

REVIEWS OF DATASETS

Ahmet Izmirliglu

The World Wealth and Income Database (WID)

Review of: The World Wealth and Income Database

Author/Compiler: Facundo Alvaredo, Anthony B. Atkinson, Thomas Piketty,
Emmanuel Saez, and Gabriel Zucman

Repository: <http://www.wid.world/#Home>:

Date of Posting: September 2013

Size: unknown

Licensing:

The World Wealth and Income Database (WID) is an open source online database that presents data series on the distribution of income and wealth in a growing list of countries. Although there are currently only thirty-three countries (mostly from Europe and the Americas) for which data series are available, there are plans to include data series for an additional forty countries in the near future. WID evolved through the international collaborative efforts of over ninety researchers to extend The World Top Incomes Database (WTID) to provide data regarding the entire distribution of income and wealth, as well as the distribution of different forms of capital assets, in the analyzed countries. This represents a significant extension in scope and purpose in WID compared to WTID, which generally focused on the concentration of wealth among the top decile in a given population. As a result, WID continues to serve as a major information hub that fuels both academic research and the raging public debate on income and wealth inequality with a growing array of novel and varied data.¹

Like its predecessor WTID, WID covers data spanning many decades and often covers the late-nineteenth and the majority of the twentieth century for many countries. This is especially useful for investigating the impact of structural changes on income and wealth distributions, which may take several decades to take form. The authors of the database emphasize that this is one of the areas in which WID sets itself apart from other databases on income and wealth distribution, which generally restrict the range of their data to the post-1970s.

Datasets such as WID that aim to capture income and wealth distributions primarily rely on tax statistics to do so, which usually evokes many sighs and raised eyebrows from weary economists—often for good reasons. For one, individuals have an incentive to minimize their tax liabilities. Thus, although incomes reported in tax documents may be satisfactory for administrative tax collection purposes, academic research attempting to identify income and wealth distributions are obviously vulnerable to distortions naturally inherent in such data. The authors of WID addressed this issue by including extensive information on capital and non-capital assets in the construction of their datasets. In addition to alleviating some of the well-known concerns with tax data, this approach also allowed the authors to include income from capital in their investigations of income and wealth distribution.

Sadly, the inclusion of capital and non-capital assets cannot resolve another problem with tax statistics. Namely, governments must first make the tax data available before researchers can use them to construct their datasets. Unfortunately, the authors of the database have not yet acquired permission to use tax data from China, Middle Eastern countries, former Soviet bloc countries, and most African countries. The available data from China relies on household surveys, which severely limits the possibilities for comparative analysis with the data available from other countries. Furthermore, although BRIC countries appear as focal points of future research (with the exception of India which has already provided strong data), no plans to focus on the Middle East or former Soviet bloc countries appear on the expansion agenda for the database (only limited efforts are planned for ten African countries). This means that the majority of the developing world will remain out of the analyses and debates regarding income and wealth distribution in the foreseeable future.

On a more optimistic note, the authors of WID state that during 2016-2017 they will produce Distributional National Accounts (DINA) for the United States. DINA are defined as distributional statistics of pre-tax and post-tax income and wealth consistent with National Income and Product Accounts, and the Flow of Funds of the United States.² In essence, national accounting (the output, expenditure, and income of all the economic actors in a country, such as households, corporations, and government) is concerned with macro-level data. Inequality analysis, on the other hand, focuses on the distributions of income and wealth, but without seeking consistency with the macro-level data of the national accounts. The successful construction of DINA would bridge the gap between the macro data of the national accounts and the micro data of inequality analysis. If the authors of WID are correct, then DINA will

serve as a significant methodological breakthrough that can determine previously unobtainable annual statistics, such as the fraction of total economic growth accruing to the top 10%, or the economic growth per person of the bottom 90% relative to the overall growth.

The simplicity and ease of use of the WID sharply contrasts with the complexity of the information reposed within the database. Data are queried through three categories in the Database section of the website: countries, years, and variables.³ Countries may be selected individually or in any combination, and one can access the data for any interval within the available years. The complexity of the data becomes apparent in the variables category, which is neatly organized into five subcategories: income, wealth, saving, prices, population. Each subcategory holds numerous choices of variables. For instance, within the income subcategory one finds variables such as *Top income shares* (with or without capital gains), *National income*, and *Average income per tax unit* (with or without capital gains), while in the wealth subcategory one can select variables like *Net private wealth*, *Financial assets*, *Non-financial assets*, *Non-profit wealth*, and *Net corporate wealth*, to name just a few. Some of the variables require familiarity with technical terms and economics methodology, such as the *Pareto-Lorenz coefficients*, but most of the variables are self-explanatory.⁴

Upon selection of the desired variables, one can construct the data series by clicking the Retrieve button. If the data series contains less than 100 variables, the series is immediately displayed on the screen with the option of downloading the data series as an Excel workbook. For series consisting of greater than 100 variables, one can only download the series as an Excel workbook. The data in the workbook is presented in two layouts. In Layout A, every column is a variable and the rows consist of countries and years. Units and notes are presented in a separate Excel sheet with the same format, so that one may easily find additional information regarding the process of calculation or estimation of a data point in a given cell in the data series Excel sheet. In Layout B, every column is a country variable and the rows consist only of the years selected. In this layout, units and notes are included in the same excel sheet as the data series.

In addition to the extensive notes and information regarding the choices made by the authors for the calculation of the variables, the authors also provide methodological notes with additional details on the logic and historical development of their processes. These notes are usually submitted by the authors along with the regular updates they submit to the database, and a list of all the methodological notes appears in the Previous News section of the website.⁵

The authors not only make an extremely commendable effort to compile and manage a very difficult database, but they also organize and present the data in a highly accessible and user-friendly manner. The innovative use of data and the ensuing methodological innovations mark a high-point for online datasets.

NOTES

¹ For surveys of some of this literature, see Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez, "Top Incomes in the Long Run of History," *Journal of Economic Literature* 49, no. 1 (2011): 3-71, and Facundo Alvaredo, Anthony B Atkinson, Thomas Piketty, and Emmanuel Saez, "The Top 1 Percent in International and Historical Perspective," *Journal of Economic Perspectives* 27, no. 3 (2013): 3-20. For a glimpse of the public

debates on these topics, see, Neil Irwin, “Thomas Piketty Responds to Criticism of His Data,” *New York Times*, May 29, 2014, <http://www.nytimes.com/2014/05/30/upshot/thomas-piketty-responds-to-criticism-of-his-data.html>.

² Thomas Piketty, Emmanuel Saez, and Gabriel Zucman, “CEG Grant: US Distributional Accounts,” Summary, <http://eml.berkeley.edu/~saez/dina-CEG.pdf>

³ <http://www.wid.world/#Database>.

⁴ The Pareto-Lorenz coefficient is a measure of the concentration of wealth among the rich. For details on its development and application, see, Anthony Atkinson, “Concentration Among the Rich,” Helsinki, Finland: UNU World Inst. for Development Economics Research (UNU/WIDER,) 2006, <https://www.wider.unu.edu/sites/default/files/rp2006-151.pdf> (accessed January 14, 2016.)

⁵ <http://www.wid.world/#PreviousNews>.



Articles in this journal are licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by the University Library System of the University of Pittsburgh as part of its D-Scribe Digital Publishing Program and is cosponsored by the University of Pittsburgh Press.