country’s statistical system. We found that the adoption of the SNA was driven, at every step in a decade-long process, by domestic policymakers. While the World Bank and other external actors encouraged the change, they did not coercively impose it. We describe this in theoretical terms as an ultimately unsuccessful case of ‘translation’ (Ban 2016). China attempted to graft aspects of the SNA onto the national accounting system already in place there, the Marxist-inspired Material Product System (MPS). This grafting was formalized in the Hybrid System of national accounting. While the Hybrid System lasted several years, it was ultimately replaced by the SNA. We find that the grafting and translation attempts failed because the national accounting system is interlinked with a country’s economic structure, policymaking apparatus, and bureaucracy.

Academic relevance

The empirical chapters have made both descriptive and theoretical contributions to the IPE and broader social science literatures. There are many gaps in existing literature, since research on the politics of statistics, although rapidly accumulating, has only started to emerge as an identifiable body of literature in recent years (e.g. Assa & Kvangraven 2021; Fioramonti 2013; Jerven 2013; Mügge 2020b). Therefore, the descriptive findings substantially increase our knowledge of the topic. Many of the descriptive findings will of course be familiar to the economic statisticians working in the field. Even so, the findings take on new significance in the frameworks of political science research.

The theoretical contributions also deepen our understanding both of the topic (statistical harmonization) and the broader IPE and global governance literatures. Theory allows us to see the political underpinnings of statistics and how they are governed, as opposed to viewing them as objective tools with which to observe empirical reality. At the same time, the policy domain of economic statistics introduces new layers to existing theoretical approaches.

The dissertation has contributed to IPE and other social science literature about the politics and sociology of numbers (e.g. Davis, Kingsbury & Merry 2012; Fukuda-Parr, Yamin & Greenstein 2014). From the existing literature, we know that numbers do not objectively capture empirical reality. Regarding economic statistics specifically, we also know quite a bit about the limitations and biases of economic indicators like GDP or FDI figures, as well as the adverse impacts on policymaking (e.g. Hoskyns & Rai 2007; Linsi & Mügge 2019; Yarrow 2020). What we did not know much about is the global governance of these numbers, as well as the causes of the limitations we observe. The previous chapters have addressed these gaps.

Returning to the sub-questions, the dissertation has inquired about the politics of statistical standard-setting, the global spread of statistical standards and practices, and their domestic adoption and implementation. Starting with standard-setting, the dissertation, especially Chapter 3, builds upon literature about biases and shortcomings in GDP, particularly from the perspectives of feminist scholars (Waring 1999). Hoskyns and Rai (2007), for example, argue that this particular statistical blind spot is a problem not only for policymaking, but also for the field of IPE. Yet, we knew much less about why the blind spot exists, and about the actual global governance of GDP measurement more broadly. By giving a detailed look into this governance, Chapter 3 explains why statisticians continue to exclude unpaid household services from the SNA even though they are aware of criticisms. This departs from critiques that imply statisticians are either consciously biased or unaware of the problem. The dissertation also identified a ‘flip side’ to the agency of experts in standard-setting. The concept of the statistical trilemma points to built-in limitations to what economic statistics can achieve. This argument directly

speaks to the existing literature on the politics of statistics. Whereas chapter 3 focuses on agency and ideational limitations, chapter 4 identifies the outer limits of what experts actually can achieve in statistical standards. The trilemma concept also speaks to IPE scholars (and those in any other disciplines) applying economic statistics in analysis. The concepts allows scholars to identify problems of validity in a more precise way.

In addressing the second sub-question, the dissertation has also contributed to academic literature on the behavior of IOs and the power of experts in global governance. This contribution was partly addressed in the chapters about standard-setting discussed above, and also in chapter 5. Chapter 5 has shed light on efforts by IOs to harmonize and govern the measurement, production and dissemination of economic statistics. For one thing, this gives insight into the statistical policies and practices of IOs. I introduced a typology of the main functions of statistical governance – standard-setting, capacity building, and compliance monitoring – that can guide further research on the topic. The concept of bounded autonomy in statistical governance also allows for a more fine-grained analysis of delegation and cooperation in highly technical issue areas.

Another academic contribution is to IR and IPE literature on the diffusion, localization and translation of norms, ideas, policies and institutions (e.g. Ban 2016; Bell & Feng 2019; Eimer et al. 2016; Lai et al. 2017). This contribution touches upon sub-questions 2 and 3, regarding the spread of statistical standards and practices as well as their domestic uptake. Chapter 6 refines existing accounts of diffusion by highlighting the features of the ‘object’ being diffused. Diffusion literature tends to treat norms, policies, ideas and institutions as interchangeable. The SNA is unique as an object of diffusion because national accounting intersects with economic ideas, policies, economic structure, and bureaucratic capacities. In other words, simply adopting or modifying a new system of national accounting is highly complicated, especially if one already exists in a country. These unique features of national accounting also challenge existing theories of localization and translation, which tend to assume modifications of foreign ideas are the end results of localization efforts (Acharya 2004). In this case, modification (in the Hybrid System) was only a temporary outcome that eventually gave way to convergence.

Suggestions for further study

The empirical findings and theoretical arguments of the dissertation point the way toward several additional lines of research. In terms of theory, the dissertation has modified existing theories and introduced new ones. Arguably the most novel theoretical arguments can be found in chapters 4 (the statistical trilemma) and 5 (bounded autonomy). Future research might extend these approaches to new empirical cases. The concept of bounded autonomy was designed to capture unique features of statistical governance.

But, in principle, it should be generalizable to other highly technical issue areas, perhaps with modifications to the theory. This might include topics as diverse as, say, artificial intelligence, biotechnology regulation, or product safety standards. The concept of the statistical trilemma could be extended first to other types of economic statistics, but also (possibly with modifications) to non-economic numbers. This might include, for instance, demographic statistics or carbon accounting practices. Future research about the politics of statistics can also build upon the trilemma concept to go beyond critiquing statistics, and instead ask why – within built-in limitations – standard-setters have made the conceptual choices that they have.

I conclude the dissertation with a number of specific research questions that could be taken up in further research. The first relates to big data, a realm of quantification that overlaps somewhat with official statistics but has many unique features. There are many ways that the approach of the dissertation could be extended to big data, but one question in particular strikes me as important: *To what extent can big data circumvent statistical capacity problems?* Low statistical capacity is a notoriously persistent problem, as it is rooted in deeper issues of state capacity and fiscal limitations in low-income countries. Big data promises to get around at least some of these constraints, since economic data can be aggregated from mobile phone financial transactions, satellite imaging, e-commerce data scraping, and other such sources. There are both practical and ethical concerns around this agenda that deserve academic scrutiny.

Why do countries participate in statistical harmonization efforts? The dissertation has focused on IO efforts to harmonize economic statistics, with less attention to domestic agency. Chapter 6 is a major exception, but we can expect China to be an atypical case in this respect. Why do countries accept, or even invite, external interventions in their statistical practices? There may be functional pressures, socialization and reputational dynamics, coercion, or other explanations. At the same time, this line of research could explore whether and why countries sometimes act as norm entrepreneurs, such as the case of Bhutan's use of Gross National Happiness (GNH) alongside GDP.

To what extent do nationalist and anti-expert sentiments affect the perceived authority and legitimacy of official statistics? This line of research would engage with political science literature on populism, anti-elite and anti-expert sentiments. The electoral success of nationalist and (at least rhetorically) populist leaders in recent years has challenged multilateralism and international cooperation. Donald Trump, for instance, painted the World Health Organization as incompatible with national interests and announced the withdrawal of the United States. We also know that citizens often question the neutrality of their own governments' official data, whether economic statistics or data on COVID-19 or migration. Do these attitudes extend to the statistics published by IOs? If so, what are the consequences for IO-member state relations?

Political theorists might also take up some questions outside the scope of empirical political science. One such set of questions revolves around GDP and proposed alternatives. Alternatives range from other indicators like GNH, to degrowth (or post-growth) as a fundamental challenge to the economic growth paradigm. These movements beg the questions: *What alternatives to GDP are possible and desirable? What is the relation of economic measurement to economic structure?* The world has seen an alternative to the System of National Accounts (and thus GDP) with the Material Product System (MPS). But the USSR and China hardly represent the only imaginable alternatives to liberal market economies, and the MPS is certainly not the only possible national accounting alternative. Polanyi's (2016) article "Socialist Accounting" proposed a third alternative rooted in market socialism, but remains hypothetical. These theoretical questions matter because the discourse around the shortcomings of GDP and other indicators assume that changing economic measurement can change society, presumably for the better. Yet, it is not clear whether targeting indicators is the most effective approach to social change. If indicators simply reflect economic structures, then the structures themselves need to be targeted. But, if indicators truly have a performative or reflexive relationship with economic structure, then it is indeed worthwhile to target them. These are questions best suited to a philosophical or interdisciplinary approach.

Social relevance and policy implications

Now more than ever, statistics take a central role in economic policy and international development agendas. At the same time, the promises of data as a tool for development are outpacing the ability to collect reliable statistics in many parts of the world. The Sustainable Development Goals, for example, greatly increase the demand for wide-ranging economic, social and environmental data. With these practical concerns in mind, the dissertation has interrogated the harmonization attempts of the UNSD, the World Bank, and the IMF. After all, whatever we know about global and national economies is ultimately influenced by these understudied political processes. Before the 1940s, internationally comparable statistics on national income, debt, unemployment, and so on simply did not exist. Nowadays, they are mainstays of news headlines, policy agendas and political campaigns. Thus, there are important societal and policy-related takeaways from the research.

Taken together, the chapters shed light on several political features of statistics. The measurement of statistics determines what gets counted and how. The production boundary of the SNA, addressed in chapters 3 and 4, is a good example. It is therefore important for policymakers and analysts to be aware of the ways that numbers frame economic objects. The theoretical arguments about standard-setting have implications for potential changes in the ways indicators are measured and applied. Chapters 3 and 4 emphasize agency and

constraints, respectively, in standard-setting. Both chapters suggest that there are limits to fundamental changes to indicators like GDP. One limitation comes from ideational path-dependency among standard-setters, and the other set of limitations comes from the constraints of the statistical trilemma.

The choice of one indicator over another also influences perceptions and policies. The case of China (chapter 6) makes it clear that statistical practices are intricately linked with economic structure, institutions, and ideologies. Even when choosing an indicator does not involve uprooting an entire national accounting system, there may still be practical consequences. For example, countries with low statistical capacity face strong external pressure to produce required economic and financial sector data (chapter 5). When staff numbers and budgets are small, that data production can be at the expense of collecting other types of data (sociodemographic data, for example) that may be more relevant to local priorities. Therefore, efforts to increase statistical capacity and data quality should be more custom-tailored to domestic needs.

With that said, the biggest societal implication is not a specific policy recommendation, but a matter of how we think about statistics. In practical terms, there needs to be greater awareness among analysts and policymakers of the conceptual biases built into economic indicators. This can start with incorporating a socially embedded perspective of statistics into economics and social science curricula. Most urgently, this embedded perspective should inform development policy at the level of international organizations and national governments.

In recent years, there has been a visible backlash against expertise in many forms, including against official statistics. This is often associated with populist attitudes, but not exclusively. Examples related to statistics include citizens' distrust in immigration data, suspicions of governments manipulating GDP and unemployment figures, and widespread skepticism and confusion surrounding COVID-19 infection and mortality rates. This trend is not confined to right-wing ideologies or conspiracy theories. GDP is a good example. Rather than simply claiming that GDP figures are made up, most critics of GDP highlight reasonable problems such as economic growth at the cost of the environment or of well-being. The dissertation has shed light on certain aspects of these trends.

The findings suggest that the production of official statistics has indeed been mostly insulated from stakeholder participation beyond a small community of experts. This was possible because of the specialized knowledge required. However, as Aragão and Linsi (2020) argue, it is not outright manipulation that critics should be concerned about. Rather, it is standard-setting where experts have the most power to shape how our economies are measured. Thus, the most important choices with distributive implications come before raw data is collected, not after. Nor is it true that these choices in standard-setting are

malicious or intended to disenfranchise certain groups. Rather, most shortcomings in economic statistics result from the banal influence of economic theories and professional norms about statistics, including statisticians' ambitions for comparability, utility for policymaking and analysis, and prescriptiveness of uniform standards.

The fact that the distributive implications of official statistics are now debated publicly is partly a response to high-profile critiques, such as the Report by the Commission on the Measurement of Economic Performance and Social Progress (Stiglitz, Sen, and Fitoussi 2010). It also results from a growing recognition in the development community and among policymakers that a wider range of data is required to meet contemporary challenges such as poverty, inequality, gender disparities, climate change, and biodiversity loss. The major increase in the number of goals and indicators from the MDGs to the SDGs is evidence of this recognition. Yet, the statistics we have at our disposal are still inadequate to properly plan for, and monitor progress on, these challenges. At the same time, abandoning economic and social measurement would leave us far worse off. Instead, there should be wider recognition of what statistics can and cannot do. The concept of the statistical trilemma, presented in Chapter 4, offers some guidelines for how data users and analysts can be more cognizant of that.

We should also continuously try to improve our measurement tools. Bringing statistics out into public and academic discussion should have the aim of offering suggestions rather than simply disparaging statistics and statisticians. There is no shadowy elite manipulating statistics behind a curtain, as some conspiracy theorists would have us believe. But it is also true that important stakeholders have long been denied a seat at the table. For example, while NGOs are invited to the UN Statistical Commission meetings, their influence is limited. One possible solution is to create dedicated forums for representatives of marginalized groups – whether that be precarious workers, indigenous peoples, or victims of systemic racism – to enter into dialogue with statistical agencies about what is important for them. Other social scientists and academics also have a role to play in thinking about how to set and monitor global goals in creative ways beyond our existing statistical toolkit. As an example, coupling the SDG indicators with ethnographic research or interviews could help us understand what a one percent increase or decrease in, say, poverty or education outcomes actually means for real people.

Bibliography

- Abbott, K., & Snidal, D. (1998). Why states act through formal international organizations. *Journal of Conflict Resolution*, 42(1), 3–32.
- Abbott, K. and Snidal, D. (2001). ‘International “standards” and international governance’, *Journal of European Public Policy* 8(3): 345–70.
- Acharya, A. (2004). How Ideas Spread: Whose Norms Matter? Norm Localization and Institutional Change in Asian Regionalism. *International Organization*, 58(2), 239–275.
- Acharya, A. (2016). Idea-shift: how ideas from the rest are reshaping global order. *Third World Quarterly*, 37 (7), 1156–70. doi: 10.1080/01436597.2016.1154433
- Acharya, A. (2004). How Ideas Spread: Whose Norms Matter? Norm Localization and Institutional Change in Asian Regionalism. *International Organization*, 58(2), 239–275.
- Ady, P. (1962). Use of National Accounts in Africa. *Review of Income and Wealth*, 1, 52–65.
- Alenda-Demoutiez, J., & Mügge, D. (2019). The lure of ill-fitting unemployment statistics: How South Africa’s discouraged work seekers disappeared from the unemployment rate. *New Political Economy*. Advance online publication. doi.org/10.1080/13563467.2019.1613355
- Allan, B. (2018). From subjects to objects: knowledge in international relations theory. *European Journal of International Relations*, 24 (4), 841–64.
- Alonso, W. and Starr, S. (eds) (1987). *The Politics of Numbers*, New York: Russell Sage Foundation.
- Alenda-Demoutiez, J., & Mügge, D. (2019). The lure of ill-fitting unemployment statistics: How South Africa’s discouraged work seekers disappeared from the unemployment rate. *New Political Economy*. Advance online publication. doi.org/10.1080/13563467.2019.1613355
- Ambrosio, T. (2012). The rise of the ‘China Model’ and ‘Beijing Consensus’: evidence of authoritarian diffusion?. *Contemporary Politics*, 18(4), 381–399.
- Appadurai, A. (1996) *Modernity at Large. Cultural Dimensions of Globalization*. Minneapolis: University of Minnesota Press.
- Aragão, R. & Linsi, L. (2020) Many shades of wrong: what governments do when they manipulate statistics, *Review of International Political Economy*, DOI: 10.1080/09692290.2020.1769704
- Árvay, J. (1971). The new system of national accounting in Hungary. *Acta Oeconomica*, 7(2), 161–176.

- Árvey, J. (1994). The Material Product System. In E. Kenessey (Ed.), *The Accounts of Nations* (pp. 218-236). Amsterdam: IOS Press.
- Assa, J. & Kvangraven, I.H. (2021). Imputing Away the Ladder: Implications of Changes in GDP Measurement for Convergence Debates and the Political Economy of Development, *New Political Economy*,.
- Ban, C. (2013). Brazil's liberal neo-developmentalism: New paradigm or edited orthodoxy? *Review of International Political Economy*, 20(2), 298–331.
- Ban, C. (2015). Austerity versus stimulus? Understanding fiscal policy change at the International Monetary Fund since the great recession. *Governance*, 28 (2), 167–83.
- Ban, C. (2016). *Ruling ideas: How global neoliberalism goes local*. New York, NY: Oxford University Press.
- Ban, C. and Patenaude, B. (2018). The professional politics of the austerity debate: a comparative field analysis of the European Central Bank and the International Monetary Fund. *Public Administration*, 2019, 1–16.
- Bardasi, E. and Wodon, Q., (2010). Working long hours and having no choice: time poverty in Guinea. *Feminist Economics*, 16 (3), 45–78.
- Barnett, M., & Finnemore, M. (1999). The politics, power, and pathologies of international organizations. *International Organization*, 53(4), 699–732.
- Barnett, M. and Finnemore, M. (2004). *Rules for the world: international organizations in global politics*. Ithaca: Cornell University Press.
- Baxandall, P. (2004). *Constructing unemployment: the politics of joblessness in east and west*. Hoboken, NJ: John Wiley & Sons.
- Bean, C. (2016) Independent Review of UK Economic Statistics. London. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf.
- Bell, S., & Feng, H. (2019). Policy diffusion as empowerment: domestic agency and the institutional dynamics of monetary policy diffusion in China. *Globalizations*, 1-15.
- Benería, L. (1992). Accounting for women's work: the progress of two decades. *World Development*, 20 (11), 1547–60.
- Benham, F. (1953). Income and product in under-developed countries: comments on the paper by Professor Frankel. *Review of Income and Wealth*, 3 (1), 169–77.
- Benjamin, R. (2019) *Race after Technology*. Cambridge: Polity Press.
- Best, J. (2012). Bureaucratic ambiguity. *Economy and Society*, 41(1), 84 –106.
- Best, J. (2017). 'The Rise of Measurement-Driven Governance: the Case of International Development', *Global Governance*, 23(2), pp. 163–181.

- Biermann, F., Pattberg, P., van Asselt, H., & Zelli, F. (2009). The fragmentation of global governance architectures: A framework for analysis. *Global Environmental Politics*, 9(4), 14–40.
- Blades, D. (1975). *Non-monetary (subsistence) activities in the national accounts of developing countries*. Paris: OECD.
- Bloch, D. and Fall, F. (2015). *Government Debt Indicators: Understanding the Data*. OECD Economics Department Working Papers 1228. Paris
- Börzel, T. A., & Risse, T. (2012). From Europeanisation to diffusion: introduction. *West European Politics*, 35(1), 1-19.
- Bos, F. (1995). *Economic Theory and National Accounting*. Statistics Netherlands Working Paper, NA-075). Voorburg, Statistics Netherlands.
- Bos, F. (2009). *The national accounts as a tool for analysis and policy: history, economic theory and data compilation issues* [online]. **Munich Personal** RePEc Archive. Available from: mpra.ub.uni-muenchen.de/23582/.
- Bottelier, P. (2007). China and the World Bank: how a partnership was built. *Journal of Contemporary China*, 16 (51), 239-258.
- Broad, R. (2006). Research, knowledge, and the art of ‘paradigm maintenance’: the World Bank’s development economics Vice-Presidency (DEC). *Review of International Political Economy*, 13 (3), 387–419.
- Breslau, D. (2003). Economics invents the economy: mathematics, statistics, and models in the work of Irving Fisher and Wesley Mitchell. *Theory and Society*, 32, 379–411.
- Brødsgaard, K. E., & Rutten, K. (2017). *From Accelerated Accumulation to Socialist Market Economy in China: Economic Discourse and Development from 1953 to the Present*. Leiden: Brill.
- 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., and Quirk, J. (2015). Governing the world at a distance: the practice of global benchmarking. *Review of International Studies*, 41: 819-841.
- Broome, A. and Seabrooke, L. (2007). Seeing like the IMF: Institutional change in small open economies. *Review of International Political Economy*, 14(4): 576–601.
- Broome, A., & Seabrooke, L. (2012). Seeing like an international organisation. *New Political Economy*, 17(1), 1–16.
- Broussard, M. (2018). *Artificial Unintelligence. How Computers Misunderstand the World*. Cambridge MA: MIT Press.

- Broussolle, D. (2015). The emerging new concepts of services in the system of national accounts and the balance of payments. *Review of Income and Wealth*, 61 (3), 573–86.
- Busuioac, E. M. (2016). Friend or foe? Inter-agency cooperation, organizational reputation, and turf. *Public Administration*, 94(1), 40–56.
- Camargo, A. de P.R. (2009). ‘The sociology of statistics: the possibilities of a new field of investigation’, *História, Ciências, Saúde-Manguinhos*, 16(4): 903–25.
- Carroll, P. (2006). *Science, Culture, and Modern State Formation*. Berkeley: University of California Press.
- Chadeau, A. (1992). What is households’ non-market production worth? *OECD Economic Studies*, 18, 85–103.
- Chander, R. (2014). *China’s Statistical System. Three Decades of Change and Transformation*. Kuala Lumpur: University of Malaya Press.
- Chen, F. (1995). *Economic transition and political legitimacy in post-Mao China: Ideology and reform*. SUNY Press.
- Christophers, B. (2011). Making finance productive. *Economy and Society*, 40 (1), 112–40.
- Christophers, B. (2013). *Banking across boundaries*. Hoboken: Wiley-Blackwell.
- Chwieroth, J. (2013). The silent revolution: How the staff exercise informal governance over IMF lending. *Review of International Organizations*, 8, 265–290.
- Chwieroth, J. (2014). Fashions and fads in finance: The political foundations of sovereign wealth fund creation. *International Studies Quarterly*, 58(4), 752–63.
- Clegg, L. (2010). Our dream is a world full of poverty indicators: The US, the World Bank, and the power of numbers. *New Political Economy*, 15(4), 473–492.
- Cobb, C. and Cobb, J. (1994). *The Green National Product*. Lanham: University Press of America.
- Cooley, A. and Snyder, J. (eds) (2015). *Ranking the World: Grading States as a Tool of Global Governance*, New York: Cambridge University Press.
- Copelovitch, M. (2010). Master or servant? Common agency and the political economy of IMF lending. *International Studies Quarterly*, 54, 49–77.
- Cottarelli, C. (2017). *What We Owe: Truths, Myths and Lies about Public Debt*. Washington DC: Brookings Institution Press.
- Coyle, D. (2014). *GDP: A Brief but Affectionate History*, Princeton, NJ: Princeton University Press.

- Coyle, D. (2017). Do-it-yourself digital: the production boundary and the productivity puzzle. ESCoE Discussion Paper 2017-01 (June), Economic Statistics Centre of Excellence.
- Coyle, D. & Nguyen, D. (2020). No plant, no problem? Factoryless manufacturing, economic measurement and national manufacturing policies, *Review of International Political Economy*. DOI: 10.1080/09692290.2020.1778502
- Dafe, F. (2019). Ambiguity in international finance and the spread of financial norms: the localization of financial inclusion in Kenya and Nigeria. *Review of International Political Economy*. DOI: 10.1080/09692290.2019.1650093
- Damgaard, J. and 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
- Damgaard, J. and Elkjaer, T. (2017). The Global FDI Network: Searching for Ultimate Investors. IMF Working Paper WP/17/258. Washington, DC: International Monetary Fund.
- Damgaard, J., Elkjaer, T. and Johannesen, N. (2019). The rise of phantom investments. *Finance & Development* 56(3): 11–13.
- Dargent, E., Lotta, G., Mejía, J.A., and Moncada, G. (2018). *Who wants to know? The political economy of statistical capacity in Latin America*. Washington, DC: Inter-American Development Bank.
- Data Revolution Group (2014). *A World That Counts: Mobilising the Data Revolution for Sustainable Development*. New York: United Nations. Available online: <https://www.undatarevolution.org/report/>
- Davis, K., Kingsbury, B. and Merry, S. (2012). Indicators as a technology of global governance. *Law & Society Review*, 46 (1), 71–104.
- De Las Casas, M. (2016). The rules of the game: Data-related mandate, obligations, and practices at the IMF (IEO Background Paper BP/16/01). Washington, DC: Independent Evaluation Office of the IMF.
- De Resende, C., and Loyola, F. (2016). On the Effect of IMF Data Standards Initiatives: Do They Affect Foreign Direct Investment, Exchange Rate Volatility, and Sovereign Borrowing Costs? (IEO Background Paper BP/16/05). Washington, DC: Independent Evaluation Office of the IMF.
- Dean, K., Joseph, J., Roberts, J. and Wight, C. (2006). *Realism, Philosophy and Social Science*. New York: Palgrave Macmillan.
- Deaton, A. (2010). Price Indexes, Inequality, and the Measurement of World Poverty. *American Economic Review*, 100(1): 5-34.

- Demombynes, G., and Vu, L. H. (2015). Demystifying Poverty Measurement in Vietnam. Vietnam Development Economics Discussion Papers. 1 (February). World Bank Vietnam Country Office.
- Deng, X. (1979). "China's Goal is to Achieve Comparative Prosperity by the End of the Century". December 6, 1979. In: *Selected Works of Deng Xiaoping, Volume II: 1975-1982*; (1984) Beijing: Foreign Languages Press. Retrieved from <https://dengxiaopingworks.wordpress.com/2013/02/25/chinas-goal-is-to-achieve-comparative-prosperity-by-the-end-of-the-century/>.
- DeRock, D. (2021). Hidden in Plain Sight: Unpaid Household Services and the Politics of GDP Measurement, *New Political Economy*, 26:1, 20-35.
- Dersnah, M. (2019). United Nations gender experts and the push to focus on conflict-related sexual violence. *European Journal of Politics and Gender*, 2 (1), 41–56.
- Desrosières, A. (1993). *La politique des grands nombres: Histoire de la raison statistique*. Paris: Éditions La Découverte.
- Desrosières, A. (2000). Measurement and its uses: harmonization and quality in social statistics. *International Statistical Institute*, 68 (2), 173–87.
- Desrosières, A. (2001). How 'Real' Are Statistics? Four Possible Attitudes. *Social Research*, 68, pp. 339-355.
- Desrosières, A. (2014). 'La statistique, outil de libération ou outil de pouvoir?', in Bruno, I., Didier, E., and Prévieux, J. (eds) *Statactivisme*. Paris: Zones. La Découverte, pp. 59–68.
- Devarajan, S. (2013). Africa's statistical tragedy. *Review of Income and Wealth*, 59, 9–15.
- DiMaggio, P. and Powell, W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48(2): 147-160.
- Dominguez, L. (1947). National income estimates of Latin-American countries. In: *Conference on research in income and wealth, studies in income and wealth*. Cambridge, MA: National Bureau of Economic Research, 160–244.
- Edward, P. (2006). The Ethical Poverty Line: A Moral Quantification of Absolute Poverty. *Third World Quarterly*, (37)2: 377–393.
- Eele, G. (2015). Building Statistical Capacity: The challenges. PARIS21. Discussion Paper No. 7 July 2015.
- Eimer, T. R., Lütz, S., & Schüren, V. (2016). Varieties of localization: International norms and the commodification of knowledge in India and Brazil. *Review of International Political Economy*, 23(3), 450–479.

- Elsig, M. (2011). Principal–agent theory and the World Trade Organization: Complex agency and ‘missing delegation’. *European Journal of International Relations*, 17(3), 495–517.
- Elson, D. (2005). Unpaid work, the Millennium Development Goals, and capital accumulation: notes for a presentation. *Conference on Unpaid Work and the Economy: Gender, Poverty and the Millennium Development Goals*, 1–3 October 2005 Annandale-on-Hudson, 1–12.
- Enns, C. (2015). Knowledges in competition: knowledge discourse at the World Bank during the knowledge for development era. *Global Social Policy: An Interdisciplinary Journal of Public Policy and Social Development*, 15 (1), 61–80.
- Espeland, W., and Stevens, M. (2008). A Sociology of Quantification. *European Journal of Sociology*, 48, 401–436.
- Esquivel, V. (2011). Sixteen years after Beijing: what are the new policy agendas for time-use data collection? *Feminist Economics*, 17 (4), 215–38.
- Eurostat (2014). *Essential SNA: Building the basics*. Luxembourg: Publications Office of the European Union.
- Ferdinand, P., & Wang, J. (2013). China and the IMF: from mimicry towards pragmatic international institutional pluralism. *International Affairs*, 89(4), 895–910.
- Ferguson, J. (1990). *The Anti-Politics Machine: ‘Development’, Depoliticization and Bureaucratic Power in Lesotho*. Cambridge: Cambridge University Press.
- Ferreira, F., Chen, S., Dabalén, A., Dihkanov, Y., Hamadeh, N., Jolliffe, D., Narayan, A., Prydz, E., Revenga, A., Sangraula, P., Serajuddin, U., Yoshida, N. (2016). A Global Count of the Extreme Poor in 2012: Data Issues, Methodology and Initial Results. Policy Research Working Paper Series, no. 7432, Washington, DC: World Bank.
- Finér, L. and Ylönen, M. (2017). Tax-driven wealth chains: a multiple case study of tax avoidance in the Finnish mining sector. *Critical Perspectives on Accounting* 48(October 2017): 53–81.
- Finnemore, M. & Sikkink, K. (1998). International norm dynamics and political change. *International Organization*, 52(4): 887–917.
- Fioramonti, L. (2013). *Gross Domestic Problem: The Politics behind the World’s Most Powerful Number*. London: Zed Books.
- Fioramonti, L. (2017). *The world after GDP*. Cambridge: Polity Press.
- Fichtner, J. (2017). Perpetual decline or persistent dominance? Uncovering Anglo-America’s true structural power in global finance. *Review of International Studies* 43(1): 3–28.

- Fletcher, A. (2017). Applying critical realism in qualitative research: methodology meets method. *International Journal of Social Research Methodology*, 20(2), 181-194.
- Folbre, N. (2009). *Greed, lust, and gender: a history of economic ideas*. New York: Oxford University Press.
- Folbre, N. (2014). The care economy in Africa: subsistence production and unpaid care. *Journal of African Economies*, 23, i128–56.
- Fougner, T. (2008). 'Neoliberal Governance of States: the Role of Competitiveness Indexing and Country Benchmarking', *Millennium*, 37(2): 303-326.
- Frankel, S., (1953). Concepts of income and welfare in advanced and under-developed societies with special reference to the intercomparability of national income estimates. *Review of Income and Wealth*, 3, 156–68.
- Freistein, K. (2016). Effects of indicator use: A comparison of poverty measuring instruments at the World Bank. *Journal of Comparative Policy Analysis*, (18)4, 366–381.
- Fukuda-Parr, S., Yamin, A. and Greenstein, J. (2014). The Power of Numbers: A Critical Review of Millennium Development Goal Targets for Human Development and Human Rights. *Journal of Human Development and Capabilities*, 15:2-3, 105-117.
- Garrett, G. (1992). International Cooperation and Institutional Choice: The European Community's Internal Market. *International Organization* 46 (spring), 533–60.
- Gehring, T. and Faude, B. (2014). A Theory of Emerging Order within Institutional Complexes: How Competition among Regulatory International Institutions Leads to Institutional Adaptation and Division of Labor', *The Review of International Organizations*, 9(4), pp. 471–498.
- Gewirtz, J. (2017). *Unlikely Partners. Chinese Reformers, Western Economists, and the Making of Global China*. Cambridge (MA): Harvard University Press.
- Gewirtz, J. (2020). 'Loving Capitalism Disease': Aids and Ideology in the People's Republic of China, 1984–2000. *Past & Present*. Advance online publication. <https://doi.org/10.1093/pastj/gtz068>.
- Giovannini, E. (2008). *Understanding Economic Statistics: An OECD perspective*. Paris: OECD.
- Goetz, A.M. (1997). *Getting institutions right for women*. London: Sage.
- Guo, J., Sonis, M. & Hewings, G.J.D. (1999). An Analysis of Internal and External Linkages of Manufacturing and Non-Manufacturing Industries: Application to Chinese Metropolitan Economies. In: Geoffrey J.D. Hewings, Michael Sonis, Moss Madden & Yoshio Kimura (Eds.), *Understanding and Interpreting Economic Structure* (pp. 317-346). Berlin: Springer.

- Gutner, T. (2010). When 'doing good' does not: The IMF and the Millennium Development Goals. In Avant, D., Finnemore, M., Sell, S. (Eds.), *Who Governs the Globe?* (pp. 266–291). Cambridge: Cambridge University Press.
- Haas, P. (1992). Introduction: epistemic communities and international policy coordination. *International Organization*, 46 (1), 1–35.
- Hacking, I. (1981) 'How should we do the history of statistics?' *I & C.*, 8(April): 15–26.
- Hafner-Burton, E., & Tsutsui, K. (2005). Human Rights in a Globalizing World: The Paradox of Empty Promises. *American Journal of Sociology* 110: 1373–1411.
- Hall, P. and Taylor, R. (1996). Political science and the three new institutionalisms. *Political Studies*, 44 (5), 952–73.
- Hamilton, C. (2003). *Growth Fetish*. London: Pluto Press.
- Hansen, H. and Porter, T. (2012). 'What do numbers do in transnational governance?', *International Political Sociology*, 6(4): 409–26.
- Harper, R. H. R. (1998). *Inside the IMF: An Ethnography of Documents, Technology and Organisational Action*. San Diego, CA: Academic Press.
- Harrison, A. (ed.) (2005). *The background to the 1993 revision of the System of National Accounts* [online]. Available from: unstats.un.org/unsd/sna1993/history/backgrd.pdf.
- Harrison, A. (2017). W(h)ither the SNA? *Review of Income and Wealth*. 63(2): S208–S222.
- Hawkins, D., Lake, D., Nielson, D., & Tierney, M. (2006). *Delegation and Agency in International Organizations*. Cambridge: Cambridge University Press.
- Hay, C. (2006). Constructivist institutionalism. In: R. Rhodes, S. Binder, and B Rockman, ed. *The Oxford Handbook of political institutions*. Oxford: Oxford University Press, 56–74.
- Hay, C. (2016). Good in a crisis: the ontological institutionalism of social constructivism. *New Political Economy*, 21 (6), 520–35.
- van Heijster, J. (2020). Imagining Modernization: The Symbolic Embrace of GDP in China. *Chinese Political Science Review*, 5(1), 50-73.
- van Heijster, J. & DeRock, D. (2020). How GDP spread to China: the experimental diffusion of macroeconomic measurement. *Review of International Political Economy*, DOI: 10.1080/09692290.2020.1835690
- Heilmann, S., & Schulte-Kulmann, N. (2011). The Limits of Policy Diffusion: Introducing International Norms of Anti-Money Laundering into China's Legal System. *Governance*, 24(4), 639-664.

- Herrera, Y. (2010). *Mirrors of the Economy. National Accounts and International Norms in Russia and Beyond*. Ithaca, NY: Cornell University Press.
- Herrera, Y., and Kapur, D. (2007). Improving data quality: Actors, incentives, and capabilities. *Political Analysis*, 15(04), 365–386.
- Hickel, J. (2016). The true extent of global poverty and hunger: questioning the good news narrative of the Millennium Development Goals. *Third World Quarterly*, 37:5, 749–767.
- Hjertholm, P. (2003) ‘Theoretical and empirical foundations of HIPC debt sustainability targets’, *Journal of Development Studies*. Taylor & Francis, 39(6), pp. 67–100.
- Hoekstra, R. (2019). *Replacing GDP by 2030: towards a common language for the well-being and sustainability community*. Cambridge: Cambridge University Press.
- Holz, C. (2004). China’s Statistical System in Transition: Challenges, Data Problems, and Institutional Innovations. *Review of Income and Wealth*, 50(3), 381–409.
- Hopewell, K. (2015). Different paths to power: The rise of Brazil, India and China at the World Trade Organization. *Review of international political economy*, 22(2), 311–338.
- Hoskyns, C. and Rai, S. (2007). ‘Recasting the Global Political Economy: Counting Women’s Unpaid Work’, *New Political Economy*, 12(3), pp. 297–317.
- Husmanns, R. (2004). *Defining and Measuring Informal Employment*. Bureau of Statistics, International Labour Office: Geneva.
- Ikenberry, J. (1992). ‘A world economy restored: consensus and the Anglo-American postwar settlement’, *International Organization*, 46(1), 289–321.
- ILO (1993). Resolution concerning statistics of employment in the informal sector, adopted by the Fifteenth International Conference of Labour Statisticians (January 1993). Available online: http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_087484.pdf
- ILO (2013a). *Measuring informality: a statistical manual on the informal sector and informal employment*. Geneva: ILO.
- ILO (2013b). Measurement of the Informal Economy. Chapter in *The Informal Economy and Decent Work: A Policy Resource Guide Supporting Transitions to Formality*. Available online: http://www.ilo.org/emppolicy/pubs/WCMS_212688/lang--en/index.htm
- ILO (2016). *Women at work: trends 2016*. Geneva: ILO.
- ILO. (2018) *Care work and care jobs for the future of decent work*. Geneva: ILO.

- IMF (n.d.) *Guide for participants and users: enhanced general data dissemination system*. Accessed online: https://dsbb.imf.org/content/pdfs/eGDDS_Guide_for_Participants_and_Users.pdf
- IMF (2012). Press Release: Statement by the IMF Executive Board on Argentina. Press Release No. 12/30, February 1, 2012. Washington, DC: International Monetary Fund. Available online: <https://www.imf.org/en/News/Articles/2015/09/14/01/49/pr1230>
- IMF (2013). Press Release: Statement by the IMF Executive Board on Argentina. Press Release No. 13/33, February 1, 2013. Washington, DC: International Monetary Fund. Available online: <https://www.imf.org/en/News/Articles/2015/09/14/01/49/pr1333>
- IMF (2014). *Government Finance Statistics Manual 2014*, Washington, DC: International Monetary Fund.
- IMF (2016). *Methodology in STA over the past 25 years*. Washington, DC: International Monetary Fund.
- IMF (2016). IMF Executive Board Removes Declaration of Censure on Argentina. Press Release No. 16/497, November 9, 2016. Washington, DC: International Monetary Fund. Available online: <https://www.imf.org/en/News/Articles/2016/11/09/PR16497-Arentina-IMF-Executive-Board-Removes-Declaration-of-Censure>
- IMF (2018a). 2018 Statistics Department at a Glance. Washington, DC: International Monetary Fund. Available online: <https://www.imf.org/en/Data>
- IMF (2018b). Overarching strategy on data and statistics at the Fund in the digital age. Washington, DC: International Monetary Fund. Available online: <https://www.imf.org/en/Publications/Policy-Papers/Issues/2018/03/20/pp020918imf-executive-board-supports-new-strategy-for-data-and-statistics-in-the-digital-age>
- IMF (2019a). 2019 Statistics Department at a Glance. Washington, DC: International Monetary Fund. Available online: <https://www.imf.org/en/Data>
- IMF (2019b). IMF Standards for Data Dissemination. Retrieved from: <https://www.imf.org/en/About/Factsheets/Sheets/2016/07/27/15/45/Standards-for-Data-Dissemination>
- IMF (2020). Data Dissemination Standards Bulletin Board. Retrieved from: <https://dsbb.imf.org>
- ISWGNA (1993). *System of national accounts 1993*. Washington, DC: Inter-Secretariat Working Group on National Accounts.
- ISWGNA (2008). *System of national accounts 2008*. Washington, DC: Inter-Secretariat Working Group on National Accounts.

- ISWGNA (2011). Guidelines for monitoring the 2008 SNA implementation. Extract of the report of the Intersecretariat Working Group on National Accounts to the forty second session (2011) of the United Nations Statistical Commission. Retrieved online: <https://unstats.un.org/unsd/nationalaccount/docs/2011Report.pdf>
- ISWGNA (2018a). Report of the intersecretariat working group on National Accounts. Statistical Commission Items for Discussion and Decision, E/CN.3/201.
- ISWGNA (2018b). ISWGNA Task Force on SNA Capacity Building. 12th Meeting of the Advisory Expert Group on National Accounts, 27-29 November 2018, Luxembourg. Agenda item: 7.a. Retrieved online: https://unstats.un.org/unsd/nationalaccount/aeg/2018/M12_7a_Capacity_Building.pdf
- Jefferies, W. (2015). *Measuring National Income in the Centrally Planned Economies: Why the West Underestimated the Transition to Capitalism*. Abingon: Routeledge.
- Jeffs, J. (2008). The politics of financial plumbing: Harmonization and interests in the construction of the international payment system, *Review of International Political Economy*, 15:2, 259-288.
- Jerven, M. (2013). *Poor Numbers. How We Are Misled by African Development Statistics and What to Do about It*. Ithaca: Cornell University Press.
- Jetschke, A., & Rüländ, J. (2009). Decoupling Rhetoric and Practice: The Cultural Limits of ASEAN Cooperation. *The Pacific Review* 22: 179-203.
- Jiang, P. (2002). 国民经济核算与政府统计改革 [National Economic Accounting and Government Statistics Reform]. *统计研究 [Statistical Research]*, 8, 28-31.
- Jolliffe, D., & Beer, E. (2016). Estimating International Poverty Lines from Comparable National Thresholds. World Bank Policy Research Working Paper 7606. Development Research Group. March.
- Jonsson, C. (1986). Interorganization Theory and International Organization. *International Studies Quarterly*, (30)1: 39-57.
- Kapstein, E. (1989). Resolving the Regulator's Dilemma: International Coordination of Banking Regulations. *International Organization* 43 (spring): 323-47.
- Kapto, S. (2019). Layers of politics and power struggles in the SDG indicators process. *Global Policy*, 10, 134-136.
- Kelley, J. and Simmons, B. (2015). 'Politics by number. Indicators as social pressure in international relations', *American Journal of Political Science* 59(1): 55-70.
- Kendrick, J. (1970). The historical development of national-income accounts. *History of Political Economy*, 2 (2), 284-315.

- Kentikelenis, A. and Seabrooke, L. (2017). The politics of world polity: script-writing in international organizations. *American Sociological Review*, 82 (5), 1065–92.
- Kerner, A. (2014). What we talk about when we talk about foreign direct investment. *International Studies Quarterly*, 58(4), 804–815.
- van Kersbergen, K., & Verbeek, B. (2007). The Politics of International Norms: Subsidiarity and the Imperfect Competence Regime of the European Union. *European Journal of International Relations*, 13, 217-238.
- Keohane, R. (1984). *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton, N.J.: Princeton University Press.
- Khan, A., Wales, J., and Stuart, E., (2015). *Country priorities for data development: what does history tell us?* London: ODI.
- Kranke, M. (2020). IMF-World Bank Cooperation Before and After the Global Financial Crisis. *Global Policy*, (11)1, pp. 15–25.
- Krasner, S. (1991). Global Communications and National Power: Life on the Pareto Frontier. *World Politics* 43 (April):336–66.
- Kunz, R., Prügl, E., and Thompson, H. (2019). Gender expertise in global governance: contesting the boundaries of a field. *European Journal of Politics and Gender*, 2 (1), 23–40.
- Kuznets, S. (1949). National income and industrial structure. *Econometrica*, 17, 205–41.
- Lai, J., Rethel, L., & Steiner, K. (2017). Conceptualizing dynamic challenges to global financial diffusion: Islamic finance and the grafting of sukuk. *Review of International Political Economy*, 24(6), 958-979.
- Le Guin, Ursula K. (1974) *The Dispossessed: An Ambiguous Utopia*. [1st ed.]. New York: Harper & Row.
- Leander, A. and Waeber, O. (2018). *Assembling exclusive expertise: knowledge, ignorance and conflict resolution in the global South*. London: Routledge.
- Lepenies, P. (2013). *Die Macht der einen Zahl: eine politische Geschichte des Bruttoinlandsprodukts*. Berlin: Suhrkamp Verlag.
- Lepenies, P. (2016). *The power of a single number: the political history of GDP*. New York: Columbia University Press.
- Lequiller, F. and Blades, D. (2014). *Understanding national accounts: second edition*. Paris: OECD Publishing.
- Lim, E. (1993). Interview with John Lewis, Richard Webb and Devesh Kapur. (World Bank Group: World Bank History Project, Washington D.C., January 11, 1993).

- Linsi, L. (2018). Foreign direct investment. In: Dingwerth K and Weinhardt C (eds) *The Language of World Trade Politics: Unpacking the Terms of Trade*. London: Routledge, 50–63. p. 53ff
- Linsi, L. & Mügge, D. (2019). Globalization and the growing defects of international economic statistics, *Review of International Political Economy*, 26:3, 361-383.
- Littoz-Monnet, A. (2017). Expert knowledge as a strategic resource: international bureaucrats and the shaping of bioethical standards. *International Studies Quarterly*, 61, 584–95.
- Lombardi, D., & Woods, N. (2008). The Politics of Influence: An Analysis of IMF Surveillance. *Review of International Political Economy*, Vol. 15, No. 5, pp. 711-739.
- Mahoney, J. (2000). Path dependence in historical sociology. *Theory and Society*, 29, 507–48.
- Marsh, D., & Sharman, J. C. (2009). Policy diffusion and policy transfer. *Policy studies*, 30(3), 269-288.
- Masood, E. (2016). *The great invention: the story of GDP and the making and unmaking of the modern world*. New York: Pegasus Books.
- Mattli, W. and Büthe, T. (2003). ‘Setting international standards: technological rationality or primacy of power?’, *World Politics*, 56(1): 1–42.
- Mazzucato, M. (2018). *The Value of Everything. Making and Taking in the Global Economy*. London: Allen Lane.
- McCormick, T. (2009). *William Petty and the Ambitions of Political Arithmetic*, Oxford: Oxford University Press.
- Mehta, L. (2001). The World Bank and its emerging knowledge empire. *Human Organization* (60)2, 189–196.
- Meyer, J., Boli, J., Thomas, G., and Ramirez, F. (1997). World Society and the Nation-State. *American Journal of Sociology*, 103, 144-181.
- Min, Q. & Xu, X. (1997). 1993 年 SNA 的发展与中国新国民经济核算体系的进一步改革 [The development of SNA in 1993 and the further reform of the new national economic accounting system in China]. *财经问题研究 [Research on Financial and Economic Issues]*, 7, 60-65.
- Miranda, V. (2011). Cooking, caring and volunteering: unpaid work around the world. *OECD Social, Employment and Migration Working Papers*, no. 116. Paris: OECD Publishing.
- Mitchell, T. (2002). *Rule of Experts: Egypt, Techno-Politics, Modernity*. Berkeley: University of California Press.

- Mitchell, T. (2005). The work of economics: how a discipline makes its world. *European Journal of Sociology*, 46 (2), 297–320.
- Momani, B. (2007). IMF staff: missing link in fund reform proposals. *Review of International Organizations*, 2 (1), 39–57.
- Momani, B. and Hibben, M. (2015). Cooperation or Clashes on 19th Street? Theorizing and Assessing IMF and World Bank Collaboration. *Journal of International Organizations Studies*, 6 (2), pp. 27–43.
- Morgan, M. (2009). Seeking parts, looking for wholes. *History of Observation in Economics Working Paper Series*, 1. Amsterdam: University of Amsterdam, 1–56.
- Morgenstern, O. (1963). *On the accuracy of economic observations*. Second ed. Princeton, N.J.: Princeton University Press.
- Moschella, M. (2012). Seeing like the IMF on capital account liberalization. *New Political Economy*, (17)1, 59 – 76.
- Mosley, L. (2003). ‘Attempting Global Standards: National Governments, International Finance, and the IMF’s Data Regime’, *Review of International Political Economy*, 10(2): 331–62.
- Mügge, D. (2016). Studying macroeconomic indicators as powerful ideas. *Journal of European Public Policy*, 23 (3), 410–27.
- Mügge, D. (2019). The Revenge of Political Arithmetick: Economic Statistics and Political Purpose. Fickle Formulas Working Paper 2019/02. Amsterdam. Available at: https://fickleformulas.org/images/pdf/FFWP_2019-02_Mgge_The_Revenge_of_Political_Arithmetick.pdf.
- Mügge, D. (2020a). International Economic Statistics: Biased Arbiters in Global Affairs?. *Fudan J. Hum. Soc. Sci.* 13, 93–112.
- Mügge, D. (2020b): Economic statistics as political artefacts. *Review of International Political Economy*. Advance online publication. DOI: 10.1080/09692290.2020.1828141.
- Mügge, D. and Linsi, L. (2020). The national accounting paradox: how statistical norms corrode international economic data. *European Journal of International Relations*. Advance online publication. DOI: 10.1177/1354066120936339
- National Bureau of Statistics (NBS) (1984). 国务院关于加强统计工作的决定 [State Council Decision on Strengthening Statistical Work]. *中国统计 [China Statistics]*, (3), 4-6.
- National Bureau of Statistics (NBS) (1996). 杨保民：对加速中国统计改革之己。 [Yang Baomin: Opinions on Accelerating China’s Statistical Reform.] Retrieved from http://www.stats.gov.cn/ztc/tjzdgg/hsyjh1/yjhxslh/hsfx/201010/t20101015_69182.html. Accessed from 29-08-2017.

- National Bureau of Statistics (NBS) (2003). 《中国国民经济核算体系 (2002) 》开始实施 [“China National Economic Accounting System (2002)” began to implement]. 23-05-2003. Retrieved from http://www.stats.gov.cn/ztjc/tjzdgg/zggmjhstx/200305/t20030527_38591.html. Accessed at 29-08-2017.
- National Bureau of Statistics (NBS) (2017). 国家统计局有关负责人就国务院批复国家统计局《中国国民经济核算体系 (2016) 》有关问题答记者问 [The person in charge of the National Bureau of statistics answered the reporter's questions on the reply of the State Council to the national economic accounting system of China (2016) issued by the National Bureau of Statistics] 14-07-2017. Retrieved from http://www.stats.gov.cn/tjsj/sjjd/201707/t20170714_1513205.html
- Naughton, B. (1990). China's Experience with Guidance Planning. *Journal of Comparative Economics*, 14, 743-767.
- Naughton, B. (1995). *Growing Out of the Plan: Chinese Economic Reform, 1978-1993*. Cambridge: Cambridge University Press.
- Ngo, B. (2015). Independent evaluation of the Marrakech action plan for statistics: Evaluation report. Washington, DC: World Bank Group.
- Nielson, D., and Tierney, M. (2003). Delegation to international organizations: Agency theory and World Bank environmental reform. *International Organization*, 57, 241 – 276.
- Nijmeijer, H., & Hiemstra, L. (2008). *Consistency and Beyond: Improving Quality of National Accounts Estimates in Practice* Discussion paper 08008. Voorburg: Statistics Netherlands.
- Nölke, A. (2015). Second Image Revisited: The Domestic Sources of China's Foreign Economic Policies. *International Politics*, 52(6): 657–665.
- OECD (2004). Glossary of Statistical Terms. Retrieved from: <https://stats.oecd.org/glossary/detail.asp?ID=1726>
- PARIS21 (2019). Statistical capacity development outlook 2019. Paris: PARIS21. Retrieved from: <https://paris21.org/flagship/2019>
- Perry, J. & Nölke, A. (2006). The political economy of International Accounting Standards, *Review of International Political Economy*, 13:4, 559-586
- Petiteville, F. (2018). International organizations beyond depoliticized governance. *Globalizations*, 15 (3), 301–13.
- Philipsen, D. (2015). *The little big number: how GDP came to rule the world and what to do about it*. Princeton: Princeton University Press.
- Pigou, A.C. (1920). *The economics of welfare*. London: Macmillan Press.

- Pilling, D. (2018). *The growth delusion: wealth, poverty, and the well-being of nations*. New York: Tim Duggan Books.
- Polanyi, K. (2001). *The great transformation: The political and economic origins of our time*. Boston: Beacon Press.
- Polanyi, K. (2016). 'Socialist accounting' [Sozialistische Rechnungslegung] by Karl Polanyi: with preface 'Socialism and the embedded economy' (Fischer, A., Woodruff, D., and Bockman, J., Trans). *Theory and Society*, 45, 385–427. (Original work published 1922).
- Porter, T. (1986). *The rise of statistical thinking: 1820-1900*. Princeton: Princeton University Press.
- Porter, T. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*, Princeton, NJ: Princeton University Press.
- Price, R. (1998). Reversing the Gun Sights: Transnational Civil Society Targets Land Mines. *International Organization*, 52 (3), 613-44.
- Qi, S. & Chen, J. (2007). The Evolution and Issues of I-O Table Compilation during Economy Transformation in China. *International Conference on Experiences and Challenges in Measuring National Income and Wealth in Transition Economies IARIW*, Beijing, September 18-21, 2007. Retrieved from <http://iariw.org/papers/2007/shuchang.pdf>.
- Qiu, D. & Song, X. (2010). 中国经济统计学60年 [Development of Economic Statistics in China over 60 Years]. *统计研究 [Statistical Research]*, 27(1), 26-34.
- Quaglia, L. (2017). The political economy of post-crisis international standards for resolving financial institutions, *New Political Economy*, 22:5, 595-609.
- Quark, A. (2012). Scientized politics and global governance in the cotton trade: Evaluating divergent theories of scientization, *Review of International Political Economy*, 19:5, 895-917.
- Radermacher, W. (1999). 'Indicators, Green Accounting and Environment Statistics: Information Requirements for Sustainable Development'. *International Statistical Institute (ISI)*, 67(3), pp. 339–354. Available at: <http://www.jstor.org/stable/1403710>.
- Rajão, R. and Duarte, T. (2018). Performing postcolonial identities at the United Nations' climate negotiations. *Postcolonial Studies*, 21 (3), 364–78.
- Rao, R.K.R.V. (1953). Some reflections on the comparability of real national incomes of industrialised and under-developed countries. *Review of Income and Wealth*, 3, 178–210.
- Ravallion, M. (2010). Poverty Lines across the World. World Bank Development Research Group, Policy Research Working Paper 5284. Washington, DC: World Bank.

- Ravallion, M., Chen, S., and Prem, S. (2009). Dollar a day revisited. *The World Bank Economic Review*, 23(2): 163 – 184.
- Ravallion, M., Datt, G. and van de Walle, D. (1991). Quantifying Absolute Poverty in the Developing World. *Review of Income and Wealth*, 37(4): 345–361.
- Reddy, S., and Pogge, T. (2005). *How Not to Count the Poor*. Mimeo. Columbia University Academic Commons.
- Reichmann, T. (2016). Progress Through Crises: The Evolution of the IMF’s Statistical Arsenal (IEO Background Paper BP/16/02). Washington, DC: Independent Evaluation Office of the IMF.
- Reid, M. (1934). *Economics of household production*. New York: J. Wiley & Sons.
- Reinold, T. (2017). The path of least resistance: mainstreaming ‘social issues’ in the International Monetary Fund. *Global Society*, 31 (3), 392–416.
- Rivera, E. (2008). *Redefining the labour force framework: Some inputs from the Philippine experience*. Seminar on Employment and unemployment: Revisiting the relevance and conceptual basis of the statistics, International Labour Organization, Room document No. 21, Geneva, Switzerland, 4-5 December.
- Röper, N. (2020). Between substantive and symbolic influence: Diffusion, translation and bricolage in German pension politics. *Review of International Political Economy*. <https://doi.org/10.1080/09692290.2020.1790405>
- Round, J. (2014). *Assessing the demand and supply of statistics in the developing world: some critical factors*. PARIS21. Discussion Paper No. 4 May 2014. Paris: PARIS21.
- Rosen, D. & Bao, B. (2015). *Broken Abacus? A More Accurate Gauge of China’s Economy*. Washington D.C: Centre for Strategic and International Studies.
- Round, J. (2014). Assessing the demand and supply of statistics in the developing world: some critical factors. PARIS21. Discussion Paper No. 4 May 2014. Paris: PARIS21
- Sakuma, I. (2013). The production boundary reconsidered. *Review of Income and Wealth*, 59 (3), 556–67.
- Salais, R., Baverez, N. and Reynaud, B. (1986). *L’invention du chômage*, Paris: Puf.
- Samuel, B. (2014). Economic calculations, instability and (in)formalisation of the state in Mauritania, 2003–2011. *Canadian Journal of Development Studies / Revue canadienne d’études du développement*, (35)1, 77–96.
- Samuels, L.H. (1962). The uses and limitations of national economic accounting, with special reference to South Africa. *Review of Income and Wealth*, 1, 162–81.
- Savage, J. (2005). *Making the EMU: The Politics of Budgetary Surveillance and the Enforcement of Maastricht*, Oxford: Oxford University Press.

- Schmelzer, M. (2016). *The Hegemony of Growth: the OECD and the making of the economic growth paradigm*. Cambridge: Cambridge University Press.
- Schmidt, V. (2008). Discursive institutionalism: the explanatory power of ideas and discourse. *Annual Review of Political Science*, 11, 303–26.
- Schmidt, V. (2010). Taking ideas and discourse seriously: explaining change through discursive institutionalism as the fourth ‘new institutionalism’. *European Political Science Review*, 2 (1), 1–25.
- Schneider, F. and Enste, D. (2000). “Shadow Economies: Size, Causes, and Consequences.” *Journal of Economic Literature*, XXXVIII, pp. 77-114.
- Scott, J. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press.
- Seabrooke, L. (2012). Pragmatic numbers: The IMF, financial reform, and policy learning in least likely environments. *Journal of International Relations and Development*, 15, 486–505.
- Seabrooke, L. and Tsingou, E. (2016). Bodies of knowledge in reproduction: epistemic boundaries in the political economy of fertility. *New Political Economy*, 21 (1), 69–89.
- Seabrooke, L. and Wigan, D. (2016). Powering ideas through expertise: professionals in global tax battles. *Journal of European Public Policy*, 23 (3), 357–74.
- Seguino, S. (2019). Engendering macroeconomic theory and policy. *Feminist Economics*, 0 (0), 1–35. doi:10.1080/13545701.2019.1609691.
- Sell, S. and Prakash, A. (2004). Using Ideas Strategically: The Contest Between Business and NGO Networks in Intellectual Property Rights. *International Studies Quarterly*, (48), 143–175.
- Sending, O. (2015). *The politics of expertise: competing for authority in global governance*. Ann Arbor: University of Michigan Press.
- Sending, O., & Lie, J., (2015). The limits of global authority: World Bank benchmarks in Ethiopia and Malawi. *Review of International Studies*, 41, 993–1010.
- Smith, J. (2012). The GDP Illusion: Value Added versus Value Capture. *Monthly Review*, 64(3), 97–113.
- Speich, D. (2008). Traveling with the GDP through early development economics’ history. *The nature of evidence: How well do facts travel?* (Working Paper No. 33/08). doi.org/10.2139/ssrn.1291058
- Speich, D. (2011). The use of global abstractions: National income accounting in the period of imperial decline. *Journal of Global History*, Special Issue, 6, pp. 7–28.

- Sridhar, D. (2007). Economic ideology and politics in the World Bank: defining hunger. *New Political Economy*, 12 (4), 499–516.
- State Council (1985). 国务院办公厅转发国家统计局关于建立第三产业统计的报告的通知 [Notice of the General Office of the State Council Forwarded the State Statistics Bureau's Report on the Establishment of Tertiary Industry Statistics]. April 5th, 1984. Retrieved from <http://www.chinaacc.com/new/63/64/80/2006/3/so894111204116360022331-0.htm>. Accessed on 20-12-2017.
- State Council (1992). 国务院关于实施新国民经济核算体系方案的通知 [Notice of the State Council on Implementing the New National Economic Accounting System Plan]. August 30, 1992. Retrieved from http://www.china.com.cn/law/flfg/txt/2006-08/08/content_7059353.htm
- Stiglitz, J., Sen, A. and Fitoussi, J.-P. (2010). *Report by the Commission on the Measurement of Economic Performance and Social Progress*, Paris: INSEE.
- Studenski, P. (1958). *The income of nations: theory, measurement and analysis past and present*. New York: New York University Press.
- Tamm, H., & Snidal, D. (2014). Rational choice and principal-agent theory. In Weiss, G. & Wilkinson, R. (Eds.), *International organization and global governance* (pp. 132-143). New York, NY: Routledge.
- Taylor, M. (2016). The political economy of statistical capacity: A theoretical approach (Discussion Paper no. IDB-DP-471). Washington, DC: Inter-American Development Bank.
- Telleria, J. (2017). Power relations? What power relations? The depoliticising conceptualisation of development of the UNDP. *Third World Quarterly*, 38 (9), 2143–58.
- Thérien, J., & Pouliot, V. (2019). Global governance as patchwork: The making of the Sustainable Development Goals. *Review of International Political Economy*. Advance online publication. doi.org/10.1080/09692290.2019.1671209
- Thomassen, E. (2017). Translating central bank independence into Norwegian: central bankers and the diffusion of central bank independence to Norway in the 1990s. *Review of International Political Economy*, 24(5): 839-858.
- Topalov, C. (1994). *Naissance du chômeur, 1880–1910*, Paris: Albin Michel, 1994.
- Tsingou, E. (2015). Club governance and the making of global financial rules. *Review of International Political Economy*, 22 (2), 225–56.
- United Nations (2015). *The world's women 2015: trends and statistics*. New York: United Nations, Department of Economic and Social Affairs, Statistics Division.

- UNECE, Eurostat and OECD (2011). *The Impact of Globalization on National Accounts*, New York and Geneva: United Nations.
- United Nations Statistical Commission (UNSC) (1981). *Report on the 21st Session [online]*. Available from: <https://unstats.un.org/unsd/statcom/reports/>.
- United Nations Statistical Commission (UNSC) (1983). Report on the 18th session. Available online: <https://unstats.un.org/unsd/statcom/reports/>
- UNSD (2013). *Guidelines on Integrated Economic Statistics*. Department of Social and Economic Affairs Studies in Methods, Series F No. 108. New York: United Nations.
- UNSD (2020a). Statistics Division in Brief 2020. New York: United Nations Department of Economic and Social Affairs. Retrieved online: <https://unstats.un.org/home/documents/brochures/2020-UNSD-Brochure.pdf>
- UNSD (2020b). United Nations Statistics Division Organizational Chart. Retrieved online: <https://unstats.un.org/home/documents/UNSD-Org-Chart-1-July-2020.pdf>
- United Nations Economic Commission for Africa (2020). Economic Statistics and National Accounts. Retrieved from: <https://www.uneca.org/acs/pages/economic-statistics-and-national-accounts>
- Vanoli, A. (2005). *A history of national accounting*. Amsterdam: IOS Press.
- Varga, M. (2018). 'Subsistence' Readings: World Bank and State Approaches to Commercialising Agriculture in Post-Communist Eurasia. *Journal of Development Studies*, 55:6, 1253-1266, DOI: 10.1080/00220388.2018.1453607
- Verdun, A. (1999). The role of the Delors Committee in the creation of EMU: an epistemic community? *Journal of European Public Policy*, 6(2), 308–28.
- Vetterlein, A. (2012). Seeing like the World Bank on poverty. *New Political Economy*, 17 (1), 35–58.
- Vetterlein, A. (2014). The discursive power of international organisations. In: D. Béland, and K Petersen, ed. *Analysing social policy concepts and language*. Bristol: Bristol University Press, 101–26.
- de Vlieger, J., and Mügge, D. (2020). 'The Ambiguity of Public Debt Measures', in Caruso, C., and Morvillo, M. (eds) *Il governo dei numeri: indicatori macro-economici e decisione di bilancio nello Stato costituzionale*. Bologna: Mulino, pp. 31-51.
- Wade, R. (2012). The Politics behind World Bank Statistics. The Case of China's Income. *Economic and Political Weekly*, 47(25), 17-18.
- Wallace, J. L. (2016). Juking the stats? Authoritarian information problems in China. *British Journal of Political Science*, 46(1), 11-29.

- Ward, M. (2004a). Some reflections on the 1968-93 SNA revision. *Review of Income and Wealth*, 50 (2), 299–313.
- Ward, M. (2004b) *Quantifying the World: UN Ideas and Statistics*. New York: Indiana University Press.
- Waring, M. J. (1999). *Counting for Nothing: What Men Value and What Women are Worth*. Toronto: University of Toronto Press.
- Waring, M. (2003). Counting for something! Recognizing women's contribution to the global economy through alternative accounting systems. *Gender and Development*, 11 (1), 35–43.
- Weaver, C. (2007). The world's bank and the bank's world. *Global Governance*, 13, 493–512.
- Wight, C. (2007). A Manifesto for Scientific Realism in IR: Assuming the Can-Opener Won't Work! *Millennium: Journal of International Studies* (35)2: 379-398.
- Wood, C. (1997). The first world/third party criterion: a feminist critique of production boundaries in economics. *Feminist Economics*, 3 (3), 47–68.
- World Bank (1983). CHINA Socialist Economic Development Volume I: The Economy, Statistical System, and Basic Data. Washington D.C.: World Bank.
- World Bank (1985). China. Long-Term Development Issues and Options. A World Bank Country Economic Report. Baltimore: Johns Hopkins University Press.
- World Bank (1992). China: Statistical System in Transition. Washington D.C.: World Bank.
- World Bank (1992b). China Country Economic Memorandum. Reform and the Role of the Plan in the 1990s. Report No. 10199-CHA. June 19, 1992.
- World Bank (1994). China: GNP per Capita. Report No. 13580-CHA, December 15, 1994. Washington D.C.: World Bank.
- World Bank (2004). The Marrakech action plan for statistics: Better data for better results. Proceedings of the second international roundtable on managing for development results, Marrakech, Morocco, February 4-5.
- World Bank, (2012). *The 2008 SNA concepts in brief: a complement to the System of National Accounts 2008*. Washington, DC: The World Bank.
- World Bank. 2017. Data for Development: An Evaluation of World Bank Support for Data and Statistical Capacity. Independent Evaluation Group. Washington, DC: World Bank.
- World Bank (2018a). *Piecing Together the Poverty Puzzle*. Washington, DC: The World Bank Group.

- World Bank (2018b). International Comparison Program (ICP). History. Accessed on 09-03-2018. Retrieved from <http://www.worldbank.org/en/programs/icp#2>.
- World Bank (2019a). Statistics for Results Facility: Catalytic Fund Annual Report May 2019. Development Data Group. Washington, DC: The World Bank.
- World Bank (2019b). Trust Fund for Statistical Capacity Building: Annual Progress Report (April 1, 2018 – March 31, 2019). Development Data Group. Washington, DC: The World Bank.
- Wu, H. X. (2000). China's GDP Level and Growth Performance: Alternative Estimates and the Implications. *Review of Income and Wealth*, 46(4), 475-499.
- Xu, X. (1991). 我国国民经济核算中的几个问题 [Several Problems in the Accounting of the National Economy]. *经济科学 [Economic Science]*, 4, 55-58.
- Xu, X. (2001). 关于《中国国民经济核算体系(试行方案)》的修订 [Amendment to China's National Account System (Pilot Scheme)]. *统计研究 [Statistical Research]*, 4, 3-8.
- Xu, X. (2009). The establishment, reform, and development of China's system of national accounts. *Review of Income and Wealth*, 55(s1), 442-465.
- Xu, X. (2014). 关于我国国民经济核算体系的修订 [About the Revisions on the System of National Accounts in China]. *全球化 [Globalization]*, 1, 14-20.
- Yarrow, D. (2020). Valuing knowledge: The political economy of human capital accounting, *Review of International Political Economy*, DOI: 10.1080/09692290.2020.1796751
- Yue W. (1989). China is setting up a new national accounting system 我国正在建立新的国民经济核算体系 [China is setting up a new national accounting system]. *中国统计 [China Statistics]*, (7), 22-24.
- Zhi, F. (1992). 中国国民经济核算体系试行方案简介之二 [A Brief Introduction to the Trial Scheme of China's National Economic Accounting System]. *中国统计 [China Statistics]*, (5), 19-21.
- Zimmermann, B. (2006). *Arbeitslosigkeit in Deutschland. Zur Entstehung Einer Sozialen Kategorie*, Frankfurt am Main: Campus Verlag.
- Zimmermann, L. (2016). Same same or different? Norm diffusion between resistance, compliance, and localization in post-conflict states. *International Studies Perspectives*, 17(1), 98-115.

Appendix: List of interviews

Appendix: List of interviews

Nearly all respondents requested to be referred to by professional title only, not by name (partial anonymity). For consistency, all names have been removed.

Interviews directly cited (in order of first appearance in manuscript)

Interview 1:

Statistician – National Accounts and Capacity Building, UNSD, January 8, 2019, New York, NY, United States

Interview 2:

Economic Statistician – National Accounts, World Bank / ISWGNA, June 14, 2017, Edinburgh, United Kingdom

Interview 3:

Chief of UNSD Economic Statistics Branch, UNSD, January 7, 2019, New York, NY, United States

Interview 4:

Statistician – National Accounts, UNSD, January 8, 2019, New York, NY, United States

Interview 5:

Former Acting Director, UNSD, March 9, 2017, Amsterdam, the Netherlands

Interview 6:

Senior Economist, World Bank Development Data Group, January 3, 2019, Washington, DC, United States

Interview 7:

Chief of Real Sector Division, IMF Statistics Department, January 2, 2019, Washington, DC, United States

Interview 8:

Deputy Director for Data and Surveillance, IMF Statistics Department, January 3, 2019, Washington, DC, United States

Interview 9:

Senior Statistician, World Bank, January 4, 2019, Washington, DC, United States

Interview 10:

Economic Statistician – National Accounts, UN ESCAP Statistics Division, June 21, 2018, Bangkok, Thailand

Interview 11:

Former World Bank official, April 11, 2018, Washington, DC, United States (Interviewed by Joan van Heijster, Chapter 6 co-author)

Interview 12:

Professor of Statistics, Beijing Normal University, September 8, 2017, Beijing, China (Interviewed by Joan van Heijster, Chapter 6 co-author)

Interview 13:

Former World Bank consultant, April 11, 2018, Washington, DC, United States (Interviewed by Joan van Heijster, Chapter 6 co-author)

Interview 14:

Former Chinese statistician (National Accounts Department), National Bureau of Statistics of China, September 1, 2017, Beijing, China (Interviewed by Joan van Heijster, Chapter 6 co-author)

All interviews

Organizational affiliation	Job title	Date	Location	In-text reference (if applicable)
UNSD	Former Head, National Accounts Branch and Interregional Adviser National Accounting	January 11, 2017	Amsterdam, the Netherlands	
UNSD	Former Acting Director	March 9, 2017	Amsterdam, the Netherlands	Interview 5
Oxford Policy Management	Portfolio Manager & Senior Consultant, Official Statistics	June 15, 2017	Oxford, United Kingdom	
Oxford Policy Management	Senior Consultant for Expert Group on Refugee Statistics, Former OPM Lead on Official Statistics	June 15, 2017	Oxford, United Kingdom	
Oxford Policy Management	Senior Consultant, Official Statistics	June 15, 2017	Oxford, United Kingdom	
IMF Statistics Department	Economic Statistician (retired)	June 15, 2017	Didcot, United Kingdom	
World Bank / ISWGNA	Economic Statistician, National Accounts	June 14, 2017	Edinburgh, United Kingdom	Interview 2
National Statistical Office of Thailand	Thai NSO Official 1	June 18, 2018	Bangkok, Thailand	

National Statistical Office of Thailand	Thai NSO Official 2	June 18, 2018	Bangkok, Thailand	
National Statistical Office of Thailand	Thai NSO Official 3	June 18, 2018	Bangkok, Thailand	
National Statistical Office of Thailand	Thai NSO Official 4	June 18, 2018	Bangkok, Thailand	
Lao Statistics Bureau	Deputy Director General, Head of International Cooperation	June 19, 2018	Vientiane, Lao PDR	
UN ESCAP Statistics Division	Director (Officer in Charge)	June 21, 2018	Bangkok, Thailand	
UN ESCAP Statistics Division	Economic Statistician, National Accounts	June 21, 2018	Bangkok, Thailand	Interview 10
UN ESCAP Statistics Division	Economic Statistician	June 21, 2018	Bangkok, Thailand	
IMF Statistics Department	Chief of Real Sector Division	January 2, 2019	Washington, DC, United States	Interview 7
IMF Statistics Department	Deputy Division Chief	January 2, 2019	Washington, DC, United States	
IMF Statistics Department	Deputy Director for Data and Surveillance	January 3, 2019	Washington, DC, United States	Interview 8
World Bank	Senior Economist, Development Data Group	January 3, 2019	Washington, DC, United States	Interview 6
World Bank	Economic Statistician, Development Data Group	January 3, 2019	Washington, DC, United States	
IMF Statistics Department	Deputy Director for Capacity Development	January 4, 2019	Washington, DC, United States	
World Bank	Senior Statistician	January 4, 2019	Washington, DC, United States	Interview 9
UNSD	Chief of UNSD Economic Statistics Branch	January 7, 2019	New York, NY, United States	Interview 3
UNSD	Statistician – Business Statistics	January 7, 2019	New York, NY, United States	
UNSD	Statistician – Environmental Economic Accounts	January 7, 2019	New York, NY, United States	
UNSD	Statistician – Environmental Economic Accounts (EU project)	January 7, 2019	New York, NY, United States	
UNSD	Statistician – Trade and Industry Statistics	January 7, 2019	New York, NY, United States	
UNSD	Statistician – Data Innovation and Capacity Branch	January 7, 2019	New York, NY, United States	
UNSD	Statistician – National Accounts and Capacity Building	January 8, 2019	New York, NY, United States	Interview 1
UNSD	Statistician – National Accounts	January 8, 2019	New York, NY, United States	Interview 4
UNSD	Statistician – Gender Statistics	January 8, 2019	New York, NY, United States	
UNSD	Statistician – Demographic Statistics	January 8, 2019	New York, NY, United States	
Ghana Statistical Services	Head of National Accounts and Economic Indicators	August 2, 2019	Accra, Ghana	

Ghana Statistical Services	Head of Monitoring and Evaluation	August 2, 2019	Accra, Ghana
Ghana Statistical Services	Data Scientist	August 2, 2019	Accra, Ghana

Interviews conducted by Joan van Heijster (co-author for Chapter 6)

Organizational affiliation	Job title	Date	Location	In-text reference
National Bureau of Statistics of China	Former Chinese statistician (National Accounts Department)	September 1, 2017	Beijing, China	Interview 14
Beijing Normal University	Professor of Statistics	September 8, 2017	Beijing, China	Interview 12
World Bank	Former World Bank official	April 11, 2018	Washington, DC, United States	Interview 11
World Bank	Former World Bank consultant	April 11, 2018	Washington, DC, United States	Interview 13

Summary

Summary

Economic statistics – such as those of economic growth, debt, inflation, unemployment, and others – are ubiquitous in global and domestic governance. Policymakers, investors, and development agencies, among others, rely on these figures to make sense out of a complex world. The kinds of statistics that are available, and the particular ways they are measured, shape actors' perceptions and thereby influence policy decisions from behind the scenes.

Because the information provided by economic statistics is so crucial for so many actors – from policymakers, to institutional investors, to NGOs – a great deal of effort goes into making statistics available, comparable, and reliable. Since the late 1940s, international organizations (IOs) such as the United Nations, the International Monetary Fund (IMF), and the World Bank have been at the forefront of efforts to harmonize economic statistics around the world. The international statistical standards that they have created, like the System of National Accounts (SNA) and the Balance of Payments Manual (BPM), are intended to ensure that official statistics are trustworthy and comparable.

However, there are gaps between the ambitions of IOs and the actual statistical practices of member countries. For one thing, international statistical standards often clash with socioeconomic realities within countries. Such issues are particularly problematic in the developing world, where statistical practices have often been imported from the highly industrialized countries in which international standards originated. Moreover, low-income countries often have insufficient statistical capacity to meet external demand for harmonized statistics. As a result, data is often missing or of poor quality, and compliance with standards might only be superficial or partial. What is more, especially in countries where these capacity constraints are present, the importance placed on internationally harmonized economic statistics can potentially crowd out other data-gathering that would be more suitable to local policy priorities. Prioritizing certain kinds of statistics over others can force statisticians to make unreliable estimates or limit policymakers' knowledge of pressing social issues if they are not measured at all.

Given the degree to which statistical practices shape perceptions and policies, combined with the remarkable challenge of reducing so much diversity into shared global measurement standards, it is not clear why global harmonization efforts take the form that they do. These observations and concerns lead to the research question: What are the drivers of international harmonization of economic statistics?

Above all, the findings demonstrate that statistical harmonization is primarily led by experts, and that the insulated nature of expertise-based governance allows for shared norms and ideas to influence the form that harmonization efforts take. However, there are

important caveats to this. First, there are built-in limitations to governance by numbers. While economic statistics are undoubtedly important and useful, there are inevitable tradeoffs and limitations. These inherent features of statistics limit what experts can actually achieve, and also leave them vulnerable to critique when statistics fall short of data users' expectations. Second, inter-organizational dynamics and the autonomy of experts vis-à-vis formal organizations influence different types of statistical governance in different ways. In line with an epistemic communities perspective, specialized expertise allows statistical standard-setting to be shaped by professional norms. But statistical capacity building and compliance monitoring are, to varying degrees, influenced by the priorities of individual IOs. This finding is in line both with constructivist perspectives and principal-agent perspectives that expect IOs to exert autonomy separate from their member states. The third caveat is that experts are not the only powerful agents in statistical harmonization. The case of GDP in China, at least historically, proves to be an exception to this main finding. In that case study, it is not experts but state officials and policymakers with the most agency over the adoption of international statistical standards. Nonetheless, transnational experts (mainly World Bank staff) and domestic government statisticians did still play important roles in transferring knowledge and norms.

The research builds upon a burgeoning literature on the politics of economic statistics and indicators. It engages with and refines international political economy (IPE) and international relations (IR) theories related to international standards, norm and policy diffusion, and international organizations. The empirical chapters shed light on many of the key actors and institutions that govern the international statistical system. They also aim to explain why the system takes the form that it does and how it has evolved over time. The theoretical contributions also deepen our understanding both of the topic (statistical harmonization) and the broader IPE and global governance literatures.

The dissertation is supported empirically by semi-structured expert interviews, archival research, official document analysis, participant-observation, and extensive literature reviews. The core of the dissertation is comprised of four empirical chapters, each based on academic articles (either published or under review). The first empirical chapter describes and explains the global governance of GDP measurement. The chapter closely investigates one of the most controversial aspects of GDP measurement, namely the exclusion of unpaid household services (or 'housework') from the production boundary of the SNA. The second empirical chapter asks whether there are universal limits to the harmonization of economic statistics. The argument is that standard-setting is constrained by a trilemma that results from contradictory demands of stakeholders, including policymakers, analysts, and citizens. The third empirical chapter addresses cooperation between IOs in global statistical harmonization. The chapter focuses on the efforts of, and interactions between, the UN, the World Bank and the IMF. It aims to explain variation between three main governance functions: standard-setting, statistical capacity building, and efforts to

monitor compliance with standards. The final empirical chapter explains the diffusion of the SNA (and thus also GDP) to China. Until the 1990s, China produced its own, Soviet-inspired, indicator of economic production, and was the last major holdout to using GDP.

There are important academic and societal implications. The findings suggest that the production of official statistics has long been insulated from stakeholder participation beyond a small community of experts. This was possible because of the specialized knowledge required. However, it is not outright manipulation that critics should be concerned about. Rather, it is standard-setting where experts have the most power to shape how our economies are measured. Thus, the most important choices with distributive implications come before raw data is collected, not after. Nor is it true that these choices in standard-setting are malicious or intended to disenfranchise certain groups. Rather, most shortcomings in economic statistics result from the banal influence of economic theories and professional norms about statistics, including statisticians' ambitions for comparability, utility for policymaking and analysis, and prescriptiveness of uniform standards. The fact that the distributive implications of official statistics are now debated publicly is partly a response to high-profile critiques and contestation. It also results from a growing recognition in the development community and among policymakers that a wider range of data is required to meet contemporary challenges such as poverty, inequality, gender disparities, climate change, and biodiversity loss.

Samenvatting

Economische cijfers – zoals die over economische groei, schulden, inflatie of werkloosheid – zijn alomtegenwoordig in nationaal en mondiaal bestuur. Zo doen onder meer beleidsmakers, investeerders en ontwikkelingsbureaus een beroep op deze cijfers om een complexe wereld te begrijpen. De soorten statistieken die beschikbaar zijn, en de manier waarop die statistieken tot stand komen, zijn zeer bepalend voor de perceptie van zulke actoren en hebben dus grote impact op beleidsbeslissingen die (achter de schermen) worden genomen.

Omdat de informatie over economische statistieken zo cruciaal is voor zoveel actoren – van beleidsmakers en institutionele beleggers, tot ngo's – wordt er veel energie gestoken in het beschikbaar, vergelijkbaar en betrouwbaar maken van de cijfers. Sinds het einde van de jaren 40 van de vorige eeuw lopen internationale organisaties (IO's) als de Verenigde Naties (VN), het Internationaal Monetair Fonds (IMF) en de Wereldbank voorop bij pogingen om economische statistieken over de hele wereld te harmoniseren. De door hen opgestelde internationale statistische standaarden, zoals het System of National Accounts (SNA) en het Balance of Payments Manual (BPM), moeten ervoor zorgen dat officiële statistieken betrouwbaar en internationaal vergelijkbaar zijn.

Er zijn echter discrepanties tussen deze door internationale organisaties geïmplementeerde standaardisatie en de daadwerkelijke statistische praktijken van hun lidstaten. Om te beginnen botsen internationale statistische normen vaak met de sociaaleconomische realiteit binnen landen. Dergelijke problemen zijn met name aan de orde in de derde wereld, waar statistische methoden en standaarden vaak zijn geïmporteerd uit de sterk geïndustrialiseerde landen waar internationale normen zijn ontstaan. Bovendien hebben lage-inkomenslanden vaak onvoldoende statistische capaciteit om aan de externe vraag naar geharmoniseerde statistieken te voldoen. Als gevolg hiervan ontbreken vaak gegevens of zijn ze van beperkte kwaliteit, en kan de naleving van normen slechts oppervlakkig of gedeeltelijk zijn. Bovendien kan het belang dat aan internationaal geharmoniseerde economische statistieken wordt gehecht, met name in landen waar deze capaciteitsbeperkingen bestaan, mogelijk andere gegevensverzameling verdringen die beter passen bij lokale beleidsprioriteiten. Het aldus prioriteren van bepaalde soorten statistieken boven andere kan ertoe leiden dat statistici genoodzaakt zijn onbetrouwbare schattingen te maken. In het uiterste geval kan het zelfs gebeuren dat kennis en informatie over dringende sociale kwesties helemaal ontbreekt bij beleidsmakers, wanneer de statistieken helemaal niet beschikbaar zijn.

Gezien het belang van statistische informatie voor percepties van beleidsmakers en het maken van beleid, in combinatie met de ingewikkelde opgave om zoveel statistische diversiteit terug te brengen tot gedeelde wereldwijde meetstandaarden, is het niet zonder meer duidelijk waarom wereldwijde harmonisatie-inspanningen de vorm aannemen die ze hebben. Deze observaties en zorgen leiden tot de onderzoeksvraag: Wat zijn de drijvende krachten achter internationale harmonisatie van economische statistieken?

Bovenal tonen de bevindingen van dit onderzoek aan dat statistische harmonisatie in de eerste plaats wordt geleid door experts, en dat het geïsoleerde karakter van op expertise gebaseerde governance de gedeelde normen, ideeën en dus de vorm van harmonisatie-inspanningen beïnvloedt. Bij deze stelling zijn echter belangrijke kanttekeningen te plaatsen. Ten eerste zijn er *inherente grenzen* aan op getallen en cijfers gebaseerde governance. Hoewel economische statistieken zonder meer belangrijk en nuttig zijn, zijn de uitkomsten niet altijd eenduidig; het is onvermijdelijk dat er soms aanvullend bepaalde afwegingen en keuzen moeten worden gemaakt. Deze inherente kenmerken van statistieken beperken wat experts daadwerkelijk kunnen bereiken, en dat maakt die experts ook kwetsbaar voor kritiek wanneer statistieken niet voldoen aan de verwachtingen van gebruikers van die gegevens. Ten tweede is de *dynamiek* tussen formele organisaties en de autonomie van experts waarmee zij werken van belang. Die dynamiek beïnvloedt verschillende soorten statistische governance op verschillende manieren. Beïnvloed door sterk gespecialiseerde expertise, kunnen statistische standaarden vorm krijgen op basis van professionele normen – dit is in lijn met het zogeheten *epistemic communities* concept. Maar ook de voorkeuren en prioriteiten van individuele internationale organisaties bepalen in belangrijke mate de statistische capaciteitsopbouw en nalevingscontrole. Deze bevinding over interorganisatorische dynamieken is in lijn met zowel constructivistische perspectieven als principaal-agent-perspectieven, die veronderstellen dat internationale organisaties hun autonomie onafhankelijk van hun lidstaten uitoefenen. Het derde voorbehoud is dat experts niet de enige machtige actoren zijn op het gebied van statistische harmonisatie. Zo blijkt het bbp in China, als historische uitzondering, niet uitsluitend te zijn gebaseerd op informatie van experts. In die casestudie zijn het niet experts, maar staatsfunctionarissen en beleidsmakers die de meeste zeggenschap hebben over de goedkeuring van internationale statistische normen. Niettemin speelden transnationale experts (voornamelijk medewerkers van de Wereldbank) en nationale statistici van de overheid ook in die casus een belangrijke rol bij het overdragen van kennis en normen.

Het onderzoek bouwt voort op de groeiende literatuur over de politiek van economische statistieken en indicatoren. Het gebruikt en verfijnt theorieën over internationale politieke economie (IPE) en internationale betrekkingen (IR) met betrekking tot internationale normen, norm- en beleidsverspreiding en internationale organisaties. De empirische hoofdstukken werpen licht op veel van de belangrijkste actoren en instellingen die het internationale statistische systeem besturen. Het doel daarbij is te verklaren waarom het systeem de vorm aanneemt die het heeft en hoe het zich in de loop van de tijd heeft ontwikkeld. Het onderzoek draagt zodoende bij aan een dieper *theoretisch* begrip van zowel het onderwerp (statistische harmonisatie) als de bredere IPE- en global governance-literatuur.

Het proefschrift is empirisch gebaseerd op semigestructureerde interviews met experts, archiefonderzoek, officiële documentanalyse, participatieve observaties en uitgebreid literatuuronderzoek. De kern van het proefschrift bestaat uit vier empirische hoofdstukken,

die elk als wetenschappelijk artikel zijn gepubliceerd of ter publicatie zijn ingediend. Het eerste empirische hoofdstuk beschrijft en verklaart de globale governance van het meten van het bruto binnenlands product (bbp). Dit hoofdstuk onderzoekt een van de meest controversiële aspecten van het meten van bbp, namelijk de uitsluiting van onbetaalde huishoudelijke diensten (of ‘huishoudelijk werk’) van de productiegrens van de SNA. Het tweede empirische hoofdstuk stelt de vraag of er universele grenzen zijn aan de harmonisatie van economische statistieken. Het argument is dat normstelling wordt beperkt door een trilemma dat het gevolg is van tegenstrijdige eisen van belanghebbenden, waaronder beleidsmakers, analisten en burgers. Het derde empirische hoofdstuk behandelt de samenwerking tussen internationale organisaties op het gebied van mondiale statistische harmonisatie. Het hoofdstuk richt zich op de inspanningen van en interacties tussen de UN, de Wereldbank en het IMF. Het hoofddoel daarbij is om de variatie tussen drie belangrijke bestuursfuncties te verklaren: het vaststellen van normen, het opbouwen van statistische capaciteit en inspanningen om de naleving van normen te controleren. Het vierde en laatste empirische hoofdstuk verklaart de verspreiding van het SNA (en dus ook het bbp) naar China. Tot de jaren 90 van de vorige eeuw hanteerde China zijn eigen, door de Sovjet-Unie geïnspireerde indicator van de economische productie, en was het de laatste grote macht die het gebruik van bbp afwees.

De bevindingen van dit onderzoek hebben belangrijke academische en maatschappelijke implicaties. Op basis van de bevindingen wordt verondersteld dat de ontwikkeling van officiële statistieken lange tijd is afgeschermd van de deelname van belanghebbenden buiten een kleine gemeenschap van experts. Dit was mogelijk het gevolg van behoefte aan gespecialiseerde kennis. Kritiek hierop moet hier echter niet uitgaan van expliciete en regelrechte manipulatie. Critici moeten eerder kijken naar de normatieve instelling waarbij experts de meeste macht hebben om vorm te geven aan hoe onze economieën worden gemeten. De belangrijkste keuzes omtrent statistieken – met belangrijke distributieve implicaties – worden dus gemaakt *voordat* de ruwe gegevens worden verzameld, niet zozeer erna. Ook is het niet zo dat deze keuzes in normstelling kwaadwillig zijn of bedoeld zijn om bepaalde groepen het recht tot deelname te ontnemen. Integendeel, de meeste tekortkomingen in economische statistieken zijn simpelweg het inherente gevolg van de invloed van economische theorieën en professionele normen over statistiek – waaronder de ambities van statistici wat betreft vergelijkbaarheid, bruikbaarheid voor beleidsvorming en analyse, en het voorschrijven van uniforme normen. Het feit dat de distributieve implicaties van officiële statistieken nu publiekelijk worden besproken, is deels een reactie op spraakmakende kritieken en betwistingen. Het komt bovendien voort uit een groeiende erkenning in de internationale ontwikkelingssamenwerking en onder beleidsmakers dat er een breder scala aan gegevens nodig is om hedendaagse uitdagingen zoals armoede, ongelijkheid, genderongelijkheid, klimaatverandering en verlies van biodiversiteit het hoofd te bieden.

