

Reassessing Tax and Development Research: A New Dataset, New Findings, and Lessons for Research

WILSON PRICHARD*

University of Toronto, Canada
Institute of Development Studies, UK

Summary. — There is growing concern with the weaknesses of economic statistics relating to developing countries, and the risks that poor data have generated misleading research findings and poor policy advice. Cross-country tax data offer a striking example, with existing datasets frequently highly incomplete, analytically imprecise, plagued by errors, and sharply lacking in transparency. This paper introduces the new Government Revenue Dataset from the International Centre for Tax and Development, which provides a more reliable, transparent, and comprehensive basis for cross-national research. This new dataset has initially been used to re-examine major questions about the relationships between tax and aid, elections, economic growth, and democratization. The results deepen some previous conclusions and call others seriously into question—notably the assertion that aid dependence consistently undermines domestic revenue efforts. Above all, the research demonstrates the value of the new dataset, the broader sensitivity of many results to changes in data quality and coverage, and the consequent importance of expanded attention to, and investments in, data quality.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Key words — taxation, cross-country data, aid, resource curse, growth, political budget cycles

1. INTRODUCTION

Concern about the low quality of many economic statistics relating to developing countries has increased greatly in recent years (Jerven, 2013a). Key research findings and policy prescriptions may rest on fundamentally problematic foundations, driving a misleading view of development challenges and, at worst, misguided development policy. The landscape of development research is littered with studies relying on highly incomplete or questionable data, frequently with little acknowledgment of these limitations. Despite some progress, an earlier argument from Herrera and Kapur (2007) appears to still hold significant truth: “Inattentiveness to data quality is, unfortunately, business as usual. . . there are serious weaknesses in many datasets used in cross-country regressions currently in vogue. . . the data sets, problematic or not, become acceptable by repetition. . . There is a certain irony in the fact that a lot of work is devoted to improving methods, but that work on methods does not necessarily translate into improved everyday use of data” (p. 366, 382, 383).

These data weaknesses are particularly acute in relation to data on government revenues. Figures are readily available from the IMF Government Finance Statistics (GFS), and from a growing array of other official international sources. They are widely used in cross-country statistical analysis. However, they suffer major limitations: very extensive missing observations; unaccounted differences between alternative sources; conflicting and ambiguous treatment of revenues from natural resource extraction; and low transparency. In addition, the most common measure of overall revenue collection performance is the ratio of tax collection to GDP. But the GDP data series are themselves sometimes unreliable. In response, an increasing number of researchers have constructed their own revenue datasets on an *ad hoc* basis, often by going back to individual country-level sources and/or by merging data from multiple overlapping international sources. This had achieved some important gains. But it has often also

carried major costs: It has reduced the scope for comparison of results and replication, while many *ad hoc* datasets have themselves suffered from significant errors. Both problems have often been exacerbated by a troubling lack of transparency.

This article correspondingly introduces the new International Centre for Tax and Development Government Revenue Dataset (ICTD GRD). It is the outcome of a four-year process¹ of analyzing data from all available international data sources, along with IMF country reports, developing a standard system for classifying that data, and combining data from mutually compatible sources into a single research dataset. It achieves demonstrably large gains in both completeness and analytical accuracy; and, critically, is both publicly available and transparent in its construction.

Nor are these abstract gains. While improved data quality is desirable in and of itself, it also offers the promise that improved data will drive more reliable research findings. With this in mind, the remainder of the paper summarizes a first round of research findings employing the ICTD GRD, which collectively illustrate that improved data do, in fact, generate new and more robust results. These studies have sought to replicate earlier results in several widely explored research areas: (a) tax and aid, (b) tax revenue, non-tax revenue and democracy, (c) electoral cycles and taxation, and (d) taxation

*Thanks to the International Centre for Tax and Development for funding this project, to Alex Cobham, Andrew Goodall, Fariya Mohiuddin, Vanessa van den Boogaard, Ava-Dayna Sefa and Kyle McNabb for invaluable contributions to constructing the dataset, to Paola Salardi, Paul Segal, Oliver Morrissey, Paul Clist, Samantha Torrance, Kyle McNabb and Philippe Le-May Boucher for leading the initial studies employing the new data, and to Mick Keen, Claudia Dziobek, Mike Seiferling, Jocelyn Pierre and others from the IMF and OECD who offered insight and support over the course of the project.
Final revision accepted: November 6, 2015.

and growth. The results are striking. In two cases access to the new data adds substantial depth to earlier findings. In the other two cases employing the new data yields substantially new findings, to the point of calling earlier findings into question. More simply, the results illustrate, in stark terms, the value of ICTD GRD specifically, but also the broader importance for development research of much greater attention to data quality.

The paper proceeds as follows. The first three sections highlight the limitations of earlier data and the construction of the ICTD GRD. The fourth section provides an extended overview of the initial set of studies employing the new data, highlighting the new research results—and policy implications—that result. The final section notes some remaining limitations of the ICTD GRD, and reflects on the lessons that it provides for efforts to strengthen development data more broadly.

2. THE LIMITS OF INTERNATIONAL SOURCES

Researchers have long been troubled by the weaknesses of cross-country revenue data. The most widely used source prior to the public release of the ICTD GRD was the IMF GFS, which assembles official revenue and expenditure data from governments across the world. However, sharp limitations of the IMF GFS for research purposes have been increasingly recognized, with even IMF researchers turning elsewhere. The most important limitations of the GFS for research purposes fall into three categories: Inadequate data coverage, inconsistent treatment of natural resource revenues, and inconsistent GDP data employed to calculate tax-to-GDP ratios.²

With respect to *data coverage*, the IMF GFS suffers from extensive missing data, as data coverage amounts to only roughly 65 per of potential country-year observations for the period 1990–2010, and declines further for lower-income countries and in earlier years. This missing data are non-random, and often concentrated in countries experiencing instability, weak governance, or a poor relationship with international organizations, and thus risks generating misleading results.³

Additional concerns are raised by *inconsistent and analytically problematic treatment of revenues from the exploitation of non-renewable natural resources*. In general, countries record corporate taxes paid by private sector resource companies as taxes, while recording royalties, export taxes, profit sharing, the profits of state-owned enterprises, and similar revenues, as non-tax revenue. However, while these distinctions may be accurate from a pure accounting standpoint, they do not appear to be strictly adhered to; and, critically, they obscure the analytical distinctions relevant to most academic research.

Researchers are most often interested in the substantive distinction between what Moore (1998) terms “earned” and “unearned” income. “Earned income” refers broadly to non-resources taxes, which are raised on a relatively broad base, are not a payment for a specific service and generally require a measure of negotiation with the population as well as the construction of complex tax bureaucracies. By contrast, “unearned income” refers to natural resource revenues and other non-tax revenues that come from comparatively captive and concentrated sources, making collection relatively low cost and independent of local populations. When researchers speak of “tax revenue” they are most often interested in *non-resource* tax revenue. However, the IMF GFS, like most other sources, does not allow a consistent distinction between resource and non-resource sources of tax revenue.

Individual country examples serve to highlight these distinctions. When employing data from the IMF GFS, Angola reports tax collection ranging from 30% to 50% of GDP depending on the year—among the highest in the world. However, while accurate in an accounting sense, this is deeply misleading for much research: *non-resource* tax revenue amounts to around 5% of GDP in Angola—among the lowest in the world. This type of discrepancy is not uncommon across resource-rich states. In Iran prior to 1990, petroleum revenues were recorded as tax revenue in the IMF GFS. Post-1990, the same petroleum revenues are recorded as non-tax revenue. While the IMF warns explicitly against merging these pre- and post-1990 data series, some researchers have done so. In all such cases there is a major risk of highly misleading data generating unreliable research results.

Finally, many studies that rely on the IMF GFS—as well as other sources—have been undermined by *inconsistent GDP series employed to calculate tax ratios*. The root of these problems lies particularly, though not exclusively, in the irregular rebasing of GDP calculations in much of the developing world, which has resulted in the dramatic underestimation of GDP in many countries prior to rebasing. This has resulted in dramatic increases in recorded GDP—sometimes amounting to 50% or more—following rebasing exercises. Rebasing in Ghana in 2010 resulted in a 60% “increase” in GDP, while rebasing in 2014 in Zambia and Nigeria resulted in “increases” of 25% and 90%, respectively.⁴ These episodes highlight the importance of regular rebasing, but also a much greater risk: international sources reporting GDP for the same country, but using different base years, can result in huge jumps in GDP from one year to the next (Jerven, 2013b), and correspondingly sharp (and entirely illusory) declines in tax-to-GDP ratios as a result.

This has been true, for example, of the IMF International Finance Statistics (IFS). At the time of writing, combining tax data from the IMF GFS with GDP data from the IMF IFS (as has been common) yields a tax to GDP ratio for Ghana of over 20% in 2005 (pre-rebasing) and less than 13% in 2006 (post-rebasing). In Iran GDP similarly became progressively underestimated over the course of the 1980s, leading the tax to GDP ratio reflected by international statistics to rise from 26% in 1982 to 119% in 1989. Meanwhile, data for Iran in 1990 witness two shifts: reliance on an updated and more realistic GDP series, and a move from recording petroleum revenue as tax revenue to recording it as non-tax revenue. The result: a fall in the recorded tax to GDP ratio from 119% to less than 6% in a single year.

These are, of course, extreme examples, but they are representative of more widespread, though less extreme, issues. They are, as importantly, indicative of insufficient attention to data quality in much existing research, and of the need for transparency of research datasets and their construction in order to weed out such potential problems.

3. RESEARCHERS' DATASETS AND THEIR LIMITATIONS

The problems noted so far have not gone entirely unnoticed. In response, researchers have increasingly relied either on regional sources, which often offer more complete data within a smaller subset of countries, or on *ad hoc* researcher datasets, which merge different data sources, clean existing data and/or draw on country-level sources in order to achieve improvements. However, these initiatives have ultimately failed to sufficiently address existing challenges, and highlight the importance of more comprehensive strategies.

Download English Version:

<https://daneshyari.com/en/article/7392847>

Download Persian Version:

<https://daneshyari.com/article/7392847>

[Daneshyari.com](https://daneshyari.com)