

Matthew Drwenski

Data on Inequality and the Inequality of Data:

The Last Two Centuries

Abstract

This paper attempts a number of tasks that will further the study of world-historical human inequality, by arguing for a comprehensive understanding of inequality and by informally comparing and aggregating multiple datasets. The paper briefly surveys and critiques the existing corpus of inequality data, noting areas of overlap, opportunities for harmonization of data, and the coverage of the historical information. The inclusion of micro-level data from historical scholarship that is not in communication with the social scientific studies is essential to further the field. The paper concludes with a regional and global narrative of human inequality over the last two centuries.

Matthew Drwenski

Data on Inequality and the Inequality of Data: The Last Two Centuries

Inequality is one of the most discussed problems in contemporary national and global politics. Despite its importance, historical national and global inequality is something about which we know little and understand even less. Inequality data, both contemporary and historical, is best understood at multiple scales, from the local to the planetary, and over long historical time frame. The major roadblock preventing a global, *longue durée* historical analysis of interpersonal inequality over the last two centuries is the lack of data from developing nations and beyond the recent past. The results of this paper take a bold step towards a truly world-historical understanding of inequality.

This paper employs a comprehensive view of human inequality including interpersonal distributions of wages, height and international measures of GDP and income. By compiling these differing datasets, I can develop several new lines of inquiry. What relationships exist between global and local scales of inequality? Can different types of inequality data, such as wages and height, be correlated and extrapolated to extend the spatial or temporal coverage of the data? How can more micro-level historical studies be inserted into these national or macro-level datasets? What broad conclusions can be made based on the aggregation and harmonization of currently-available historical data on inequality? What priorities for further opportunities for archival and methodological improvements exist?

This paper attempts a number of tasks that will further the historical study of human inequality. First, I argue for a comprehensive understanding of inequality, including both economic and social categories of data as well as comparing aggregate and distributional data. Second, I compare several ways of measuring inequality. Third, I briefly survey and critique the existing corpus of inequality data, noting areas of overlap, opportunities for harmonization of data, and coverage of the historical information. Fourth, I advocate for the better inclusion of micro-level data from historical scholarship that is not in communication with the social scientific studies generally associated with the study of inequality and suggest several priorities for further research and sites for methodological improvement. Fifth, I present a tentative narrative of human inequality over the last two centuries.

Definitions and Descriptions

Inequality among humans is entangled with divisions by age, health, geography, citizenship, gender, race, and class. Defining inequality is difficult. Observing the measurable effects on relational categories and difference

among humans is more feasible and more useful for historians.¹ We can study measured inequality even if we don't have a satisfactory definition for what actual inequality among humans is. However, we can say that historical inequality research is the description and analysis of relational differences among human societies and among humans.

Researchers must adopt an encompassing definition of inequality. The pioneering work in distributional studies of well-being was first done by examining national accounts data, tax-returns records and wage surveys. Yet, these types of data only cover a very small fraction of humans over the last two centuries. Even in the 21st-century, over a billion people are outside the datasets constructed using these data.² As one move further back in time, these data narrow even more. Fifty years ago, less than twenty percent of humans are measured this way and one hundred years ago, less than ten percent are covered.³ This means that inequality research must also encompass other measures of social well-being, caloric intake, life expectancy, and even height. Several promising analyses have shown how well-being has been historically correlated with economic inequality.⁴ Encompassing understandings of inequality, which include non-income data, are the only way to construct a global and historical narrative.

Two perspectives are useful when analyzing data that measure inequality. First are data that compare aggregations among national or local administrative units. For example, much of historical information we have on wealth is based on measures of national income per capita, an average. Because of the surfeit of these aggregate measures, and the dearth of more individual records, historical studies of inequality must include them. Indeed, some types of measures of well-being, exposure to pollution, for example, are very difficult if not impossible to refine down past a certain level of aggregation. Second, for historians interested in the interplay among scales, distributional measures must also be considered alongside aggregate measures. Distributional measures of inequality are those that look at the relational difference within a locality or administrative unit.

It is important to briefly note the common ways in which the social sciences have described distributional inequality. The Gini index is a useful and simple starting point for measuring the distribution of any type of inequality. First, it is a single number between 0, representing perfect equality, and 1, meaning perfect inequality. Second, it is invariable to scale and sample size, but sensitive to any redistribution within the system.⁵ Unfortunately, since the index is a single number, it is possible that widely different distributions—distributions with different shapes and skews—can have an equal Gini coefficient. And, the Gini index is not additive, one cannot sum different Gini indices from administrative units together, or subtract out the Gini of specific locality. The Theil index, while less sensitive to any movement within the distribution is additive. Because both of these summary statistics are limited, inequality research must also consider distributions by percentiles, deciles, or quantiles. Historical inequality research must also consider less social scientific descriptions, such as simple counts and divisions of persons by profession, class, or income bracket.

Survey of Existing Inequality Data

Inequality data can come in many forms: probates, tax records, and lists of professions. At present, there are seven overlapping longitudinal databases, compiled by both NGOs and inequality researchers that address the topic of world-historical inequality. They are the Maddison Project, the Luxembourg Income Survey (LIS), the Clio Infra project, the World Bank Open Data (WB), the World Top Incomes Database (WTID), the Socio-Economic Database for Latin America and the Caribbean (SEDLAC), and the UNU-WIDER World Income Inequality Database (WIID).⁶ All seven are briefly summarized in Table 1. First, there are several broad commonalities

among these databases. Each uses national units as the basis for organizing most of the data. It should be noted that each makes no attempt to account for changing national borders and administrative control. In each database, European states and European settled colonies form the majority of the data with less information from states outside of Europe.

Source	What is measured?	Spatial Coverage	Temporal Coverage	Units
World Bank	Gini Indices, GDP, GNI per capita, Population, Gov't Salaries	Global, but focusing on developed nations	1950-2009	National Units
Luxembourg Income Survey	Income deciles	Europe, US, Japan, Australia, and New Zealand	1960-2011, most countries have 5-10 surveys over the time period	National Units
Clio-Infra	Average Height, Height Gini, Income, Income Gini	Global, ~160 countries for the most recent dates, ~40 for the earliest dates	1820, 1850, 1870, 1890, 1910, 1929, 1950, 1960, 1970, 1980, 1990, 2000 (benchmark dates)	National Colonial Units
UNU-WIDER World Income Inequality Database	Mean and median income, income Gini, income deciles, income quantiles	Global, ~90 countries for the latest dates, ~13 for mid-century	1951-2011	National Units
World Top Income Database	Income and wealth shares for top 10, 5, 1, .5, .1, and .01% of the population, Pareto-Lorenz coefficients	Global (European focused, but also including the US, India, Argentina), ~40 countries	~1912 to 2010	National Units
The Maddison Project	GDP per capita estimates, often calculated through backwards projection	Global, ~180 countries	Year 1 to 2010 at benchmark dates that increase in frequency as one approaches the present.	National Units (including estimates for national units before those national units existed)
SEDLAC	Income deciles, Gini indices, wages	Latin America and the Caribbean, 26 countries	Usually ~1980 to 2013	National Units, Metro-area units

Table 1. Summary of the major inequality datasets.

This is especially true for data from farther in the past. As I abstractly illustrate in Figure 1, these seven databases are by no means a comprehensive survey of all of humanity for the last two centuries. However, they do illustrate the need for further creation of longitudinal data for countries outside of Europe as well as an opportunity for the inclusion of other sources of historical information.

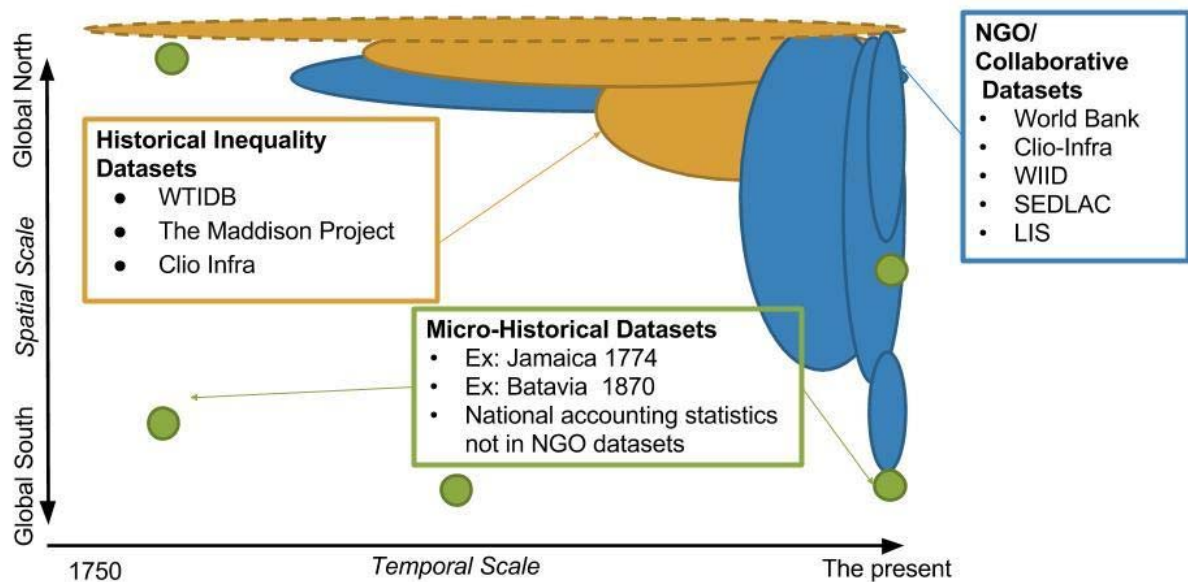


Figure 1. Spatial and temporal coverage of inequality datasets.

Incorporation of micro-data

Despite the drawbacks—the changing boundaries and limited historical coverage—of national data, these measures remain the only reasonable way to begin to construct global estimates.⁷ A consequence of relying on national data is a Eurocentric bias in studying inequality. As one goes back in time coverage for areas outside Western Europe and North America narrows and then disappears. Scholars who have estimated global levels of inequality for the non-West have relied on guesswork and conclusion-crippling assumptions. For example, as seen in the Maddison Project, on GDP growth and change in wealth distribution for China, South Asia, Central Asia, Oceania, and Africa have been assumed to be static, or only changing at fixed rate.⁸ Most of these measures are obtained by projecting recent statistics backwards in time. Clio Infra, building on Maddison’s research has more nuanced estimates for these data-poor regions. One potential work-around is the inclusion of local, micro-economic studies within these global measures.

We must “account for inequality as a complex set of (for example, interactions that occur simultaneously within and between countries) that have unfolded over space and time as a truly *world-historical* phenomenon.”⁹ Both national statistics and local research must be placed in a global framework. When moving from the data-rich 20th-century back towards 1780, national and global estimates must similarly be augmented with more micro-level studies. Recent scholarship has emphasized the interconnectedness of the global economy since at least the late

18th-century and inequality cannot be an exception to a global-local relationship. No attempt has been made to integrate data from local studies, done in traditionally data-poor spaces, into the broad macro-level estimates of global, historical inequality.¹⁰ Micro-level data, from sub-polities, regions, or cities, are also more accurate than estimations and offer a new avenue of connection between scales for analysis. First, micro-level aggregate data can be used to correct and improve historical datasets that have sketchy estimations for most of the world. Second, micro-level data serve as a basis for interpolating distributions for similar localities for which distributions are currently unknown.

An Example Priority for Future Inequality Research

The Caribbean as a stand-in for the world will offer historians of human inequality a revealing site of analysis. The Greater Caribbean, defined here as nations that border the Caribbean Sea or the Gulf of Mexico, would be case-study in the interplay among local, regional, and planetary scales of inequality. The Greater Caribbean before 1950 is missing from all of the major studies on inequality.¹¹ Even though it is a smaller proportion of human population than other regions, the Caribbean is ideal for this study for a number of reasons. The statistical units, both colonial and sovereign states, are remarkably consistent throughout the last two and a half centuries. Caribbean elites of the 17th and early 18th-century also occupied the top-level of the global distribution but have since dropped out of the upper echelons. Throughout the examined period, this regional economy was closely linked to economic changes in North America, Europe, Africa, and Asia. Nearly all of the historical processes that have causal links to inequality were articulated in the Caribbean. For example, both the early 19th-century and mid-20th-century independence movements reshaped the politics of the region. The Caribbean was also the epicenter of the transition from slave to free and semi-free labor, as well as the rise and fall of the plantation agriculture. Millions of Africans, followed by South and East Asian migrated either permanently or temporarily migrated into Caribbean, and more recently, millions have migrated out of the Greater Caribbean to North America and Europe. Not only is economic history Caribbean a microcosm of the last two centuries of global capitalism, its historical dynamics will help us understand the historical mechanisms that create, sustain, and shift inequality at the regional and global scales.

Work done by the Socio-Economic Database for Latin America and the Caribbean (SEDLAC) has surveyed income distributions in the regions, but only for the last two decades.¹² A more in depth view of the Caribbean will allow historians of inequality to move beyond broad and inaccurate extrapolations. For example, Thomas Piketty uses Argentina's distribution of wealth as a stand in for most of Latin America.¹³ When studied over the longer temporal scope, the Caribbean can also serve as a test of two of the most important, but limited, works in the study global inequality.¹⁴ Did the migrations among the more wealthy North Atlantic economies of the late 19th-century that the reason for the convergence of wages among these countries? Without data from other regions with close economic connections to industrializing nations, such as the Caribbean, it is difficult to access the broader conclusions for the world inequality or know if factors outside the North Atlantic affected the convergence that created what would become the Global North today. Are last fifty years of the 20th-century, a period of stability in the distribution of world as a whole unit, caused by institutions of "selective exclusion" of the high-income nations?¹⁵ A needed counterpoint to this argument is an examination of the global system during a period of *changing* distributions. How do the same state and international institutions function when the economies of the Caribbean are rapidly converging or diverging? This would help illuminate whether the flows of permanent and temporary laborers and international capital in the 19th and early 20th centuries do indeed correlate with changes in the global distribution of wealth among-countries.

A World-Historical Narrative through Inequality Data

Despite all of the work that remains to be done, both in the collection and aggregation of inequality data, the seven longitudinal databases and the several micro-historical studies cited in this article, provide the critical mass of inequality data necessary to sketch a rough historical narrative centered on inequality. This narrative is an example of the conclusions my conceptualizations of inequality and my survey of inequality data. Here, I use Patrick Manning's example of a global historical narrative through data as a model, albeit with a more focused description on inequality.¹⁶ Starting with the world at the end of the 18th-century and the turn of the 19th-century, there are some broad similarities in global inequality. At first average GDPs among Europe and its offshoots in Latin America and South Africa are not widely divergent, with the exception of two low-population polities of the United Kingdom and the Netherlands.¹⁷ Data on height and caloric intake also show interregional similarities between Europe and West Africa, and between Europe and East Asia.¹⁸ Though not covered by the data, several micro-level studies on the Caribbean also exist that argue that this region was the most stratified in the world.¹⁹

In the 19th-century, we begin to see strong global divergence in GDP among countries globally and between Western Europe and Eastern Europe.²⁰ When rough estimates are available for South and East Asia at the beginning of the century, along with more accurate data at the end of the century, the total output of these nations lag significantly behind those of Western Europe and Latin America. For height, nutrition, and general standard of living data, there is broad parity between Asia and Europe until the very end of the 19th-century.²¹ However, the Gini coefficients of income, representing inequality within countries continue to show a remarkable similarity across spaces for much of this time period.²² Within locality data for Africa, the Caribbean, and most of Asia is missing from all these datasets. This 19th-century change illustrates the complex relationships among different types of measurable inequality. A growth early in the century in the output in Western Europe leads to a difference in standard of living after a significant time-lag. However, we should be quick to question this as a simple explanation. Changes in the structure of inequality within Asian, African, and Caribbean could contribute to a systemic change in the global inequality regime: one populous region (Europe) as an outlier in every type of measured inequality.

The 20th-century history of inequality is the story of two large changes. One trend is the economic shocks of the first half of the century, the World Wars and the global depression, that greatly reduced within-country inequality, in Europe, much of the Americas, China, and Japan. To a lesser degree somewhat of the opposite trend took place in Africa and the rest of Asia. The second change is a stabilization of inequality among countries (though this obscures the upward movement of the output of many nations) at a historically high level, coupled with the slow but steady increase of the inequality within countries.²³ Height, nutrition, and other standards of living are distributed more unevenly among countries and world regions at the end of the 20th-century than ever before. In this very brief narrative, different types and scales of inequality continue to react with one another as the global inequality regime changed. Among the world regions, the Caribbean, Africa, and Southeast Asia all need comprehensive longitudinal estimates of all the major measures of inequality. In the meantime, an incorporation of micro-level data can bridge the missing information and allow for the construction of world-historical narrative that incorporates the local-to-global and global-to-local consequences of historical change.

NOTES

¹ Charles Tilly, *Durable Inequality* (Berkeley: University of California Press, 1998), 6-9.

² Branko Milanovic, "Global Inequality Recalculated: The Effect of New 2005 PPP Estimates on Global Inequality," Policy Research Working Paper 5061, The World Bank Development Research Group (2009), 13.

³ See datasets in Table 1, or the abstract representation of coverage in Figure 1.

⁴ Robert C. Allen, Tommy Bengtsson, and Martin Dribe, *Living Standards in the Past: New Perspectives on Well-Being in Asia and Europe* (Oxford: Oxford University Press, 2005); François Bourguignon and Christian Morrisson, "Inequality among World Citizens: 1820-1992," *The American Economic Review* 92 (2002).

⁵ Roberto Patricio Korzeniewicz and Timothy Patrick Moran, *Unveiling Inequality: A World Historical Perspective* (New York: Russell Sage Foundation, 2009), 120-2.

⁶ The sources for these datasets are as follows:

Angus Maddison, Jutta Bolt, and Jan Luiten van Zanden, The Maddison-Project (2013 version), accessed January 10, 2015, <http://www.ggd.net/maddison/maddison-project/home.htm>; Clio Infra, Clio Infra: Reconstructing Global Inequality, accessed January 20, 2014, <https://www.clio-infra.eu/>; Janet Gornick, Thierry Kruten, Branko Milanovic, David Leonhardt, and Kevin Quealy, LIS / New York Times Income Distribution Database (2014), accessed January 19, 2015, <http://www.lisdatacenter.org/wp-content/uploads/resources-other-nyt.xlsx>; The World Bank, World DataBank, accessed January 20, 2015, <http://databank.worldbank.org>; Facundo Alvaredo, Anthony B. Atkinson, Thomas Piketty and Emmanuel Saez, The World Top Incomes Database, accessed January 15, 2015, <http://topincomes.g-mond.parisschoolofeconomics.eu/>; Centro de Estudios Distributivos Laborales y Sociales and The World Bank, SEDLAC - Socio-Economic Database for Latin America and the Caribbean, accessed January 20, 2015, <Http://sedlac.econo.unlp.edu.ar/>; United Nations University, UNU-WIDER, "World Income Inequality Database (WIID3.0b)," accessed November 5, 2014, http://www.wider.unu.edu/research/WIID3-0B/en_GB/database/.

⁷ I can imagine a point that micro-level case studies are accumulated to such a degree that older historical estimates are largely abandoned, but that point is very far in the future.

⁸ Angus Maddison, *The World Economy: Historical Statistics*, (Paris: OECD, 2001).

⁹ Roberto Patricio Korzeniewicz and Timothy Patrick Moran, *Unveiling Inequality: A World Historical Perspective* (New York: Russell Sage Foundation, 2009), xviii.

¹⁰ Branko Milanovic, "A short history of global inequality: The past two centuries," in *Explorations in Economic History* 48 (2011), 494-506.

¹¹ Several studies and databases simulate data for the Caribbean region from national economies that are quite different, Argentina, for example from Caribbean economies.

¹² SEDLAC.

¹³ World Top Incomes Database.

¹⁴ Jeffrey G. Williamson and Kevin H. O'Rourke, *Globalization and History: The Evolution of a Nineteenth-Century Atlantic Economy* (Cambridge: MIT Press, 1999); Thomas Piketty, *Capital in the Twenty-First Century* (Harvard University Press, 2014).

¹⁵ Korzeniewicz and Moran, *Unveiling Inequality*, 76-8.

¹⁶ Patrick Manning, *Big Data in History* (New York: Palgrave Macmillan, 2013), 20-5.

¹⁷ The Maddison Project.

¹⁸ Joerg Baten and Matthias Blum, "Human height since 1820," in *How Was Life?: Global Well-being since 1820, Height and Standards of Living in the Global Past*, ed. by J.L. van Zanden et al., (OECD Publishing, 2014), 124-8.

¹⁹ T. G. Burnard, “‘Prodigious riches’: the wealth of Jamaica before the American Revolution,” *Economic History Review*, LIV, 3 (2001), 506-24.

²⁰ The Maddison Project.

²¹ Allen, Bengtsson, and Dribe, *Living Standards in the Past*.

²² Clio Infra, GDP per Capita and Inequality datasets.

²³ Korzeniewicz and Moran, *Unveiling Inequality*, 68-72; Joerg Baten and Michail Moatsos, “Income inequality since 1820,” in *How Was Life?: Global Well-being since 1820, Height and Standards of Living in the Global Past*, ed. by J.L. van Zanden et al., (OECD Publishing, 2014), 205-10.



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.