UNU-WIDER | 30 YEARS OF RESEARCH FOR DEVELOPMENT

WIDER Working Paper 2015/003

The global consumption and income project

An introduction and preliminary findings

Rahul Lahoti,[1] Arjun Jayadev,[2] and Sanjay G. Reddy[3]

January 2015

United Nations University World Institute for Development Economics Research

wider.unu.edu

Abstract: We introduce two separate datasets—the Global Consumption Dataset and the Global Income Dataset—containing an unprecedented portrait of consumption and income of persons over time, within and across countries, around the world. The benchmark version of the dataset presents estimates in purchasing power parity units of monthly real consumption and income for every decile of the population (a 'consumption/income profile') for 133 countries and more than half a century (1960–2012). We describe the construction of the datasets and demonstrate some possible uses by presenting preliminary results concerning the consumption distribution, poverty, and inequality for the world and specific country aggregates.

Keywords: consumption, growth, global income distribution, global poverty, inclusive growth, inequality
JEL classification: B41, C80, I30, I32

Acknowledgements: We are grateful for the important contribution to this project made by Michalis Nikiforos, who among other things, executed much of the work required to construct an earlier version of the database. We are also most grateful to Brandt Weathers for helpful research assistance. We thank participants in a seminar at the United Nations Department of Economic and Social Affairs for their suggestions. We acknowledge support for this project by the Schwartz Center for Economic Policy Analysis (SCEPA) at the New School and Azim Premji University.

Future versions and related materials will be made available on www.globalconsumptionandincomeproject.org.

[1]Department of Economics, Georg-August Universität Göttingen, rahul.lahoti@gmail.com; [2]Department of Economics, University of Massachusetts at Boston and Azim Premji University, arunjayadev@gmail.com; [3]Department of Economics, the New School for Social Research, CUNY Advanced Research Collaborative and Initiative for Policy Dialogue, reddysanjayg@gmail.com

This paper has been presented at the UNU-WIDER 'Inequality—Measurement, Trends, Impacts, and Policies' conference, held 5–6 September 2014 in Helsinki, Finland.

ISSN 1798-7237 ISBN 978-92-9230-888-9

Typescript prepared by Lesley Ellen for UNU-WIDER.

UNU-WIDER gratefully acknowledges the financial contributions to the research programme from the governments of Denmark, Finland, Sweden, and the United Kingdom.

The World Institute for Development Economics Research (WIDER) was established by the United Nations University (UNU) as its first research and training centre and started work in Helsinki, Finland in 1985. The Institute undertakes applied research and policy analysis on structural changes affecting the developing and transitional economies, provides a forum for the advocacy of policies leading to robust, equitable and environmentally sustainable growth, and promotes capacity strengthening and training in the field of economic and social policy-making. Work is carried out by staff researchers and visiting scholars in Helsinki and through networks of collaborating scholars and institutions around the world.

UNU-WIDER, Katajanokanlaituri 6 B, 00160 Helsinki, Finland, wider.unu.edu

1 Introduction

Increases in mean per capita income are often used as a first approximation for a society's economic development. However, it is a metric that is widely recognized to be insufficient for a general description of increasing social well-being. Public debate is increasingly concerned with whether growth experiences are 'delivering' in terms of increasing social well-being. Some recent work has focused on the extent to which gross domestic product (GDP) growth fails as an adequate measure of the social value of economic activity (see for example Stiglitz et al. 2010). Such concerns encompass whether there has been adequate recognition of the ways in which growth has been distributed.

Over the last two decades the increased availability of high quality data has allowed researchers to track the existence and persistence of widespread inequalities both among people within countries and between countries. To date, however, the issues of global and regional inequality and global and regional income growth have by and large been dealt with separately. We describe our effort to create resources that can help address these questions together: that is, to give plausible estimates of the extent to which income and consumption are enjoyed differentially across and within countries and regions over a reasonably long time-span. Specifically, we introduce the Global Consumption and Income Project (GCIP), which has as its foundation the creation of two separate datasets—the Global Consumption Dataset (GCD) and the Global Income Dataset (GID)—containing a portrait of consumption and income of persons over time, within and across countries, and around the world, and aims at analysing these data in future work. The benchmark version of the dataset presents estimates of monthly real consumption and income (in US$2005 PPP) of every decile of the population (a 'consumption/income profile') for the vast majority of countries in the world (more than 130) for every year for more than half a century (1960–2012). The methodology of construction of the dataset allows for comparable data to be presented for an arbitrary number of quantiles.

Using the GCIP one can estimate a Lorenz curve, mean and consumption/income profile for any given year and country or aggregate of countries. This enables us to create a synthetic population from which any poverty measure (headcount ratio, poverty gap ratio, Foster-Greer-Thorbecke measure etc.), inequality measure (Gini coefficient, ratio of mean to median, Palma ratio, Theil index etc.), or measure of inclusiveness of growth and development (anon) can be calculated.

The resulting nearly continuous portrait of the evolution of the world consumption and income pattern is unique. It goes beyond the Penn World Tables in presenting estimates of the distribution of consumption within countries and it goes beyond recent analyses of the world consumption distribution both in greatly extending the period covered and in presenting estimates for every year. Further, whereas with rare exceptions (for example Lakner and Milanovic 2013) such databases and studies based upon them have focused on relative inequalities alone, we provide data on levels of consumption so as to enable assessment of level and distribution together, as is required for analyses in areas such as the inclusivity of growth and development. We have also developed, and intend to publicly provide, in-built tools for filling in missing data and creating portraits of aggregates of countries. Our intent is that the GCIP should meet a high standard of transparency, allowing for third-party replication, modification, and updating and the adoption of alternate assumptions for the selection and treatment of data from the underlying universe, unlike any of the current databases. Among other benefits of such an approach is likely to be that the database can eventually be kept up-to-date through the involvement of multiple users, ensuring that it remains current.

Constructing the dataset involves undertaking several decisions with regard to the selection of data as well as to the manner in which estimates are generated for country-years in which no household survey was undertaken. In a more comprehensive, planned companion paper (Jayadev et al. forthcoming) we document the process of construction and specific choices concerning data in greater detail. Some of the other methods we have developed (e.g. for Lorenz curve estimation and aggregation) and which we intend to make available through freely available software will also be described in further accompanying papers. The current paper briefly describes the methods we have employed in the construction of the benchmark version of the database and presents preliminary results for a few countries and aggregates. Extensions of the primary database (for instance involving quintiles or ventiles rather than deciles or different PPP base years) are created using analogous methods.

2 Existing databases

Ours is certainly not the first dataset that can be used to illuminate issues related to global poverty or inequality. Since the mid-1990s, with the release of the Deininger and Squire dataset (Deininger and Squire 1996), economists have had data on the distribution of income across many countries. This availability in turn has led to greater efforts to try and extend the data (for example, through the World Income Inequality Database (WIID)[1] developed by UNU-WIDER (2008, updated in 2012 and 2014), to harmonize it, as for example with the Standardized World Income Inequality Database (Solt 2009) and to extend the data backwards in time (Pinkovskiy & Sala-i-Martin 2009). The World Bank has been developing global poverty estimates on the basis of its own data collection since the late 1970s, and the World Bank's Povcalnet database has been available to the general public since 2001 as a result of demands for greater data access and transparency by the world public. This institutional collection of data has also been the basis for the influential work of Milanovic (2002, 2005).

Our work seeks to go beyond these efforts in at least four ways. First, we construct estimates of both consumption and of income. It is well known that consumption and income display have different levels for individuals and distributions for populations. We therefore create separate income and consumption estimates for each country-year observation and quantile in the database. Second, as noted above, we collect information on both the level of advantage (income or consumption) as well as the distribution of advantages for quantiles of the population within and across countries and over time, interpolating where necessary to create a complete time-space tableau. Third, we allow for the aggregation of estimates of the level and distribution of income for user-defined regions and groups of countries. This capability relies both on our having previously created estimates which are aligned exactly in time in a given year, through interpolation where necessary. This aspect of our effort therefore builds on the preceding one. We have developed our own software and methods to merge distributions for these user-defined aggregates, providing very useful and flexible capability for researchers and policy makers. Fourth, we aim to provide full documentation of our methods and tools for the ready adoption of alternate assumptions underlying the database and for the ongoing improvement of methods, tools, and data choices through engagement of specialists and the general public.

A recent exercise furthering related objectives is that undertaken by Lakner and Milanovic (2013), which builds upon Milanovic (2005) and seeks to describe the global income distribution between 1988 and 2013, analysing the evolution of levels of income as well as the distribution of

---

[1] World Income Inequality Database Version 2.0c. Accessed May 2014. Available at: http://www.wider.unu.edu/research/Database/en_GB/database/.

income. They choose a few benchmark years and describe the change in the global distribution over the period using surveys based on observations near to those years. We employ a standardized income concept in our comparisons and employ a longer time series, although much of the increased length comes from extrapolation of data. We also, as mentioned, allow for the static and dynamic portrayal of distributions for regional and other groupings of countries, and for other dimensions of variation (such as the use of alternate PPPs, corresponding to distinct base years or other factors).

In another recent exercise Dykstra et al. (2014) queried the Povcalnet database using automated methods to create a cumulative distribution of income or consumption (lumped together in that database) for a large number of survey years (from each of 942 surveys spanning 127 countries over the period 1977 to 2012). The resulting database can (as the GCIP can) be used for diverse purposes, some of which would have been very difficult without downloading the data in this comprehensive way. The exercise highlights the difficulty in accessing even nominally public data for research and replication, the prevalence of poor documentation and the value of fully publicly accessible datasets.

In creating an earlier version of the GCIP, we undertook a very similar exercise. However, we abandoned that effort because: (a) the computational effort for the exercise was very high[2] and the cumulative distribution could simply be replicated for the entire distribution for as many points as desired, and more flexibly and transparently, by replicating the reported parametric regressions that underlay the data; (b) the Povcalnet database is confined to developing countries from the early 1980s onwards; and (c) there was no reason to privilege Povcalnet as a source of survey data even for developing countries. Accordingly, the GCIP differs in key respects. The GCIP has wider area and time coverage (due to inclusion of surveys from other secondary sources), it incorporates a standardized welfare concept (consumption or income) making within- and cross-country comparisons meaningful, it allows for the estimation of all measures for every year (not just the survey year), it provides access to tools for creating user-defined composites of countries in any given year, and provides flexibility in updating the dataset and in choosing specific parametric or non-parametric estimation methods for the Lorenz curve (as opposed to accepting the version which happens to be chosen by Povcalnet, which may reflect not only variable methods but sometimes invalid estimated Lorenz curves). One of the key goals of GCIP is to provide complete documentation, access to data and code, the possibility of applying alternative assumption in database creation or analysis, and transparency of methods.

We do not attempt to discuss the merits and demerits of previous efforts but instead seek to focus on the distinguishing features of this dataset. It is nevertheless useful to clarify the differences between our approach and existing efforts (see Table 1). As is evident, the GCIP provides data for a much wider set of countries and regions as well as concepts than other existing databases. Creating this database of course requires many assumptions and decisions, which we discuss further later.

[2] So much so that in a memorable but regrettable incident we caused the World Bank's computer servers to 'crash' temporarily when we attempted this some years ago.

Table 1: Comparison of various global datasets

Database features	Penn world tables	WIID (Version 2.0C)	SWIID	Povcalnet	GCIP
Coverage by type of country	Both developing and developed countries	Both developing and developed countries	Both developing and developed countries	Developing countries only	Both developing and developed countries
Temporal coverage	1950–2011 (not all countries)	1960–2008	1960–2005	1980–2012	1960–2012
Level of consumption/income	Both, based on national accounts	Only one of consumption or income, and not for all surveys	Neither	Consumption or income only, based on surveys wherever possible	Both, based on surveys and national accounts.
Distribution by quantile	No	Yes	No	Yes, only survey years	Yes, all years
Adjustment of data on distributions to achieve greater comparability	N/A	No	Yes (through econometric estimation of Gini coefficients, adopting LIS as 'gold standard')	No	Yes (through econometrically estimated quintile-specific consumption-income ratio)
Interpolation for non-survey years	Yes	No	No	Not of reported quantiles or means, but implicit in reported poverty estimates	Yes
Flexibility in modifying database according to alternate assumptions	No	No	In certain respects	In certain respects	Transparent about sources and methods so as to be flexible
Inequality measures	No	Gini only	Gini only	Selected, for survey years only	All
Aggregate over countries	Yes	No	No	Yes, but only for poverty measures	Yes, for poverty, inequality and the complete consumption or income profile (arbitrary number of quantile means)

Source: Authors' compilation.

3 Construction of global consumption and income datasets

Constructing a consumption (or income) profile for a given country-year requires two distinct pieces of information: the relative distribution and the mean in that year. These two are sufficient to create a unique profile of actual consumption (or income) levels of each decile in each country-year. We divide the process of creating the database into four distinct steps.

In the first step, we collect data on relative distributions and levels for each country from various existing sources and select a unique set of surveys for the various country-years. Next, we standardize the distributions by converting all distributions that are not already in the required format (consumption or income distributions depending on the database) into estimated equivalents. The selected surveys for country-years consist of both consumption and income surveys. Where surveys of both kinds are available they differ, as the share of income consumed tends to be higher for lower quantiles as compared to higher quantiles. Hence, to make any meaningful comparison among distributions across and within countries and over time, we must transform the distributions into a single type. Although the conceptual case for doing so is strong, this is rarely, if ever done, in international comparisons. In the third step, where necessary, we estimate a consumption mean for the GCD and an income mean for the GID so as to place the means in comparable units. Using the mean and distributional data previously generated, we estimate a Lorenz curve for the survey years (using both standard parametric methods and, where these do not suffice, a method of our own design). Finally for non-survey years we estimate the consumption/income profile by interpolation or extrapolation by using the appropriate per capita growth rate figures from the World Development Indicators (WDI)[3] and to create a time-weighted average of the 'perspectives' on the estimation year that are associated with the nearest survey years. We describe each step in detail below.

3.1 Create the universe of surveys

The GCIP draws data on relative distributions from diverse sources, in particular the World Income Inequality Database (henceforth WIID), World Bank's Povcalnet database and the LIS (previously Luxembourg Income Study).[4] We are committed in principle to integrating historical and contemporary data from all relevant other sources, including country statistical offices, UN agencies, and academic studies and hope that users will help to extend the database in this way in the future. Povcalnet is a collection of surveys from developing countries starting from the early 1980s and is maintained by the World Bank. WIID is a collection of surveys from various other secondary sources compiled by UNU-WIDER. It covers both developed and developing countries and spans the period 1960–2012. Our third source, LIS has harmonized data according to its chosen protocols from primary surveys for over 40 countries mostly from upper and middle income countries.

We initially pursue a 'union approach', seeking to collect all available distributional and level data for the country-years of interest. Note here that we may thus import errors from the original data, although we try to identify and correct egregious errors, as we discuss below. The initial database thus constructed sometimes contains more than one observation for a country-year since multiple household surveys were undertaken in certain country-years and the same data might be reported in multiple sources. The first task is therefore to refine the observations so as

---

3 World Development Indicators. Accessed Feb 1st, 2014. Retrieved from http://data.worldbank.org/data-catalog/world-development-indicators.

4 www.lisdatacenter.org (accessed June 2014).

to arrive at one observation for each country and year. Every survey contained in GCIP is reported as having certain coverage of geographical area, population and age, a certain assigned quality rating, income definition, and unit of analysis. To choose one observation for country-years where there are multiple we apply a lexicographic ordering to a set of selection criteria. The criteria and their sequence in the ordering are based on what we consider important considerations for common usage scenarios for the database. These can be altered if other usage scenarios are envisioned or indeed if users' judgments as to the relevance and importance of specific selection criteria differ from our own.

Before applying the various criteria, we restrict the universe of surveys to only per capita surveys. Per capita surveys are simple to compute and understand and have a corresponding concept in the national accounts. They are also most common in secondary data and used by several other global datasets. The drawback of using them is that they ignore any economies of scale due to household size and composition. Limiting our focus only to per capita surveys also makes them more comparable (even when it is reported that a survey uses an equivalence scale, typically insufficient detail about the scale that was used is presented, making it difficult or impossible to compare distinct surveys meaningfully). For LIS surveys, which report data using an equivalence scale, we obtain data in per capita terms using micro-data.

The lexicographic ordering of various criteria which we employ is as follows: whether a mean is present, type of survey (consumption/income), the nature of the income/consumption definition, database source, area coverage, population coverage, quality as defined in the source database, currency unit, and survey source. As we are interested in both levels and distribution, we prefer surveys with mean information over ones for which means are not reported. For the GCD, which focuses on consumption estimates, we prefer consumption surveys to income surveys (and vice versa for the GID). Among income definition concepts, we prefer concepts that are closer to arriving at total income net of taxes and transfers. The order of preference of income definition concepts employed in the underlying databases, drawing upon the classification scheme and related definitions presented in the WIID, is as follows, from most preferred to least preferred: disposable income, disposable monetary income, gross income, gross monetary income, taxable disposable income, primary income, net earnings, gross earnings, and finally a residual category for concepts that are not fully specified, e.g. we do not know if the reported data refers to net, gross, or disposable income.

Povcalnet and LIS surveys are often compiled using primary data, while WIID is a collection of secondary data. We judge that Povcalnet and LIS may be more rigorously scrutinized and have a smaller probability of transcription or other errors as compared to WIID surveys and hence we prefer these two sources to the WIID. Since LIS surveys have until recently included few if any developing countries and Povcalnet does not include developed countries, the overlap in terms of country-years covered by both of these is small. However, when there is an overlap we prefer Povcalnet to LIS for the reason that this ensures greater internal comparability across developing-country surveys and enables greater external comparability with Povcalnet based estimates for developing countries deriving from other sources (in particular, World Bank poverty estimates). We prefer surveys with broader area and population coverage and surveys deemed higher quality by the source database to others. WIID surveys report a quality rating but Povcalnet and LIS surveys do not report any quality rating. Given that Povcalnet and LIS are constructed using primary data and have stricter inclusion requirements, we assign them the highest quality rating (but it must be remembered that this is only an ordinal characterization). We prefer surveys that report means in local currency units over those which are reported in other units because the method of conversion into international units by the source can often be non-transparent. For the GCIP we prefer surveys in which the source of the survey is known over those for which it is missing. Even after applying all of these criteria we find that some country-years still have multiple surveys. At this stage we pick that survey which leads to the survey source being more compatible with the portrait presented by other years' observations for the same country (especially the nearest survey years for which data are available).

3.2 Standardizing the distributions

Surveys vary widely by the type of achievement measured, which makes comparability between countries difficult. The surveys of interest to us can estimate consumption or income. Furthermore, the definition of income varies widely between surveys (some report gross income, others after tax income and others still wider or narrower categories, often with somewhat obscure definitions. Table 2 presents the various income/consumption concepts used in surveys in GCD with their frequencies, adopting the classification used in the WIID. Atkinson and Brandolini (2001) provide an earlier account of problems of comparability of surveys across countries.

Table 2: Income/consumption concept used in GCD surveys

Income/consumption concept used	Number of surveys	Percentage
Consumption	100	7.5
Consumption/expenditure	482	36
Earnings, gross	19	1.4
Earnings, net	25	1.9
Expenditure	18	1.3
Factor income	1	0.1
Income	243	18.1
Income, disposable	123	9.2
Income, disposable gross	123	9.2
Income, disposable net	42	3.1
Income, gross	102	7.6
Monetary income	2	0.2
Monetary income, disposable	36	2.7
Monetary income, gross	20	1.5
Taxable income	4	0.3
Total	1,340	100

Source: Authors' calculations.

As is well known, the distribution of consumption is expected to be less unequal than the distribution of income. Those concerned with estimating global inequality or poverty almost universally recognize this concern but make no correction for it. Comparing measures of inequality or poverty across countries can therefore be highly misleading. Similarly, aggregating information for groups of countries to obtain a measure of poverty or inequality for, say Sub-Saharan Africa, becomes difficult and results obtained from combining income and consumption-based surveys may lead to misleading results.

One effort to overcome these disparities is the work of Solt (2009) who makes the assumption (plausible at least for developed countries) that the LIS may be treated a 'gold standard' and then tries to adjust other surveys using a regression-based method to estimate a 'standardized' summary measure of the distribution of income (the Gini coefficient) in other countries. We take a different approach here. As it turns out there exist in the WIID database about 120 instances in which there is both consumption and an income survey reported by the same statistical agency in the same year for a country. From the WIID notes we are not able to tell whether, in each case, information on both income and consumption was collected in a single survey or through separate surveys undertaken in the same year.

We use this information to estimate the expected relationship between income and consumption. We begin by employing an extremely simple bivariate regression between income and consumption quintile shares reported to obtain an implied relationship.[5] The regression formula is:

$$CQ_x = \alpha IQ_x + \varepsilon \qquad (1)$$

where IQ is the income quintile share, CQ is the consumption quintile share and x =1, 2, 3, 4, 5 for each quintile.

Table 3: Regression for conversion from income to equivalent consumption quintile shares

Quintile	Coefficient on income quintile (alpha)	Adjusted R-squared of regression	Lower limit of 95% confidence interval	Upper limit of 95% confidence interval
1	1.185	0.89	1.11	1.26
2	1.15	0.95	1.1	1.2
3	1.12	0.97	1.09	1.16
4	1.06	0.99	1.04	1.09
5	0.86	0.98	0.84	0.88
N	120			

Source: Authors' calculations.

As it turns out, there is a very tight relationship observed across the sample between consumption and income quintile shares. Table 3 below provides the details from the regressions. The R-squared for each regression varies from 0.89 to 0.99. In all quintiles, the bounds of the 95 per cent confidence interval lie very close to the estimated mean, giving us confidence that one can reasonably estimate the income share of various quintiles given consumption quintile shares and vice versa. We tried diverse alternate formulations including ones involving regional dummies but found that this did not much improve upon the performance of this basic regression and so did not change it.

We use this regression formula to obtain a derived implied consumption distribution when one has only an income distribution available for a country and a derived implied income distribution when one only has information on the consumption distribution. We undertake this exercise for the whole dataset so that every country can be assigned an income and consumption distribution (at least one original and at most one derived) for every survey year.

5 We plan to further develop our estimation methodology in subsequent revisions. Presently, we are working on replacing these univariate Ordinary Least Squares regressions with Seemingly Unrelated Regressions which would include additional control variables reflecting, for instance, income concepts, regional groups, and country income levels.

Table 4: Stages of the standardization process for Mexico 1989 income survey

Quintile	Original income shares	Implied consumption shares after application of regression coefficients	Implied consumption shares after adjustment for the adding-up constraint
1	3.93	4.66	4.81
2	7.97	9.17	9.46
3	12.28	13.79	14.23
4	19.39	20.61	21.27
5	56.66	48.67	50.23
Sum of shares	100	96.89	100

Source: Authors' calculations.

However, prior to the final assignment we must make an adjustment for the adding-up constraint that the sum of percentage shares in the derived distribution must sum to one hundred. Typically, one is left with income or consumption that is unaccounted for by the simple application of the regression coefficients, for the reason that the regressions were undertaken independently. The sum of shares might be above or below 100. We think it reasonable that the unaccounted for income may be added or subtracted (depending on the direction of the error in the total) proportionally equally across quintiles. This is admittedly only one possible choice: we could apply other rule of apportionment. However, in the absence of compelling reasons to do otherwise, we think this a sound choice. An example of application of this method can be provided for Mexico in 1989. GCIP has an income survey for Mexico for 1989, which we convert to an estimated 'equivalent' consumption distribution. After application of the regression coefficients the sum of the shares of quintiles is 96.89. The unaccounted for share, of 3.11 points is assigned proportionally to all the quintiles so that each quintile's share is increased by the same percentage. The shares at various stages of the process are shown in Table 4.

3.3 Standardizing the means

While there has been substantial interest among researchers in the variance between survey and national accounts means (see Deaton 2005), there has been little or no examination to the best of our knowledge of the variance *between* survey means carried out in the same year for a given country. Our initial examination suggests that these can be extremely wide. For example, Bolivia has two surveys in WIID for 1997 which report monetary income means that differ by 30 percentage points (414 vs. 538 Bolivianos per month). This in turn means that although our lexicographic ordering gives us a particular mean, a slightly different ordering might have led us to choose a dataset with a very different level of income or consumption. This problem will plague any attempt to choose surveys. The mean number of surveys per country-year is 2.95 and the country-years with more than one survey have on average 3.78 surveys. Thirty per cent of country-years have only one survey. In future work, we hope to provide a more comprehensive examination of the issue of disparate survey means. For now, we simply note the problem and attempt to standardize the means for the surveys that our ordering leads us to. As noted before, the universe of surveys provides various definitions of income and consumption. Furthermore, these are often reported in non-comparable units (for example by providing the information in real terms and nominal terms, in local currency or international currency units, and for different time periods). Our next task is therefore to construct a consumption and income mean for every country-year in comparable units. In order to do this, we seek to generate an estimate of the consumption or mean for each country-year for which we have an observation. Whenever an estimate of the mean was available from the survey with which we obtained the relative distribution, this was the preferred source of data. This mean, usually expressed in local currency units (LCUs) of the survey year, was then converted to 2005 LCUs using local consumer price indices wherever available (and in rare cases, where unavailable, the GDP deflator).

In order to make the estimates comparable across countries, we then converted them into common units by applying 2005 PPP exchange rates and converting all data into monthly per capita units (for example if the survey estimate of consumption is for a weekly amount, we multiply it by 30/7). In the future we hope to be able to provide these estimates for diverse PPP base years (1985, 1993, 2005, 2013, and others appearing later) and alternate PPP concepts (e.g. PPPs for income rather than consumption and PPPs constructed in alternate ways).

Outlier detection

Despite our best attempts at corroboration, the survey means data that we are left with contain outliers. These are means that are implausible *prima facie* given other existing data on the subject. We are unclear about the source of the discrepancies given that we use secondary data. We identify outliers using two criteria described below. A survey mean that is identified as an outlier by both the criteria is marked as an outlier and adjusted.

We first run a separate regression for each country to identify the time trend in survey means for that country. In this step, we regress the survey mean with respect to time (years elapsed since 1960). If the survey mean is above or below two studentized residuals from the regression line we mark it as a potential outlier. We find that about 8 per cent of our observations are marked as potential outliers using this criterion. Applying this 'internal' criterion in isolation would mark cases in which a country's economy *actually* experienced sudden growth spurts or severe sharp declines as outliers since the linear time trend may not be able to account for sudden transitions. To avoid this we impose a second 'external' condition, namely that the annualized survey mean growth rate is within some bounds of the national accounts-based growth rate in per capita gross domestic product. The acceptable band for the survey mean growth rate, as currently defined, is between the growth rate of GDP per capita minus plus or minus twice the growth rate. (For instance, if the GDP per capita growth rate is 10 per cent then the band is -10 per cent to +30 per cent). This criterion, while hardly restrictive, helps us to anchor the outlier detection mechanism to a measure of external validation provided by the economy's growth rate. About 60 observations (5 per cent of surveys with means data) are marked as outliers using both the criteria. Instead of completely discarding the outliers we view them as still providing relevant information and therefore adjust them. The outlier means are adjusted (decreased or increased) up to the acceptable outer bounds of the time trend-line. For example, outliers that are higher than the trend-line are adjusted so that they have a value equal to the trend-line plus two studentized residuals. Our reasoning for doing so is that if we were to adjust the means to a higher level they would remain outliers according to our criteria, which would not serve the purpose of adjustment. At the same time, adjusting them to a level lower than the bounds would lead to treating outliers differently from means which are above the adjusted value of the mean but below the outlier detection bounds.

3.4 Generate Lorenz curve and consumption profile

Having obtained or constructed means and distributional data for every survey year chosen, we estimate a Lorenz curve in parametric form using a standard regression framework (see Datt 1998; Miniou and Reddy 2009) for some discussion of the methods, (also employed by Povcalnet). We prefer the generalized quadratic Lorenz curve estimation of Villasenor and Arnold (1989) for its theoretical properties but when the procedure fails to generate a valid Lorenz curve we utilize the Beta Lorenz curve estimation due to Kakwani (1980) as applied to quintiles.[6] When both of these methods fail (very rarely) we create a piecewise linear consumption profile based upon 'connecting the dots' defined by the quantile means, following a method we have developed). We can also calculate the associated Lorenz curve, which is strictly convex (as required for its validity).

Once we arrive at an estimated Lorenz curve, we use it in combination with the estimate of the mean to generate a consumption profile consisting of a mean income or consumption for each decile of the country-year (although in the case of the piecewise linear method for the estimation of the consumption profile, we need not generate a Lorenz curve at all). Specifically, the mean income of each decile is calculated by taking the share of total income accounted for by that decile, and multiplying it by the survey mean times the number of deciles (10). For example if the Lorenz ordinates for the first 2 deciles are 0.02 and 0.05 respectively and the mean income is US$15, then the mean income of the first decile is US$15*10*.02=US$3, while the mean income of the second decile is US$15*10*(.05-.02)=US$4.5.

Our goal is to estimate the consumption profile or set of quantile means for every country-year for the entire period covered by our database in order to obtain a 'consumption profile tableau'. In order to attempt to fill in the consumption profile tableau, we estimate the profile for intermediate years using growth rate figures from the WDI in order to interpolate or extrapolate consumption or income profiles for non-survey years. As noted below, the survey coverage is very limited before 1980. This is why several researchers prefer to begin their empirical efforts after that date. Moreover, whether before or after that date they typically confine themselves to survey year estimates, which may not be temporally aligned across countries, thus limiting the possibilities for comparison and aggregation across countries. However, we are interested in trying to extend coverage as fully as possible, so as to facilitate these tasks. We fully recognize the concerns that such extension may raise, and accordingly try to do so according to carefully chosen assumptions. A substantial amount of the data before 1980 is extrapolated and thus has to be treated with special caution.

There are two methods used to calculate the consumption profile for the non-survey year, viz.

Extrapolation

If the non-survey year lies before or after the first/last survey year for which we have consumption or income profile, then the consumption or income profile of that year is extrapolated (forwards or backwards) based on the survey year and the relevant per capita growth rates. For example, if we want to estimate the consumption or income profile for a country and the last survey year happens to be in a given prior year, then for the subsequent years, we extrapolate the consumption profile using the following formula iteratively:

$$M_t = M_{t-1} * (1 + g) \tag{2}$$

where M is the estimated mean consumption/income of a decile, t is the year and g is the growth rate of mean consumption/income per capita between the two years.

[6] In practice, when generating a valid Lorenz Curve, both procedures typically provide a reasonably good fit to the data. The Beta Lorenz curve fails the test of giving rise to a valid Lorenz Curve more often.

Interpolation

If the non-survey year lies between two survey years for which we have the consumption or income profile, the consumption or income profile for this non-survey year is a time-weighted average of the growth-adjusted consumption or income profiles (arrived at by extrapolating respectively backwards and forwards through applying the observed growth rates of mean per capita consumption or income) of the two survey years. This procedure is the same as described in Chen and Ravallion (2004) to impute means for non-survey years except that we extend the procedure to the overall distribution and estimate decile means in an analogous manner.

Since the consumption/income profiles for survey years are already expressed in comparable units (US$2005 PPP in the benchmark version of the database) we therefore use the growth rates of real (inflation adjusted) per capita consumption to arrive at an estimated consumption profile for each non-survey year and similarly use the growth rates of real per capita income to arrive at an estimated income profile for each non-survey year. For consumption, as our primary source, we use the growth rates of real 'per capita final consumption expenditure, etc.' from the WDI. When this is not available, we use the growth rate of per capita real GDP in LCU from the WDI. If neither of these is available, we use the growth rates of real per capita GDP (US$2005 PPP) from the Penn World Tables and Total Economy Database (Total Economy Database, The Conference Board 2010[7]). For income, we use the growth rate of real per capita GDP from the WDI.

The earliest year to which we extrapolate our data backwards is 1960. This is because annual growth rates of mean consumption from national accounts for a wide variety of countries are available only starting then. In some cases (typically the ex-Soviet countries) the earliest year available is 1991. Other cases in which the earliest available year is after 1960 are as follows: Djibouti (1971), Lao (1971), Mali (1967), and Swaziland (1971). The result in all of these cases is that there are gaps in the tableau. This not only affects the ability to define trends over the entire period but also to construct regional or global aggregates which are fully comparable over time. We seek to fill these gaps over time, in part by drawing on broad public participation. In the meantime, one option is to discard from consideration those entities for which we do not have data over a sufficient period and another is to restrict the temporal scope of the analysis. For certain purposes, it may be tenable to compare alternatives which both do and do not contain certain countries, but one must be aware of the potential distortions arising from this source. The empirical examples we provide in this paper do not include any adjustments for such non-comparability.

[7] http://www.conference-board.org/data/economydatabase/

Table 5: Summary statistics for surveys in Global Consumption Database

	All surveys (1960–2012)	1960s	1970s	1980s	1990s	2001–2012
# of country-year observations	1340	67	67	196	444	566
# of countries[8]	133	35	39	85	121	122
% consumption surveys	45	16	12	29	46	57
% with all area coverage	97	94	97	92	97	99
% with all population coverage	92	58	63	86	96	98
% surveys with means data	82	30	42	69	85	95
# of countries with no means	0	125	116	67	17	11
Database source (%)						
LIS	13	3	15	14	13	14
Povcalnet	62	0	1	25	41	75
WIID	38	97	84	60	46	11

Source: Authors' calculations.

3.5 Description of Global Consumption Database surveys

Tables 5 to 7 present summary statistics for the set of surveys in the GCD. The total number of surveys is 1,340 over the 52-year period, from 133 countries in the world. About 45 per cent of the surveys are consumption surveys and more than 90 per cent of surveys are nationally representative and cover the entire population. The coverage of surveys is sparse in the 1960s and 1970s with less than 40 countries with surveys in each of these decades. The number of countries with at least one survey and the number of surveys with information on means both increase steadily in each decade, with rapid growth from the 1970s through the 1990s. Povcalnet is our biggest source of survey information, accounting for 62 per cent of surveys in the GCD, followed by WIID (38 per cent) and LIS (13 per cent). However, Povcalnet has almost no surveys in the first two decades, for which we instead rely heavily on WIID and to a lesser extent on LIS.

The density of surveys is lowest among low income countries (10 per cent of all country-years have surveys) and highest among upper income countries (25 per cent of all country-years have surveys). For all income groupings, this density is lowest in earlier decades (the 1960s and 1970s) and highest in the recent period (1990s and later), see Table 6. We observe a similar pattern when we examine the evolution of the density of surveys by region. Latin America and the Caribbean has the highest density of surveys over the entire period among all the regions while Sub-Saharan Africa has the lowest density of surveys, see Table 7. For all income and categories the density of surveys in the 1960s and 1970s is low (always less than 10 per cent of country-years) and for regional categories in the same period it is a little higher but still low (a high of 30 per cent for South Asia in the 1960s). Care must be taken in interpreting these numbers, however. An average of a single survey for each country in a region during a decade will result in

[8] In any decade, not all countries have surveys. Hence the total number of countries in the database (133) is greater than those represented in any one decade.

a density indicator of 10 per cent for that country, but that might still suffice to conclude something about living standards in the countries concerned in the decade in question.[9]

Table 6: Availability of surveys by income group[10] and time period

Countries by income Group	All surveys (1960–2012)	1960s	1970s	1980s	1990s	2001–2012
Low income	26	6	3	9	21	25
Lower middle income	34	7	6	19	33	32
Upper middle income	37	12	14	26	32	33
High income	36	10	16	31	35	32
Density of Surveys (# of surveys / # of country-years)						
Low income	0.10	0.03	0.01	0.05	0.19	0.20
Lower middle income	0.19	0.06	0.03	0.13	0.34	0.32
Upper middle income	0.25	0.07	0.06	0.21	0.40	0.44
High income	0.21	0.04	0.08	0.17	0.37	0.32

Source: Authors' calculations.

3.6 Aggregation module

We have developed a module that can be used to obtain a consumption profile for an arbitrary grouping of countries. This helps us determine trends in poverty, inequality, or growth in consumption or income for a set of countries defined by region, income level, association membership, or indeed any other criteria of interest. These patterns can be juxtaposed with individual country experiences to understand how the set of countries is performing. We can perform various analytical exercises with data aggregated in this way such as decomposing contributions to levels (or changes in) inequality, poverty, or other statistics into within-country and between-country components. The evolution of a group of countries can be surprising as it necessarily reflects the relative growth performance of different countries as well as their internal distributional dynamics. For instance, the evolution of inequality within a region (such as Latin America in recent years) may for this reason be different from what might be suggested by the evolution of inequality within individual countries. Several Latin American countries have experienced a dramatic decline in inequality in recent years: between 2000 and 2010, according to our estimates, the consumption Gini coefficient for Brazil has dropped 8 points (from 53 to 45). Chile's consumption Gini coefficient has dropped by 7 points (from 50 to 43). Contrastingly, our estimates indicate that in the same period the overall consumption Gini coefficient for Latin America and the Caribbean has dropped only by 4 points (from 51 to 47). This is because of the contribution of differential growth rates of different countries, which might not be apparent at first. This is an insight only made possible by looking at the composite of countries, as we are able to do. A few illustrative examples of applications of the aggregation module are provided in the results section. Here we briefly describe the method used to combine countries and obtain a single consumption/income profile for the set of countries.

[9] The figures reported here do not take note of country unifications and splits in order to facilitate inter-temporal comparison. For this reason, the density indicator is more meaningful than the absolute number of surveys.

[10] Countries are classified according to World Bank's Income groupings as of early 2014.

Table 7: Availability of surveys by region and time period

Countries by region	All surveys (1960–2012)	1960s	1970s	1980s	1990s	2001–2012
East Asia & Pacific	14	2	6	8	13	12
Europe & Central Asia	44	8	11	33	42	42
Latin America & Caribbean	24	11	8	19	22	20
Middle East & North Africa	10	2	4	6	9	9
North America	2	0	2	2	2	2
South Asia	5	4	4	5	5	5
Sub-Saharan Africa	34	8	4	12	28	32
Density by region						
East Asia & Pacific	0.16	0.03	0.08	0.21	0.26	0.21
Europe & Central Asia	0.23	0.05	0.05	0.15	0.39	0.45
Latin America & Caribbean	0.30	0.07	0.05	0.23	0.53	0.51
Middle East & North Africa	0.11	0.03	0.05	0.08	0.20	0.15
North America	0.21	0.00	0.20	0.15	0.35	0.31
South Asia	0.27	0.30	0.14	0.28	0.36	0.26
Sub-Saharan Africa	0.09	0.03	0.01	0.06	0.19	0.16
Total	0.19	0.05	0.05	0.15	0.33	0.33

Source: Authors' calculations.

We first obtain a consumption profile for all the individual countries within the grouping of countries and for a given year using the procedure described in previous sections. Next, employing a 'poverty-line sweep' method, we obtain consumption levels for the 0.5 and 1.5 percentiles of the group. Specifically, we start at an arbitrary income/consumption level and calculate the percentage of population of each country that has income/consumption below this level. Then, using the population share of each country in the aggregate grouping we obtain the percentage of the group population at this level. We adjust the level and iterate until we obtain the income/consumption level below which the desired percentage of the group population lies, to a specified level of tolerance. Using the 0.5 and 1.5 percentile income/consumption levels as starting points, we then raise the income/consumption level progressively in steps to obtain income/consumption levels at just over 100 points along the spectrum, using error corrections to adjust the size of the steps as we proceed so as to arrive at points within every or nearly every percentile interval. The resulting set of percentile points and the corresponding income/consumption levels are then connected linearly to obtain a consumption profile and to create a 'synthetic population', i.e. a model population with the requisite profile. Using the synthetic population we can calculate any poverty or inequality measure, measure of inclusivity of growth etc. which we may wish to calculate for the group.

4 Preliminary results

In this section, by way of conclusion we provide a few figures and tables that offer more specific indications of the kinds of analysis that are possible with the dataset. We limit ourselves here to data from the global consumption distribution.

4.1 Evolution of world consumption distribution

Figure 1 (density functions) shows the evolution of the world consumption distribution in three 'snapshot' years, 1960, 1980, and 2010. The figure shows twin peaks in the 1980s (identified by Quah 1996 among others). However, the period since then has seen the transformation of world consumption from a bimodal to a unimodal distribution and one in which the overall distribution has narrowed. This is undoubtedly largely due to China's growth in the period but also because of the rapid growth in India in the 2000s. Given the paucity of surveys in the 1960s, the log density function for that year should be viewed with caution. Even so, the data suggests that the distribution in the 1960s was also relatively unimodal, putting a different light on the pattern of evolution of the world distribution in later years that has been discussed by others. The factors underlining the changing world distribution are underlined rather dramatically if one looks at the evolution of the global Gini coefficient including and excluding China as in Figure 2 (Gini coefficient with and without China). Rapid Chinese growth and the country's large population have meant that the global consumption Gini coefficient has fallen monotonically from its peak of .71 in the 1970s and 1980, to a low of 0.63 in 2010. However, excluding China suggests an altogether different picture. Without China, global consumption inequality rose sharply to a maximum in 2000 before declining moderately in the last decade (presumably due to the rapid growth in the other country with a giant population, India). Given the paucity of coverage in earlier years, there are several countries for which the only distributional data are interpolated backwards from later surveys. Additionally, some countries that emerged from the break-up of the Soviet Union are not included in the data prior to 1990. There are several other examples of countries for which we do not have reliable data (for example, East Germany prior to reunification or Cuba). We hope to try and acquire such information in future versions of the database, drawing on specialist and public engagement.

4.2 Inequality in India and China taken together

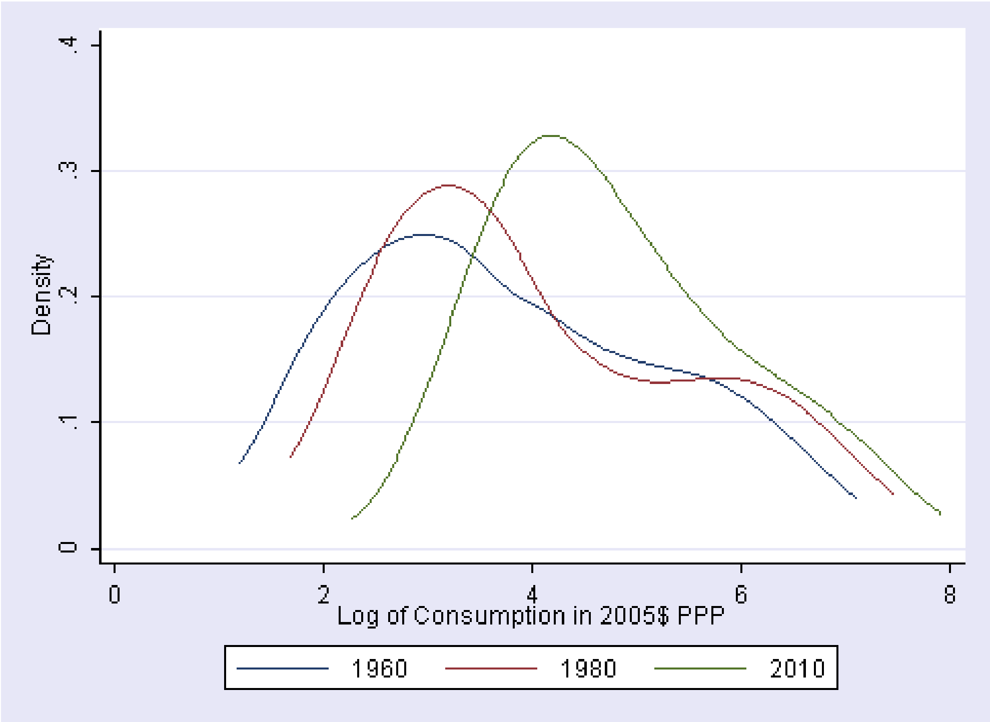
Of course, as others have noted, this has come at a time of increasing inequality in both India and China.[11] One of the advantages of the GCD is its flexible aggregation module and its capacity to straightforwardly generate a consumption distribution for any multi-country aggregate.[12] In Figure 3 (Inequality in India-China), we show how inequality has changed in the aggregate of China and India together. In order to do so, Chinese income surveys have been transformed into equivalent estimated consumption surveys as described earlier. The Theil Index for the India-China composite suggests some very interesting patterns. First, in the period 1980–1990, one observes a decline in inequality followed by a rise thereafter. We may speculate that this pattern can be linked to the more equalizing (than subsequently) and rapid growth widely characterized as having taken place in the early 1980s in China (a period in which the rate of poverty reduction was extremely high) and in the mid-1980s in India. Since that decade, however, in both countries inequality has risen and that also is the case for the aggregate since in 2010 the Theil registers a higher level of inequality than in any other period.

---

11 Rising interpersonal inequality in India has been (somewhat) disputed. In China, it is uncontested.

12 Some computational time and power can be required, however, especially for aggregates involving a large number of countries.

Figure 1: Global consumption density for select years

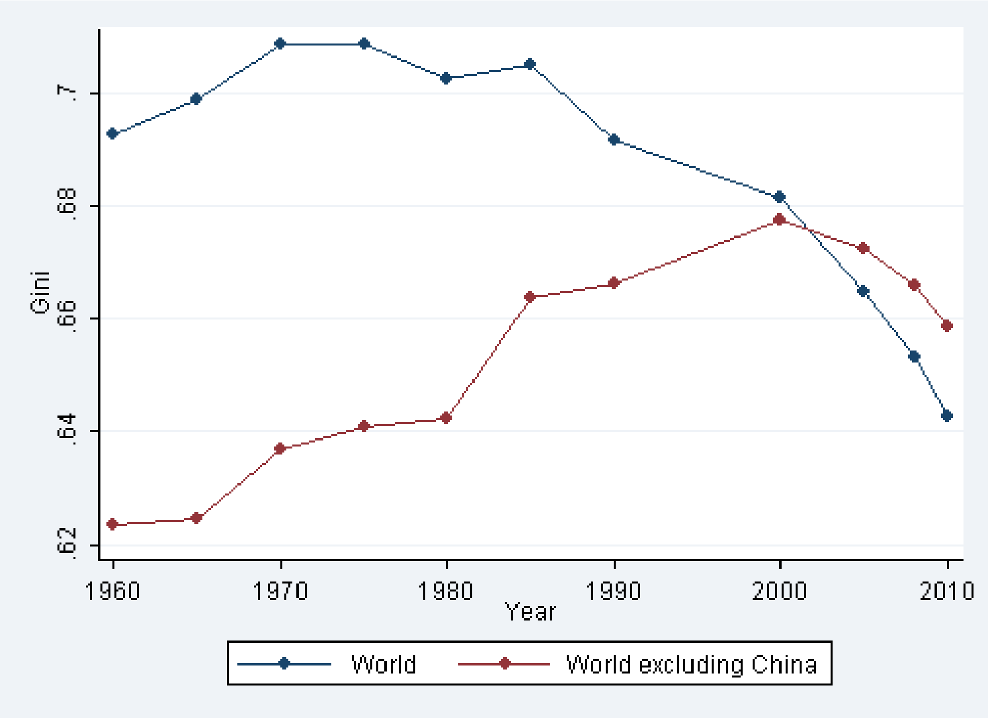


Source: Authors' calculations.

4.3 Poverty headcount ratios

Since we are interested in the use of the data for levels as well as distribution, the database can also be used to assess global poverty trends, using the US$2.5 and US$1.25 dollar a day (2005 PPP of consumption) poverty lines popularized by the World Bank. Figure 4 (poverty trends) depicts the fall in the headcount ratio since 1980. As is evident, there is a sharp decline starting in the 1980s, again initially propelled by growth in China and later in India. In 2010, the estimated headcount ratio according to this measure stands at 17 per cent, corresponding to about 1.2 billion people in absolute poverty. The remarkable impact of China is highlighted by the portrait of poverty reduction in the world (according to the World Bank's measures) with and without China. We note that our estimates differ from those provided by the World Bank for a number of reasons. We explicitly harmonize surveys to reflect the 'consumption concept'. In comparison to the World Bank's earlier method, which scaled down all quantile income estimate by the consumption to income ratio in the national accounts, this would tend to lower poverty estimates by raising estimated consumption for lower quantiles when income surveys are the source. In comparison to the World Bank's current method, which makes no such adjustments, the impact is unknown. In addition, we report the poverty headcount ratio across the world, rather than only for developing countries. The World Bank estimates that 22 per cent of the developing countries' population lived below US$1.25/day in 2008. We estimate that 18 per cent of *the world's* population (1.21 billion out of 6.7 billion) lived below US$1.25/day in 2008. If we assume that none of the people in the developed world are poor by this definition, then we arrive at a headcount ratio of 21 per cent for the developing world, which is very close to the World World Bank estimate. In the future, we will directly estimate a poverty rate for the developing countries and for the developed countries for better comparison, rather than axiomatically assuming (as does the World Bank) that there are no poor in the developed countries.

Figure 2: Global Gini coefficient with and without China

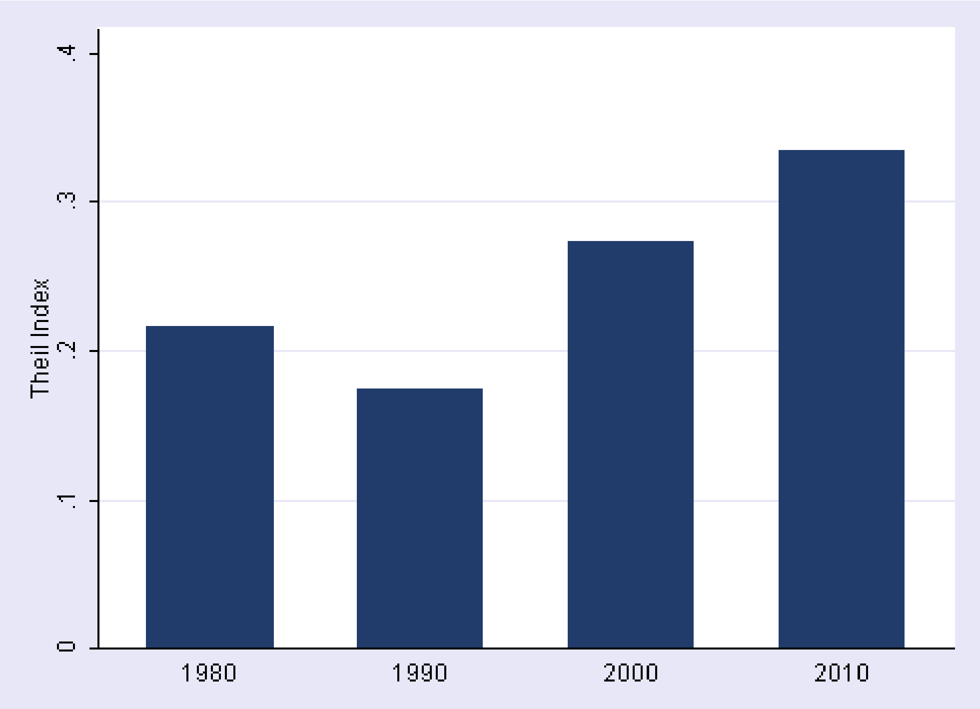


Source: Authors' calculations.

When a survey provides only income (as is the case with China), as noted above we create a synthetic 'consumption' survey, which reflects a regression-based estimate of what consumption may have looked like in that country-year. The World Bank mixes consumption and income surveys in its estimates of poverty because they do not (any longer) make adjustments for survey type.[13] Our procedure allows us to harmonize concepts that allow for cross-country comparison but is only as good as the validity of the regression model for out-of-sample prediction. However, as noted above the good performance of the regression we use gives us some confidence. The World Bank procedure on the other hand relies only on raw survey data or grouped data as reported by national statistical offices (as in the case of China) and may be better in that sense, but strictly speaking, neither poverty nor inequality across countries can be meaningfully compared (or therefore constructed for aggregates) when the achievement concepts are different. An example of the difficulty the World Bank's current procedure entails is offered by Peru, which is one of the few countries that has both a consumption and income survey available for a year and which appears in Povcalnet. In 1997, the headcount ratio for poverty when estimated on the basis of the income survey was 13.8 per cent while it was less than 1 per cent when estimated using the consumption survey, despite employing the same poverty-line concept (US$1.25 PPP per day of consumption).

[13] Correspondence with Shaohua Chen (16 June 2014): The previous approach of the World Bank as documented in published papers was to uniformly adjust income levels from income surveys downward by multiplying by the overall consumption to income ratio in the national income accounts. The present approach appears to be motivated by the idea that the higher mean in income surveys is roughly counter-balanced for lower quantiles by the smaller proportion of total income attributed to these quantiles. It is important to note, however, that even if income surveys are employed without any adjustment, they are in principle being used to *estimate* mean consumption, since the Povcalnet database is constructed to be employed in the estimation of poverty using a poverty line (e.g. US$1.25 PPP) defined in units of consumption.

Figure 3: Theil index for India-China composite for select years



Source: Authors' calculations.

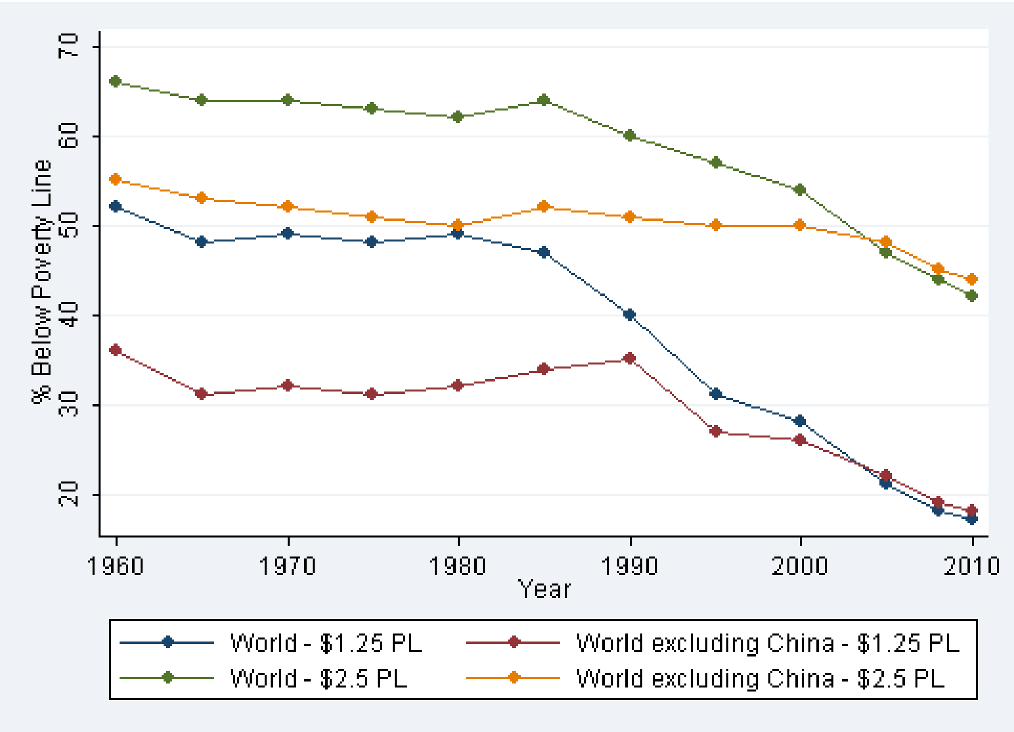
4.4 Proportions of country populations in various quintiles

China's acceleration relative to the world can be depicted in yet another illuminating way as in Figure 5 (country's population in bottom quintile of world). The panel shows the proportion of a country's population that was in the bottom quintile of the world's consumption distribution over time for a selection of countries. In 1980 over 50 per cent of the Chinese population lay in this group. By contrast, by 2010 less than 20 per cent (the horizontal red line) were part of the bottom 20 per cent of the world. Other developing countries have now occupied the space left behind by China. India, notably now has about 40 per cent of its population in the lowest quintile of the world consumption distribution.

Again, one must be careful about the data prior to 1980 since for China there were no surveys before that period and it has a large effect on this calculation. There is also no adequate data for many other centrally planned economies. With these points underlined, however, it is still striking to note the main process of China overtaking all other developing countries in this respect during the period.

Figures 6a-c show the proportion of a country's population that was in the top decile of the world's consumption distribution over time for a selection of countries. The rich countries as expected always have a high proportion. (We should note that our figures for the income of the top decile are based only on our surveys and do not at this stage include further adjustments for other sources of information on top incomes, although we would like to extend the database in the future in this way). Here, the USA as a rich country with a large population has the lion's share of the world's top decile: throughout the period around 70–80 per cent of its population are in this category. As inequality has begun to rise in the USA and other OECD countries have experienced sustained growth over the decades, a larger proportion of other OECD populations now inhabit the top 10 per cent of the world's consumption distribution. A very small proportion of non-OECD populations are even now in the top 10 per cent of the world's consumption distribution. OECD countries such as Spain, Italy, and Korea that were once relatively poorer now have about 40 per cent of their population in the richest decile, as a result of national economic growth. Korea's extraordinarily rapid growth over the bulk of the period makes it the only really new entrant into the club, offering a window that goes beyond population averages, on the difficulty of breaking into the rich country club and the exceptional nature of Korea's achievement.

Figure 4: Global headcount ratios for various poverty lines



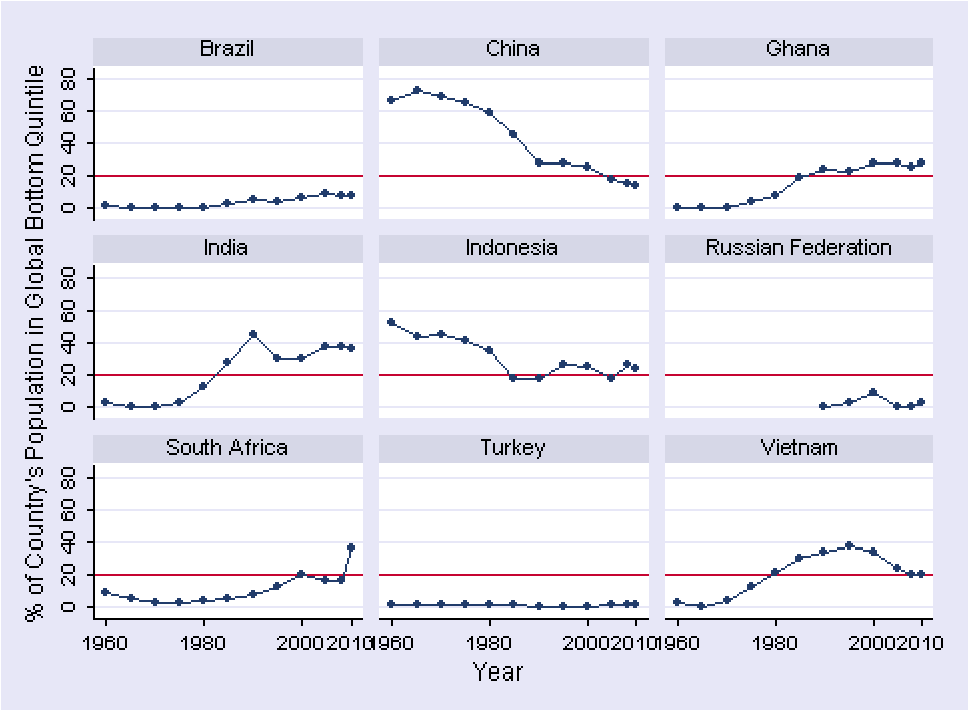
Source: Authors' calculations.

Even for the relatively rich within countries, to remain among, let alone, join the global relatively rich is no small achievement, and requires growth in national incomes as well as maintenance of their relative positions within nations.

Despite its exceptional influence on the world consumption distribution, China remains a decidedly middle-/lower-middle-income country across its population, and less than 1 per cent of its population has consumption levels in the top decile of the world population. In highly unequal and slightly richer countries such as Brazil and the Russian Federation, non-negligible fractions—between 5 per cent and 10 per cent of the population—enjoy rich country level incomes.

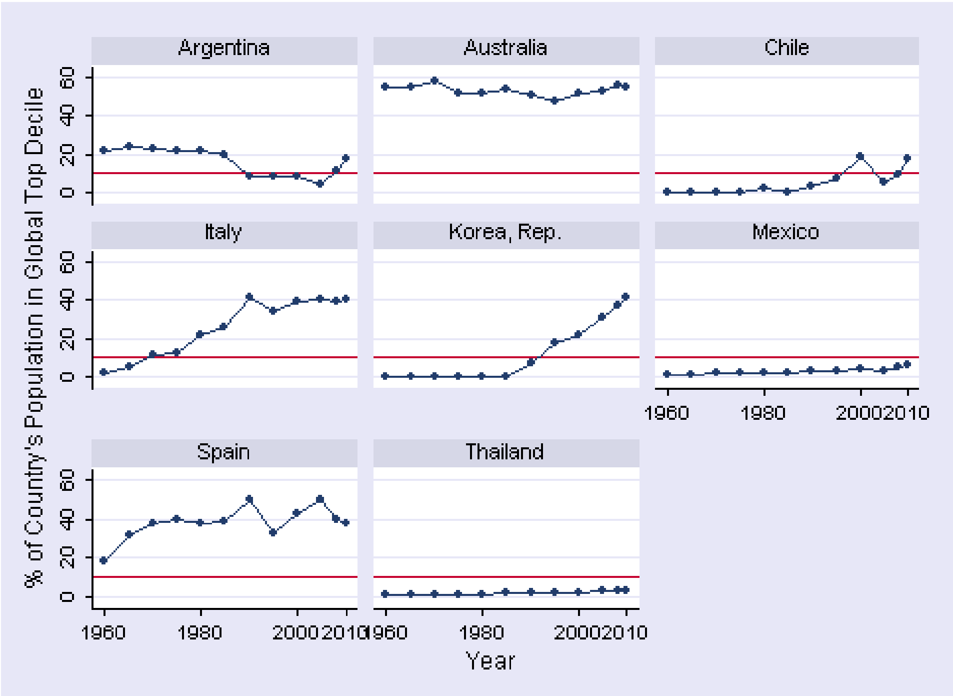
We can also look at the emergence of a group of rapidly growing 'emerging countries' such as the BRICS using the GCD, exploring the ways in which growth has been experienced differently across the income distribution. While all of these have been relatively rapidly growing economies, Figure 7 shows that the lion's share of growth has occurred at the top decile. In these countries taken as a whole, inequality has risen and growth is led by growth in the consumption of the relatively rich. The mean to median ratio has also increased markedly, though less dramatically.

Figure 5: Proportion of country's population in bottom quintile of global consumption



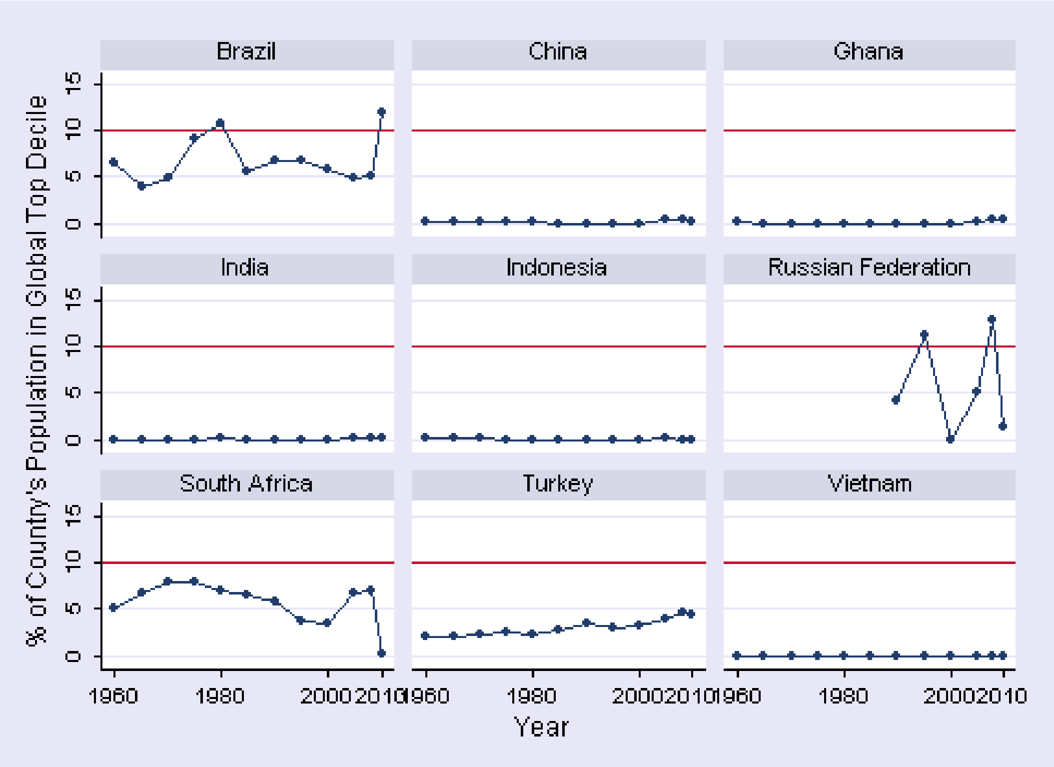
Source: Authors' calculations.

Figure 6a: Proportion of country's population in the top decile of global consumption



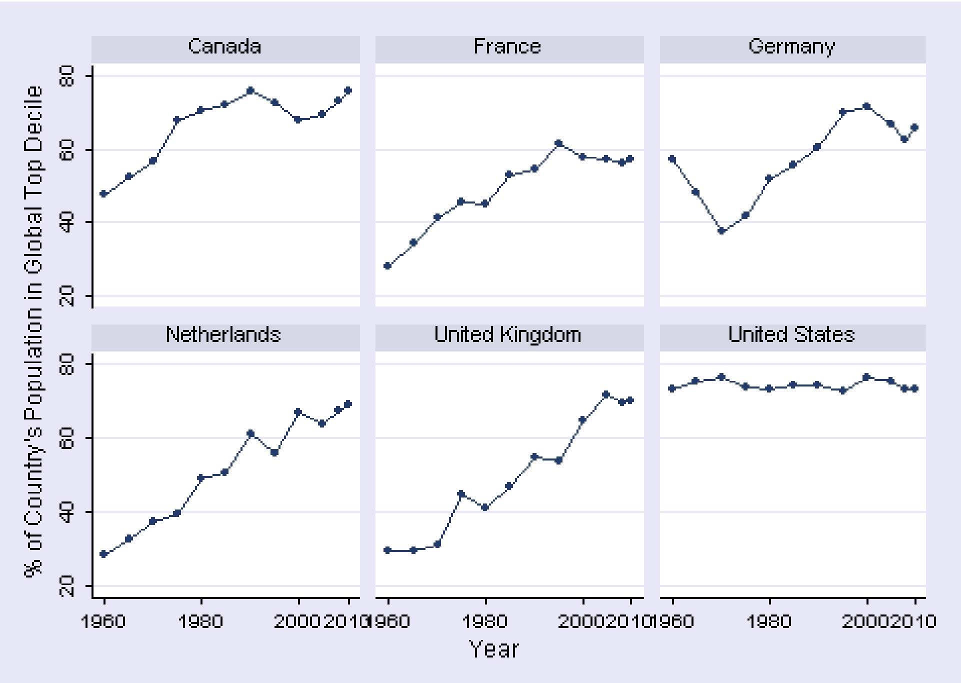
Source: Authors' calculations.

Figure 6b: Proportion of country's population in the top decile of global consumption



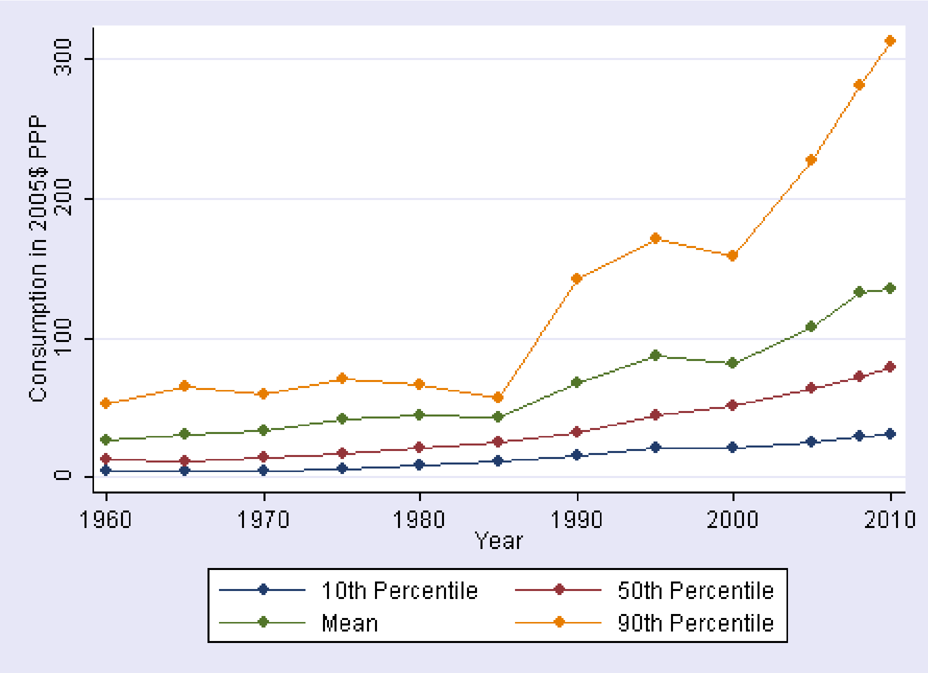
Source: Authors' calculations.

Figure 6c: Proportion of country's population in the top decile of global consumption



Source: Authors' calculations.

Figure 7: Consumption profile for BRICS composite

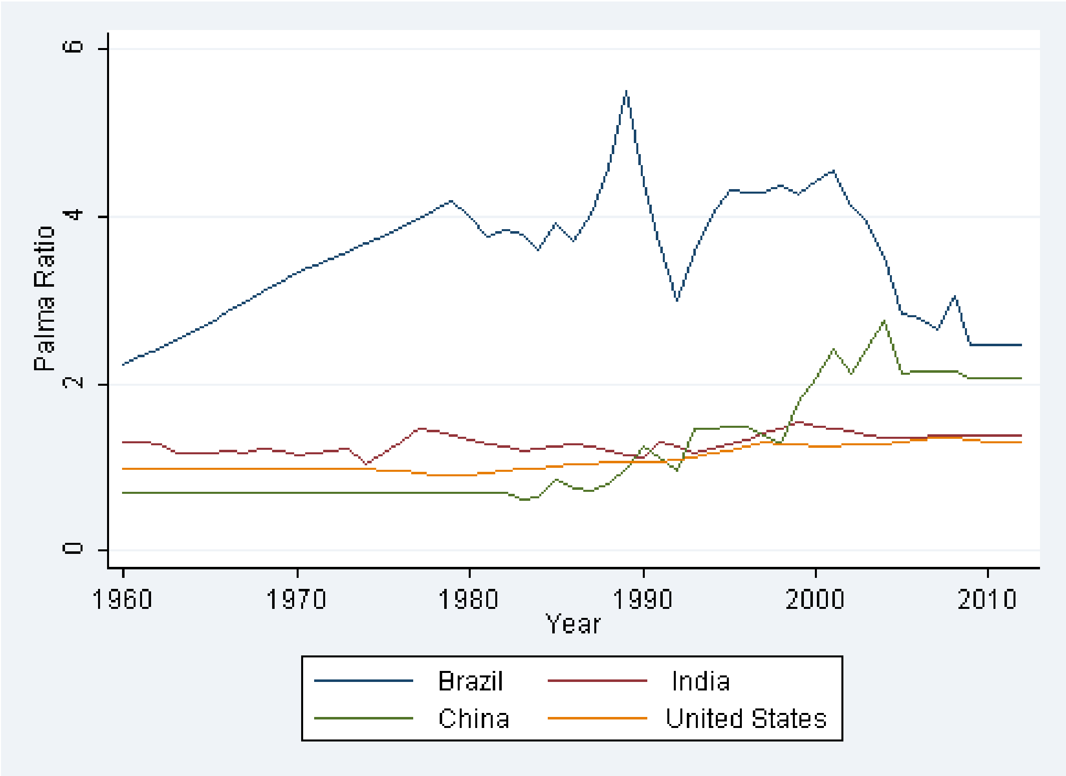


Source: Authors' calculations.

4.5 The Palma ratio: examples

Finally, as we have noted, while we have been describing standard measures of inequality such as the Gini coefficient and Theil index, because of our approach involving synthetic populations, the GCD can be easily used to produce any other desired measure of inequality. In Figure 8 below we show the evolution of the 'Palma ratio' (the ratio of the share of the top 10 per cent to that of the bottom 40 per cent) for the USA, China, India, and Brazil from 1960 to 2010. Inequality in Brazil, while beginning at a very high level, has, if we focus on this ratio, declined substantially. Contrastingly, in the case of China, the USA, and India, there have been increases, although, apart from the first, the change appears to be relatively modest compared to that in Brazil.

Figure 8. Palma ratio 1960–2013: selected economies



Source: Authors' calculations.

5 Conclusion

The lottery of birth—to whom one is born, when, and where—accounts for the majority of variation in the resources and opportunities available to human beings. Within nations, other influences—one's gender, ethnic, or racial category and other such factors—serve to disadvantage some individuals in myriad, often invisible, ways from before they are born until their deaths. These patterns of inequality can be reinforced over generations through the effects of structural barriers, differences in political power, or social discrimination, limiting the potential of persons to flourish.

One recent estimate suggests that the richest 8 per cent of individuals in the world enjoy the same income as the other 92 per cent of the population (Milanovic 2013) and this is likely to be an underestimate as the incomes of the rich are poorly reflected in household surveys and even in tax records.[14] Prominent social movements across the world (from the Indignados in Spain to the Movimento dos Trabalhadores Rurais Sem Terra in Brazil to the Occupy movement across the world and the Arab Spring protests) have all been at least partly driven by the concern of the perceived illegitimacy of economic and political inequality; and these are only the most well-known such instances. Governments in many parts of the world, it seems, are faced with dissatisfied citizenries that object to inadequate life chances. At the same time, the middle class is also burgeoning in many countries and, especially if modestly defined, arguably also in the world as a whole. Poverty appears to have fallen by certain measures although in a very geographically uneven way. These diverse facts give rise to a complex picture of a changing global reality. Better

[14] Some recent attempts have been made to try and include additional data from alternative sources such as tax records and the top incomes database when estimating inequality. In future versions of our database we hope to include information from such exercises.

research and data are needed to begin to capture the gross contrasts as well as the necessary nuances. Such data must be used not only for purposes of description but in order to better understand the determinants of the changing relative and absolute fortunes of people. We have presented some results from our initial (benchmark) global consumption distribution dataset. There are myriad applications that can be imagined, separately or together, for this dataset as well as its twin global income distribution dataset. We present a work in progress that offers possibilities for a deeper understanding of the evolution of material well-being both within and across countries, for regions and the world as a whole, and that extends from description to explanation. It is to this end that we introduce our project as a whole, and seek to build and improve the database that is its foundation, with the involvement of interested specialists and the world public.

References:

Atkinson, A, and A. Brandolini (2001). 'Promise and Pitfalls in the Use of "Secondary" Data-Sets: Income Inequality in OECD Countries As a Case Study'. *Journal of Economic Literature, American Economic Association,* 39(3): 771–99.

Chen, S., and M. Ravallion (2004). 'How Have the World's Poorest Fared since the Early 1980s?'. *World Bank Research Observer*, 19(2): 141–69.

Datt, G. (1998). *Computational Tools for Poverty Measurement and Analysis*. Washington, DC: International Food Policy Research Institute.

Deaton, A. (2005). 'Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)'. *Review of Economics and Statistics*, 87(1): 1–19.

Deininger, K., and L. Squire (1996). 'A New Data Set Measuring Income Inequality'. *The World Bank Economic Review*, 10(3): 565–91.

Dykstra, S., B. Dykstra, and J. Sandefur (2014). 'We Just Ran 23 Million Queries of the World Bank's Website'. Working Paper 362. Washington, DC: Center for Global Development.

Jayadev, A., R. Lahoti, M. Nikiforos, and S.G. Reddy (forthcoming). 'The Global Consumption and Income Project: Methods and Contents'.

Kakwani, N.C. (1980). 'Functional Forms for Estimating the Lorenz Curve: A Reply'. *Econometrica*, 48(4): 1063–64.

Lakner, C., and B. Milanovic (2013). 'Global Income Distribution: from the Fall of the Berlin Wall to the Great Recession'. Working Paper 6719. Washington, DC: World Bank.

Luxembourg Income Study Database (LIS), www.lisdatacenter.org (multiple countries; 1–3 June 2014). Luxembourg: LIS.

Milanovic, B. (2002). 'True World Income Distribution, 1988 and 1993: First Calculation Based on Household Surveys Alone'. *The Economic Journal*, 112(476): 51–92.

Milanovic, B. (2005). *Worlds Apart: Measuring International and Global Inequality*. Princeton, NJ: Princeton University Press.

Milanovic, B. (2013). 'Global Income Inequality in Numbers: in History and Now'. *Global Policy*, 4(2): 198–208.

Minoiu, C., and S.G. Reddy (2009). 'The Estimation of Poverty and Inequality through Parametric Estimation of Lorenz Curves: An Evaluation'. *Journal of Income Distribution*, 18(2): 160–78.

Pinkovskiy, M., and X Sala-i-Martin (2009). 'Parametric Estimations of the World Distribution of Income'. Working Paper 15433. Cambridge, MA: NBER.

Quah, D.T. (1996). 'Twin Peaks: Growth and Convergence in Models of Distribution Dynamics'. *Economic Journal*, 106(437): 1045–55.

Solt, F. (2009). 'Standardizing the World Income Inequality Database'. *Social Science Quarterly*, 90(2): 231–42.

Stiglitz, J.E., A. Sen, and J.-F. Fitoussi (2010). *Mismeasuring Our Lives: Why GDP Doesn't Add Up*. New York: The New Press.

Total Economy Database, The Conference Board. Available at: http://www.conference-board.org/data/economydatabase (Accessed 1 February 2014).

UNU-WIDER World Income Inequality Database, Version 2.0c, May 2008. Available at: http://www.wider.unu.edu/research/Database/en_GB/database (accessed May 2014).

Villasenor, J., and B.C. Arnold (1989). 'Elliptical Lorenz Curves'. *Journal of Econometrics*, 40(2): 327–38.

Povcalnet. Available at: http://iresearch.worldbank.org/PovcalNet/index.htm (accessed 1 April 2014). Washington, DC: World Bank.

World Development Indicators: Available at: http://data.worldbank.org/data-catalog/world-development (accessed 1 February 2014). Washington, DC: World Bank.