Data Literacy in Economic Development

Building up to development data analysis

Simon Halliday[[1]](#footnote-1)

This version: May 2016

Preliminary and Incomplete. Please do not cite without permission.

**Abstract:**

When studying economic development students regularly encounter a variety of economic development measures of absolute and relative deprivation: from poverty measures like the headcount ratio and Foster-Greer-Thorbecke measures, to income dispersion measures like the Gini index. Though exceptions exist, students seldom practice the application of poverty measures to real world data enabling them to go about "doing economics" (Bartlett and King, 1990). In this paper, based on a lower-level elective in economic development, I propose a series of exercises that start with stylized 10 household economies for basic poverty analysis and proceed to an assignment that culminates in students applying relevant measurements of income, poverty and inequality to a nationally representative cross-sectional survey from South Africa using Excel or Google Spreadsheets. Students employ the skills they develop from the exercises to a team research project analyzing real-world data from a developing country to reach a policy conclusion (McGoldrick, 2008; Imazeki, 2014). The data sources are easily tailored to alternative household surveys in developing countries that include the five main variables that are used (household income, household size, government transfers, remittance receipts and a unique household identifier), for example, a survey from Mexico has already been used (provided by a textbook publisher).

Keywords: development, poverty measurement, inequality, National Income Dynamics Survey, South Africa, Sub-Saharan Africa.

JEL Codes: A22, I32

**Introduction**

A generation ago, Bartlett and King (1990:182) argued “we expect students to eventually learn to think like economists without providing them with any real opportunity to learn how economists go about “doing” economics.” Though much has changed in undergraduate economics education, it often remains the case that students seldom encounter data to come to a conclusion about policy, or to make a decision using economic insights. Yet, at the heart of economic development and development policy is the necessity of making choices based on data used to construct measures to understand both the status quo and the effects of policy on the relevant development outcomes. Why, then, is the instruction of economic development so often divorced from data? This paper aims to fill this gap providing a step-by-step set of exercises for a variety of economic development classes that teach topics on income, poverty and inequality and which aspire to use real-world data.

Speaking to the existence of this gap, though development economics and the study of poverty are featured in the economics education literature, few examples exist of deliberate and sustained development of the understanding of economic measurement issues in an undergraduate development economics or economics of poverty classroom. Diduch (2012) is the main exception. She demonstrates that a detailed engagement with the question of how to measure poverty motivates students and provides them with the nuts and bolts of policy analysis in the United States. However, no similar demonstrations exist for a course in development economics or in the use of data from middle and low-income countries.

To be sure, a variety of courses have attempted to address questions of *content* in the economic development curriculum. Singh and Russo (2013) provide insight into how the use of a ‘dream experiment’ motivates students at a small liberal arts college to investigate closely and deeply the prospect of running an experiment (a randomized controlled trial) in a developing country to focus on a particular policy-relevant topic. In discussing economic growth theories in the undergraduate curriculum, two articles provide a relevant perspective. Chen (2005) showcases the great diversity in theories to explain economic growth. Acemoglu (2013) makes the case that teaching growth and development in the undergraduate curriculum more broadly exposes students to facts about the world economy that are inherently interesting and that provide motivation for instructors to teach growth models, such as the Solow-Swann model, in a more nuanced and historically situated manner. This paper, on the other hand, assumes that an economic development course will teach students ideas of income, poverty and inequality, and innovates in *how* those topics are taught with the goal of eventually grappling with real-world data.

I begin by outlining the main measures that are taught in an economic development classroom – income, poverty and inequality – and explain some of their intuitions and history. Second, I explain the initial exercise that I use to introduce students to the primitives of data analysis for development: individuals with incomes who form an economy, the aggregate properties of which need to be measured. The exercise employs stylized 10 person (or 10 household) economies for students to start learning data literacy and development theory. Third, the encounter with data literacy proceeds to an advanced stage by providing students with data in Excel that are publicly available from the 2008 wave of the South African National Income Dynamics Study (<http://www.nids.uct.ac.za>) as well as the World Bank Development Indicators (WDI). Students use the data to calculate a variety of measures and to compare aggregate national measures with measures from household survey data. Students use these data in problem sets and a take-home midterm exam. Lastly, the semester concludes with students working on team projects on a developing country that they choose and for which they use the Excel skills they have learned to measure poverty and inequality or other relevant household (or other group) characteristics that are pertinent to their research questions.

**Measuring Income, Poverty and Inequality: The Theory**

Most economic development courses and textbooks introduce the measurement of income, poverty and inequality by starting with national measures of gross domestic product (GDP) and gross national product or income (GNP/GNI), that is, in either case computing the value of goods and services of final products sold at market prices (either domestically or by national citizens).

## Income

Typically, input-output tables are used to teach simplified calculations of gross domestic product, relying on either the expenditures that different sectors have on the different factors of production, or the values of final goods produced minus imports. These can then be used to calculate total GDP and then divided by the total population of an example economy to provide per capita GDP. Though it is possible to use artificial input-output tables to teach the basics of GDP calculation, it is not feasible to get students to calculate GDP with data from an economy – students lack the skills and the data are not available.

Apart from the inability to teach students how to go about teaching students how to calculate GDP with actual data, national income measures also have a variety of shortcomings in terms of understanding poverty and inequality. First, that the per capita GDP of a country is a certain number, say $10,000 in 2014, does not imply that each person in an economy has that amount of money to spend, or that the *average* (median) consumer in an economy has disposable income even approximately equivalent to that amount. To be sure, household expenditures on final goods plus goods exported ought to provide an accounting relationship that is approximately equal to GDP in a given year, but this will certainly be affected by inventories, purchases of used goods, and other factors that make household expenditures and aggregate measures differ from each other.

Second, and importantly for economic development, per-capita GDP measures say nothing about the *distribution* of income or the *skewness* of that distribution. For example, two economies may have similar per capita GDP, but different poverty headcounts or Gini index.

Third, GDP measures are compiled from data with which most students are unfamiliar and from which they are likely to remain disengaged. Though some students may have worked in retail, restaurant services, or construction, few students have worked in manufacturing, non-restaurant services, or agriculture where final goods and services may be produced and aggregated up for final measurement by national statistical agencies. Many students, in their work and family life, though, will be familiar with the idea that the family as a unit has a certain amount of income that can be spent on the goods and services that the family wants and with which the family may compare itself to other families, or through which the family may be compared by policy-makers with other families.

It therefore becomes useful to think about economies comprising households or individuals with certain levels of daily (or monthly) income that corresponds more closely to the measures of poverty and inequality that are used in economic development, while also being the unit of analysis that students themselves are more familiar with (the individual and household levels).

What are the measures of income, poverty and inequality that are used and what are their roles in understanding economic development?

## Poverty

Foster, Greer and Thorbecke (1984) were engaged in the ongoing debate about how to measure poverty in a way that was decomposable. That is, if we can separate all people who are poor into different groups (by gender, ethnicity, etc), then a decrease (increase) in the poverty of one of the groups should result in a decrease (increase) in overall poverty (see Sen, 1999).

The Foster-Greer-Thorbecke formula provides a generalizable way to teach each of the poverty measures. Poverty is indexed in the following way:

$$FGT\left(α\right)= \frac{1}{Nz^{α}}\sum\_{i=1}^{q}z-y\_{i})^{α}$$

Where *N* is the total population, *z* is the poverty line, *yi* is the income of individual *i*, and *q* is the total number of individuals with incomes below the poverty line.

FGT(α) provides three different measures of poverty depending on the value of α.

* α = 0 results in the poverty head count ratio where $FGT\left(0\right)=\frac{q}{N}$ where *q* is the number of poor individuals *q*, that is, those individuals with income yi below the poverty line, *z*, divided by the total population *N*.
* α = 1 results in the poverty gap index, measuring the normalized distance of each poor individual’s income from the poverty line, i.e. $FGT\left(1\right)= \frac{1}{Nz}\sum\_{1=1}^{q}(z-y\_{i})$.
* α = 2 results in the poverty gap-squared index, measuring each individual’s squared poverty gap, which is then summed and normalized by the number of individuals in the population multiplied by the poverty line squared, i.e., $FGT\left(2\right)= \frac{1}{Nz^{2} }\sum\_{i=1}^{q}(z-y\_{i})^{2}$.

The poverty headcount ratio takes the intuitive absolute number of poor individuals in an economy, that is, those whose incomes (*yi*) are beneath the poverty line (*z*) and counting them. But the total number of poor individuals is not comparable across countries because countries have populations of different sizes. Therefore, the poverty headcount is turned into a ratio by dividing it by the size of the population – city, state, or country – to make it comparable across economies. The poverty headcount ratio must fall between 0 and 1: that is, if there are no people living beneath a poverty line, then FGT(0) = 0 and if all people fall beneath the poverty line then FGT(0) = 1. For most students, the normalization by the population size is intuitive and immediately understood.

The poverty gap index first calculates the sum of all poor individuals’ income gaps: that is, the difference between their income and the poverty line. Each individual’s poverty gap is then summed across all poor individuals. This poverty gap provides a first approximation for the amount of money (in local currency) that would be required to lift all poor individuals out of poverty. As a result, students learn the *policy* aspect of poverty measurement: that is, a first order question of how much money in absolute terms the government would have to spend to eliminate poverty (assuming no change in the number of poor individuals).

Though students find this insight itself worthwhile, an instructor needs to follow up with the next insight: is this absolute number comparable across economies? Students will typically answer with important ideas, such as the value of the exchange rate, the size of the population, and the different kinds of goods people consume locally (especially given just-seen discussions of purchasing power parity under the topic of income). It then becomes clear that the first part of the FGT(1) formula helps policy-makers (economists or students) to make comparisons across economies because it takes the local poverty line, multiplies it by the total population and therefore provides a measure between 0 and 1 which can be compared across economies. Though the poverty gap in one country may seem small in dollar terms, if in local terms (with a small population) many people are far from the poverty line, then the size of the FGT(1) measure will be large, whereas in another country with an initially larger poverty gap, but also a larger population and higher poverty line, the FGT(1) measure may be significantly smaller and closer to zero. In this way, students come to understand the extent of an individual’s *depth* in poverty, the total money required to raise people out of poverty (the absolute poverty gap) and therefore the total depth of poverty, and the ways in which poverty gaps can be compared across countries (using the indexed version of the poverty gap or the FGT(1)).

Similar insights apply to the FGT(2). Many students, when first encountering the poverty gap-squared are able to compute the poverty gap, but the make the error of squaring the total gap rather than squaring each individual’s gap and summing across the squared gaps. Students can readily identify that the poverty gap-squared (before indexing) gives higher values to those who are further from the poverty line. The poverty gap-squared therefore shows the *severity* of poverty in an economy because it gives greater weight to those individuals with income most distant from the poverty line. Students also identify that the poverty gap-squared cannot be compared across economies because of differences highlighted earlier for the poverty gap. By learning that the value can be indexed by normalizing it by the total population multiplied by the poverty line squared, students are able to compare FGT(2) measures across countries, as is often done in the economic development literature.

Though values of α larger than 2 can be conceived, their value in terms of instruction are limited. Unless a student has taken political philosophy, for example, she would not understand that as α tends to infinity the value of the FGT index tends towards a Rawlsian Rule by which society judges poverty only on the basis of the poorest of the poor.

## Inequality

Many ways to measure inequality exist in the economics literature. Notwithstanding the usefulness of measures like the Atkinson index and the Theil Index (or the class of generalized entropy measures), I focus on two measures in my classroom because of their ease of understanding either graphically, as with the Gini index, or mathematically, as with the Palma measure.[[2]](#footnote-2) The Gini index is often taught graphically as the difference between the Lorenz curve, or the cumulative distribution of income in an economy, and the line of perfectly distributed income, divided by the total area beneath the curve of the line of perfect equality. Thus, we have that the Gini (G) index is given by:

$$G=\frac{A}{A+B}$$

Where A is the area between the Lorenz curve and the line of equality and A + B is the area beneath the line of equality.

The Gini index can also be measured with household and individual data using covariance as follows (Pyatt, et al, 1980):

$$G=cov\left(y,F\left(y\right)\right)\frac{2}{\overbar{y}}$$

Where *cov* is the covariance of income level *y*, the cumulative distribution of *y* given by *F(y),* and $\overbar{y}$ is average income.



Figure 1: The horizontal axis is the cumulative population share. The vertical axis is the cumulative income. The Lorenz Curve for a fictional economy is shown by the solid black line. The line of perfect equality for the lowest ten percent to the highest ten percent is shown by the dashed black line. Area A is the area between the line of perfect equality and the Lorenz curve. The sum of areas A and B provides the total area below the line of perfect equality and the Gini index, G = A/(A+B). The data for this curve are given by ‘Economy J’ in Table 1.

The graphical nature of the Gini index lends it an intuitive appeal to students who can understand the shape of the Lorenz curve (and as a result, typically, of having taken calculus and pre-calculus courses). Students grasp how the Gini depends on the extent to which those individuals or households at different parts of the cumulative income distribution deviate from equal income shares. But, the Gini index, by construction, is not particularly sensitive to changes at the tails of the income distribution.[[3]](#footnote-3) The Gini therefore fails to correspond to the intuitions that students have about the definition of inequality: that it should be *higher* for differences between the highest and lowest parts of the income distribution rather than differences within the middle of the distribution.

Though the Gini is intuitively appealing because of its graphical depiction to those who understand the issues of cumulative distribution functions, it is not easily translatable to many people in terms of the underlying mathematics. Part of teaching a course in economic development is also training students in the skills to communicate an idea of a development outcome intuitively and quickly and the Gini index is not amenable to this outcome. But, the Palma measure (named for Gabriel Palma’s work in Palma, 2011) is an intuitive measure of inequality for whichever outcome you wish to measure, i.e. wealth or income.

$$P=\frac{Y\_{10}}{\sum\_{i=1}^{4}Y\_{i}}$$

Where incomes shares (divide by decile) are ranked from lowest (Y1) to highest (Y10) the Palma takes the income share of the top ten percent of the income distribution (Y10) and divides it by the income share of the bottom 40 percent of the income distribution (Y1 + Y2 + Y3 + Y4). Cobham and Sumner (2013a) show that the Palma ranges from 0.757 (Belarus in 1988) to 15.081 (Namibia in 1993) using the World Bank’s PovCal dataset. Cobham and Sumner argue that this measure captures the main insights of what people intuitively grasp when thinking about inequality and that it reflects the fact that the ‘middle 50%’ (income shares 5 through 9 in ten percent increments) has remained relatively stable over time.

**The Exercise**

The course is a lower-level (post introductory microeconomics and macroeconomics) course in economic development that does not require students to have taken calculus or statistics. The course is taught at a small liberal arts college where the development economics elective will have anywhere from 20 to 55 students depending on enrollments in a given semester. In collaboration with the quantitative learning center at the college, students are able to attend workshops on how to use MS Excel and have access to a statistical consultant (though she was rarely used: she recorded two visits for this course).

Three main ideas: income, poverty and inequality using initial in-class exercises with ‘artificial economies’ then applying those skills to real-world data from an internationally recognized panel data survey. We use one wave of the survey as a cross-sectional dataset. The data the students use are a subset of the original data that includes government transfers and remittance data. Students analyze the data in MS Excel or Google Spreadsheets using the relevant formulas.

The artificial economies comprise ten ranked person-households (one person in each household) to provide a basic guideline for thinking about incomes, poverty and inequality. Let us start with thinking about income and consider the case of the economies presented in Table 1. The table presents the economies given to the students in the exercise. Groups of 3 to 5 students are each given an economy, corresponding to the rows in the table. The economies (A through J) comprise 10 people with incomes given by the columns (P1 to P10). Each economy has a total income ranging from 35 monetary units to 100 monetary units (numbers between 0 and 100 are used to ease student understanding, not for their relationship to reality). People (columns) in each economy have incomes ranging from 1 to 30 per day. Ten people are used because each person therefore corresponds to the idea of a decile, which is necessary for later computation of the Palma and Gini measures.

## Income

Students first compute the total daily income in the economy by summing each person’s daily income over the ten individuals. They then compute the per capita daily income by taking the total and dividing it by the total number of people in the economy (10). The per-capita incomes therefore range from 3.5 to 10. The relevance of computing per-capita income and *how* per capita income is different to individual income is immediately apparent, an insight students sometimes fail to grasp when first thinking about GDP data in the abstract. Students do this task *before* their class on poverty, typically at the end of a class on income and GDP to begin thinking about the exercise.

## Poverty

Having seen the income calculations in a previous class, students are then told that in each of the economies there are two poverty lines: $2 per person per day and $3 per person per day (these roughly correspond to the current $1.90 and $3.10 World Bank poverty lines). The purpose then becomes to illustrate four ideas encapsulating important student learning outcome

1. The choice of poverty lines is crucial to understanding *how much* poverty exists in an economy.
2. Two countries that have *identical* income per capita (low or high) need not have identical rates of poverty.
3. Countries with *different* income per capita (low vs. high) need not have *different* rates of poverty. On the contrary, rates of poverty and income per capita may be unrelated.
4. Fiscal policy and redistribution of income can alleviate poverty. (I defer this to after a discussion of inequality).

To meet the first outcome, students first calculate the different poverty measures (head count ratio, poverty gap index, poverty gap-squared index) for each of the two poverty lines. A person who has income *yi = z* is not considered poor. Students then check their calculations with other members of the class who have been allocated the same economy (A to J) and they engage in a think-pair-share exercise about their successes and failures in calculating the different poverty measures (the focus here is on the process of calculation and a growth mindset rather than the ‘correct’ answer; in the groups students have always managed to find the correct answer in assisting each other).

To address the second outcome, students are *paired* with other students who were given another country that has the *same* income level in total daily income and per-capita terms. Though students often understand the abstract idea that different economies with similar incomes can have different rates of poverty, seeing ‘data’ for individuals in an economy and comparing economies with other people makes them realize that significant differences that can exist in income distributions and absolute deprivation across two economies. For example, as shown in Table 2, countries A and B both have the same incomes, but starkly differing rates of poverty at the two different poverty lines: A has an FGT(0) = 0.3 with a poverty line of 2, whereas at the poverty line of z = 2, B’s FGT(0) = 0.

To address the third outcome, students are paired with students given other countries with *different* incomes and asked to discuss the rates of poverty. Students immediately see that though a country’s per capita income may be lower (higher) than another country’s per-capita income, that does not imply that rates of poverty are higher (lower). To be sure, students see that with ‘very low income’ countries (such as A and B in Table 1) there is likely to be higher poverty, but having higher income does not mean poverty does not exist in high-income countries (such as I or J). Students also identify that what produces these differences is the distribution of income or inequality, such that even though some countries have higher average income, incomes may be unequally distributed such that there are high levels of poverty.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Person/Country | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | GDP | PCGDP |
| A | 1 | 1 | 1 | 2 | 2 | 3 | 4 | 5 | 6 | 10 | 35 | 3.5 |
| B | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 5 | 6 | 8 | 35 | 3.5 |
| C | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 10 | 12 | 18 | 50 | 5 |
| D | 1 | 2 | 2 | 3 | 4 | 4 | 4 | 8 | 10 | 12 | 50 | 5 |
| E | 1 | 1 | 1 | 1 | 6 | 6 | 9 | 10 | 15 | 20 | 70 | 7 |
| F | 2 | 3 | 3 | 4 | 4 | 4 | 5 | 6 | 9 | 30 | 70 | 7 |
| G | 1 | 2 | 3 | 3 | 4 | 4 | 6 | 12 | 20 | 30 | 85 | 8.5 |
| H | 1 | 1 | 5 | 5 | 8 | 9 | 10 | 12 | 16 | 18 | 85 | 8.5 |
| I | 2 | 2 | 3 | 5 | 7 | 9 | 13 | 15 | 18 | 26 | 100 | 10 |
| J | 1 | 1 | 2 | 2 | 8 | 8 | 14 | 15 | 19 | 30 | 100 | 10 |

Table 1: Individual incomes in each artificial economy with their corresponding daily incomes and per-capita daily incomes.

|  |  |  |
| --- | --- | --- |
|  | Poverty Line, z = 2 | Poverty Line, z = 3 |
| Country | HC | PG | PG2 | FGT(0) | FGT(1) | FGT(2) | HC | PG | PG2 | FGT(0) | FGT(1) | FGT(2) |
| A | 3 | 3 | 3 | 0.3 | 0.15 | 0.075 | 5 | 8 | 14 | 0.5 | 0.4 | 0.35 |
| B | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 5 | 5 | 0.5 | 0.25 | 0.125 |
| C | 4 | 4 | 4 | 0.4 | 0.2 | 0.1 | 7 | 11 | 19 | 0.7 | 0.55 | 0.475 |
| D | 1 | 1 | 1 | 0.1 | 0.05 | 0.025 | 3 | 4 | 6 | 0.3 | 0.2 | 0.15 |
| E | 4 | 4 | 4 | 0.4 | 0.2 | 0.1 | 4 | 8 | 16 | 0.4 | 0.4 | 0.4 |
| F | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0.1 | 0.05 | 0.025 |
| G | 1 | 1 | 1 | 0.1 | 0.05 | 0.025 | 2 | 3 | 5 | 0.2 | 0.15 | 0.125 |
| H | 2 | 2 | 2 | 0.2 | 0.1 | 0.05 | 2 | 4 | 8 | 0.2 | 0.2 | 0.2 |
| I | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 0.2 | 0.1 | 0.05 |
| J | 2 | 2 | 2 | 0.2 | 0.1 | 0.05 | 4 | 6 | 10 | 0.4 | 0.3 | 0.25 |

Table 2: Poverty measures for each economy at the two poverty lines. HC = head count, PG = poverty gap, PG2 = poverty gap-squared, FGT(.) corresponds to the Foster-Greer-Thorbecke measure for each economy at the relevant poverty line (z = 2, 3) for each of α = 0, 1, 2.

## Inequality

As discussed earlier, there are two main measures of income inequality that I refer to in the course: the well-established Gini index and the lesser known, but upcoming Palma measure proposed by the Center for Global Development.

Students can very intuitively grasp how the Palma measure, taking the income of the top decile and dividing it by the sum of incomes by the bottom four deciles, provides a useful measure of inequality when judging inequality as the dispersion of income between the poorest and the richest parts of an economy (in terms of either income or wealth). The idea that a higher Palma corresponds to more inequality is intuitive to them.

As with the poverty exercise, the main emphasis in the inequality exercise is for students to achieve the following

1. Find a line of equality and sketch that line, therefore computing area B for the Gini index
2. Find each decile’s income share and sketch that as a bar chart or line graph, therefore using it to find the value of area A as one would to calculate the Gini index (the area under the line of equality minus the cumulative share at that income share).
3. Use the income shares to compute the Palma

For all students, step 1 is normally intuitive and easy. If the total GDP in an economy is, for example, $100, then all students understand pretty quickly that the lowest 10% should have $10, the lowest two should accumulate $20, etc. Because the numbers provided to them in the toy economies are relatively small they can either compute the line of equality using fractions or a calculator.

Step two has been more challenging and requires students accurately to consider what a cumulative share means *empirically*. Consider economy D as an example. The total income is 50. The bottom person’s income is 1, meaning their cumulative share is 1/50 = 0.02. If the poorest 10% were to have had an equal share of 5/50 = 0.1, then they would require an additional 4/50 = 0.08 of income. Therefore, the portion of area A (for the Gini index) for the lowest 10% is 4/50. This share of A needs to be computed for each of the ten shares, it is then summed over the ten shares to find a total area of A. This total area A must be divided by the total area of B (the area under the line of perfect equality). If each ten percentage point share is given its share, then the total area of A + B is 5.5 (the area under the line of perfect equality). Each economy’s share of A divided by 5.5 results in the country’s Gini Index, as shown in Table 4. Students in their groups, in calculating the Gini Index values, will likely sketch Lorenz curves that look approximately like those in Figure 1, which shows each artificial economy’s Lorenz curve compared with a line of perfect equality.

Having completed these exercises, students are put in groups and re-paired to identify the following learning outcomes:

1. Countries with the same/similar incomes can have different levels of inequality.
2. Countries with different total and per-capita incomes can have similar levels of inequality.
3. Fiscal policy to redistribute income can address income inequality.

|  |  |
| --- | --- |
|  | Cumulative Share of Income |
| Country | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
| A | 0.03 | 0.06 | 0.09 | 0.14 | 0.20 | 0.29 | 0.40 | 0.54 | 0.71 | 1.00 |
| B | 0.06 | 0.11 | 0.17 | 0.23 | 0.29 | 0.37 | 0.46 | 0.60 | 0.77 | 1.00 |
| C | 0.02 | 0.04 | 0.06 | 0.08 | 0.12 | 0.16 | 0.20 | 0.40 | 0.64 | 1.00 |
| D | 0.02 | 0.06 | 0.10 | 0.16 | 0.24 | 0.32 | 0.40 | 0.56 | 0.76 | 1.00 |
| E | 0.01 | 0.03 | 0.04 | 0.06 | 0.14 | 0.23 | 0.36 | 0.50 | 0.71 | 1.00 |
| F | 0.03 | 0.07 | 0.11 | 0.17 | 0.23 | 0.29 | 0.36 | 0.44 | 0.57 | 1.00 |
| G | 0.01 | 0.04 | 0.07 | 0.11 | 0.15 | 0.20 | 0.27 | 0.41 | 0.65 | 1.00 |
| H | 0.01 | 0.02 | 0.08 | 0.14 | 0.24 | 0.34 | 0.46 | 0.60 | 0.79 | 1.00 |
| I | 0.02 | 0.04 | 0.07 | 0.12 | 0.19 | 0.28 | 0.41 | 0.56 | 0.74 | 1.00 |
| J | 0.01 | 0.02 | 0.04 | 0.06 | 0.14 | 0.22 | 0.36 | 0.51 | 0.70 | 1.00 |
| Equality | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.00 |

Table 3: Cumulative income shares for each economy used to compute the Gini index.

|  |  |  |  |
| --- | --- | --- | --- |
|  | A at Each 10% Share | Total |  |
| Country | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 | Area A | Gini |
| A | 0.07 | 0.14 | 0.21 | 0.26 | 0.30 | 0.31 | 0.30 | 0.26 | 0.19 | 0.00 | 2.04 | 0.37 |
| B | 0.04 | 0.09 | 0.13 | 0.17 | 0.21 | 0.23 | 0.24 | 0.20 | 0.13 | 0.00 | 1.44 | 0.26 |
| C | 0.08 | 0.16 | 0.24 | 0.32 | 0.38 | 0.44 | 0.50 | 0.40 | 0.26 | 0.00 | 2.78 | 0.51 |
| D | 0.08 | 0.14 | 0.20 | 0.24 | 0.26 | 0.28 | 0.30 | 0.24 | 0.14 | 0.00 | 1.88 | 0.34 |
| E | 0.09 | 0.17 | 0.26 | 0.34 | 0.36 | 0.37 | 0.34 | 0.30 | 0.19 | 0.00 | 2.41 | 0.44 |
| F | 0.07 | 0.13 | 0.19 | 0.23 | 0.27 | 0.31 | 0.34 | 0.36 | 0.33 | 0.00 | 2.23 | 0.41 |
| G | 0.09 | 0.16 | 0.23 | 0.29 | 0.35 | 0.40 | 0.43 | 0.39 | 0.25 | 0.00 | 2.59 | 0.47 |
| H | 0.09 | 0.18 | 0.22 | 0.26 | 0.26 | 0.26 | 0.24 | 0.20 | 0.11 | 0.00 | 1.82 | 0.33 |
| I | 0.08 | 0.16 | 0.23 | 0.28 | 0.31 | 0.32 | 0.29 | 0.24 | 0.16 | 0.00 | 2.07 | 0.38 |
| J | 0.09 | 0.18 | 0.26 | 0.34 | 0.36 | 0.38 | 0.34 | 0.29 | 0.20 | 0.00 | 2.44 | 0.44 |

Table 4: Value of area A at each decile, providing the total A area for each economy and therefore each economy's Gini index.



Figure 2: Lorenz curves for artificial countries with the line of equality shown diagonally. This is a draft figure. Cumulative population share is along the horizontal axis. Cumulative income share is along the vertical axis.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Country | A | B | C | D | E | F | G | H | I | J |
| Gini | 0.37 | 0.26 | 0.51 | 0.34 | 0.44 | 0.41 | 0.47 | 0.33 | 0.38 | 0.44 |
| Palma | 2.00 | 1.00 | 4.50 | 1.50 | 5.00 | 2.50 | 3.33 | 1.50 | 2.17 | 5.00 |

Table 5: Comparing the Gini Index and Palma Measure for Artificial Countries

# Proceeding to real-world data

We use two kinds of data: aggregated national data from the World Bank Development Indicators (WDI) and a relevant household dataset, the South African National Income Dynamics Study (NIDS). NIDS is a panel data survey that has been run consistently since 2008 (SALDRU, 2015). I discuss the ways students use and think about the NIDS data for each of the three topics: income, poverty and inequality.

In the NIDS dataset, students are given a spreadsheet with eight variables as shown in the MS Excel spreadsheet in the supplementary materials: a household identifier (hhid), household size (hhsize), household income net of remittances and government transfer (hhy\_net), government transfers received by the household (transfers), remittances received by the household (remittances), per capita net income which is measured as net income divided by household size, the household’s income rank from lowest to highest (rank(pcy)), and the distribution of the ranks F(pcy).[[4]](#footnote-4) The instructor only needs a dataset with five of these to compute the others (I provide pcy, hhrank and F(pcy) as aides to the students for initial attempts to work with the data).

Prior to giving students the full data set for problem sets and the midterm exam, students are first shown a ten row dataset. The ten row data set comprises ten household randomly selected from the full NIDS or ENHRUM dataset and the incomes are linearly transformed (so that they appear different to the equivalents in the original survey). Each household has the variable as shown in the final dataset and I explain to students the correspondence between the ten-person (or ten-household) dataset that we used in the exercise and this sample dataset from the data we shall use. We discuss how each of the measures we have already found for income, poverty, and inequality can be used to understand inequality with this mini dataset. Students can then re-apply these insights to the full data set and think through, particularly, the roles of government transfers and household transfers (remittances).

## Income

First, for income, one learning outcome that students can gain substantially from is to understand the difference between a country’s per-capita GDP during a given year and the average incomes from a household survey for that same country. Students can be asked to identify how a country should be classified according to the WDI data and how that some country would be classified if the income classification were done based on the average incomes from a household survey, such as NIDS from South Africa or ENHRUM from Mexico.

Using the spreadsheet from a subset of data from the 2008 wave of NIDS, students compute the average per-capita income using the net household income, summing it over all the rows and dividing it by the total population in the survey to find the average per-capita income. This provides a *monthly* number as the net household income is measured on a monthly basis. But, World Bank data are based on annual measures, so students have to multiply the number they find by 12 to find an annual number (on this basis you can also start a valuable conversation about what social scientists would assume if we take one month as representative of all months over a year).

Having computed the annual number, students must convert it using the exchange rate equivalents in US dollars, and then convert these once again to Purchasing Power Parity dollars to compare with the WDI data. Students gain two skills from doing this: they understand, first, the difference between exchange rates without purchasing power parity, which is the number with which they are most familiar if *they* were to convert their own money to a foreign currency, and, second, they realize that local prices and consumption patterns must be controlled for in order to compare countries internationally. The PPP comparison is used to compare the data to the World Bank categorizations of income per capita based on PPP GDP with aggregated World Bank Development Indicator (WDI) data and make a decision on whether the country would still be classified as the World Bank Classified it based on GDP data. As average incomes from household surveys are often lower than the per-capita income computed from GDP data, students can see that the country may be classified differently to how it would be classified using GDP data.

## Poverty

Having computed and compared incomes, students use the per capita income data to compute the Foster-Greer-Thorbecke index measures to understand basic poverty measures. There are two basic capabilities that students need to do this:

1. Familiarity with using mathematical formulas in Excel.
2. Familiarity with IF statements in Excel.

Students need to be familiar with if statements to ensure that when they calculate the poverty measures, they only count individuals in households with per-capita income below the poverty line. They then use mathematical formulas independently of the if statement (or within the if statement) to compute the household’s poverty gap and the household’s poverty gap-squared.[[5]](#footnote-5) For all of the following calculations, please refer to the annotated spreadsheet in the supplementary materials to see how they are computed.

Let us start with FGT(0) or the head count ratio.

1. In one column, use an if statement to count whether the household is classified as poor or not based on the per-capita income. This if statement takes a 1 if the household is poor and 0 otherwise.
2. In the next column, multiply the household size by the outcome of whether the household is poor or not. This is the *total number of poor individuals* in the household.
3. In a new cell (preferably in a new sheet), sum all of the *rows* showing the total number of poor individuals in each household. This is the *poverty head count*.
4. Divide the *poverty head count* by the *total population* in the sample to find the poverty head count ratio or FGT(0).

For the poverty gap and the FGT(1) we use steps 1 and 2 from previously, but add 4 more steps:

1. Subtract the household’s per capita income from the poverty line (in South Africa in 2008, this was ZAR515 per person per month). This is the *individual poverty gap*.
2. Multiply the individual poverty gap by the number of poor household members (from step 2). This is the *total household poverty gap*.
3. Sum all of the household poverty gaps to find the *total poverty gap*.
4. Divide the total poverty gap by the number of individuals in the sample multiplied by the poverty line to find the *poverty gap index* or FGT(1).

For the poverty gap-squared, we replicate steps 1 though 2 from previously, but replace steps 3 through 6 with the following:

1. Subtract the household’s per capita income from the poverty line (in South Africa in 2008, this was ZAR515 per person per month) and square this number (raise it to the power 2). This is the *individual poverty gap-squared*.
2. Multiply the individual poverty gap-squared by the number of poor household members (from step 2). This is the *total household poverty gap-squared*.
3. Sum all of the household poverty gaps-squared to find the *total poverty gap-squared*.
4. Divide the total poverty gap-squared by the number of individuals in the sample multiplied by the poverty line-squared to find the *poverty gap-squared index* or FGT(2).

Having completed these calculations for the initial per-capita income data, students are asked to include the monthly household transfers from government and remittances into the household total income to re-calculate per capita income. They repeat the processes used to calculate the poverty head count ratio, poverty gap index, and poverty gap-squared index to evaluate the effects of transfers and remittances on poverty. In so doing, they can see the ways in which anti-poverty policy can alleviate the proportion of the population that is poor, the depth of poverty of poor individuals, and the severity of poverty for the most poor.

## Inequality

Having examined poverty and the effects of transfers and remittances on poverty, students use the same data to compute a Gini index on *household* data.[[6]](#footnote-6) Unlike with the artificial household exercise, students do not compute each household’s share of the difference between the Lorenz Curve and the line of perfect equality. Instead, they use the covariance formulation of the Gini index. To do this, they require two pieces of information:

1. The rank of each household in terms of the overall income distribution to form a distribution.
2. Each household’s per-capita income

The covariance of these terms and the average per-capita income are required for the Gini index. Students can then combine these using the equation to find the Gini index for the data, G = 0.6727 (or 67.27). This is roughly consistent with more rigorous estimates of the Gini index from the NIDS data and with estimates of South Africa’s Gini index provided by the World Bank.[[7]](#footnote-7)

Having completed this aspect of the exercise, students proceed to include remittances and government transfers in household total income and divide it by household size to find the new per capita income (as they did with the poverty measures). They use the new per-capita income to re-rank the households to find a new F(pcy) and go through the same process they went through originally to find the Gini index value. Including remittances and government transfers, the students find that the Gini decreases to 0.4787 (or 47.87). This is an approximate decrease of 20 percentage points or a 40% decrease in inequality as a result of the transfers and remittances, showing the extent to which transfers and remittances alleviate inequality. Students can then compare the post-transfer and post-remittance Gini index to comparable Gini index values internationally.

# Conclusion and Extensions

Using only five variables (household income, household size, government transfers, remittance receipts and a unique household identifier), students are able to contrast the effects of government policy and private remittances as anti-poverty and inequality-alleviating measures, as well as identifying the differences between national income measures (such as GDP and per capita GDP) with more granular measures of income, poverty and inequality.

Though I have not proposed how students should use these steps to begin their empirical team projects, the in-class exercises and the Excel-based problems provide the chassis on which such a project can be constructed. Students glean the main insight that, with few economic variables, they can come to significant conclusions about what the status quo is prior to a policy, such as government transfers to the poor, and what the effects of such a policy might be on household outcomes such as poverty, consumption, or health. In the course in which these skills were taught, students went on to use household data sets to understand urban vs. rural health insurance rates in China, levels of HIV/AIDS, Malaria and TB awareness and knowledge in Nigeria, the distribution of investments in fertilizer in Malawi, a comparison of within and between race inequality in South Africa, and a variety of other topics. In each case, they were enabled to investigate these questions through the knowledge that with a few variables they might be able to come to powerful conclusions.

Filling a gap in the literature with respect to the teaching of data literacy in economic development, I hope to contribute to an ongoing discussion about transparency in research, knowledge and application of basic statistics, and the uses of empirics in undergraduate economics field courses. Future research should focus on the related ways in which activities such as these can impact learning directly, rather than on the construction and motivation of these activities such as I have provided here.

**References**

 Acemoglu, Daron, 2013, Economic Growth and Development in the Undergraduate Curriculum, *Journal of Economic Education*, 44(2), 169-177.

 Bartlett, Robin L. and Paul G. King, 1990, Teaching Economics as Laboratory Science, Journal of Economic Education, Spring 1990, 181-193

 Chen, Edward K., 2005, Teaching and Learning Development Economics: Retrospect and Prospect, *Journal of Economic Education*, Summer, 236-248.

Alex Cobham and Andy Sumner, 2013a, ‘[Putting the Gini Back in the Bottle? “The Palma](http://www.kcl.ac.uk/aboutkings/worldwide/initiatives/global/intdev/people/Sumner/Cobham-Sumner-15March2013.pdf)” as a Policy-Relevant Measure of Inequality’, *Mimeograph*, London: King’s College London.

 Alex Cobham and Andy Sumner, 2013b, ‘[Is It All About the Tails? The Palma Measure of Income Inequality](http://www.cgdev.org/sites/default/files/it-all-about-tails-palma-measure-income-inequality.pdf)‘, *CGD Working Paper,* Washington DC: Center for Global Development.

Cowell, F. A. (2000). Measurement of inequality. In A. B. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, Chapter 2, pp. 87-166. Amsterdam: North Holland.

Finn, A., Leibbrandt, M., & Levinsohn, J. (2014). [Income mobility in a high-inequality society: Evidence from the first two waves of the National Income Dynamics Study](http://www.tandfonline.com/doi/pdf/10.1080/0376835X.2). Development Southern Africa, 31(1), 16-30.

Finn, A., & Leibbrandt, M. 2013a. [Mobility and Inequality in the First Three Waves of NIDS](http://www.opensaldru.uct.ac.za/bitstream/handle/11090/684/2013_120_Saldruwp.pdf?sequence=1). SALDRU Working Paper no 120.

Finn, A., & Leibbrandt, M. 2013b. [The dynamics of poverty in the first three waves of NIDS](http://www.opensaldru.uct.ac.za/bitstream/handle/11090/683/2013_119_Saldruwp.pdf?sequence=1). SALDRU Working Paper No. 119.

Leibbrandt, M., & Levinsohn, J. 2011. [Fifteen Years on: Household Incomes in South Africa](http://www.nber.org/papers/w16661.pdf). NBER Working Paper, no. 16661.

Leibbrandt, M., Finn, A., Argent, J. & Woolard, I. 2010. [Changes in income poverty over the post-apartheid period: an analysis on data from the 1993 Project for Statistics on Living Standard and Development and the 2008 base wave of the National Income Dynamics Study](http://reference.sabinet.co.za.ezproxy.uct.ac.za/webx/access/electronic_journals/bersee/bersee_v34_n3_a3.pdf). Studies in Economics and Econometrics. 34(3):25-43.

Leibbrandt, M., Woolard, I., Finn, A. & Argent, J. 2010. Trends in South African income distribution and poverty since the fall of Apartheid. OECD Social, Employment and Migrations Working Paper, no. 101.

Litchfield, J., 1999, ‘Inequality: Methods and tools’, World Bank.

Jansen, A., Moses, M., Mujuta, S., & Yu, D. (2014). [Measurements and determinants of multifaceted poverty in South Africa](http://www.tandfonline.com/doi/pdf/10.1080/0376835X.2014.984377). *Development Southern Africa,* (ahead-of-print), 1-19.

Leibbrandt, M., Finn, A. & Woolard, I. 2012. [Describing and decomposing post-apartheid income inequality in South Africa](http://www.tandfonline.com.ezproxy.uct.ac.za/doi/pdf/10.1080/0376835X.2012.645639). Development Southern Africa. 29(1):19-34

McCormick Diduch, Amy, 2012, Using the Consumer Expenditure Survey to Teach Poverty Measurement, *Journal of Economic Education*, *43*(1), 99–106.

Palma, Gabriel, 2011, ‘Homogeneous middles vs. heterogeneous tails, and the end of the ‘Inverted-U’: The share of the rich is what it’s all about’ Development and Change, 42, 1, 87-153

Pyatt, G., Chen, C.-N., Fei, J., 1980, The Distribution of Income by Factor Components. *Quarterly Journal of Economics*, November, 451-73.

Southern Africa Labour and Development Research Unit. National Income Dynamics Study 2008, Wave 1 [dataset]. Version 5.3. Cape Town: Southern Africa Labour and Development Research Unit [producer], 2015. Cape Town: DataFirst [distributor].

Singh, Prakarsh and Alexa Russo, 2013, A Dream Experiment in Development Economics, *Journal of Economic Education*, 44(2), 158-168.

 Taylor, J. Edward and Travis J. Lybbert, 2015, *Essentials of Development Economics,* 2nd Edition, U. California Press.

1. Smith College, Department of Economics, 107 Pierce Hall, 21 West St. Northampton, MA, 01062, USA. shalliday@smith.edu. +1413-585-3529. [↑](#footnote-ref-1)
2. Five axioms for inequality measurement are considered important in the literature: 1) The Pigou-Dalton transfer principle, 2) income scale independence, 3) Dalton’s principle of population, 4) anonymity or symmetry, and 5) decomposability. See Cowell (2000) and Litchfield (1999). [↑](#footnote-ref-2)
3. See Palma (2011) and Cobham and Sumner (2013a, b) for a fuller explanation as to why the Gini index is insensitive to changes in the tails of the income distribution: the basic intuition is that substantial weight is given to deciles 2 through 5 of the income distribution and the tails are not given substantial weight in the Gini as each part of the distribution is equally weighted. This characteristic of the Gini is likely to be axiomatically true (Shorrocks, 1980). [↑](#footnote-ref-3)
4. Taylor and Lybbert (2015) include the Mexico Rural Household Survey in their supporting material for their textbook, *Essentials of Development Economics*, along with some exercises similar to those I ask students to perform with these data. However, my exercises are grounded in initial exercises in-class as presented earlier, along with additional questions and ideas that Taylor and Lybbert do not include. [↑](#footnote-ref-4)
5. A variety of work has used the NIDS to understand South African poverty. See Leibbrandt, Woolard, Finn and Argent (2010), Leibbrandt and Levinsohn (2011), Finn and Leibbrandt (2013b), and Jansen, Moses, Mujuta and Yu (2014). [↑](#footnote-ref-5)
6. Because each row corresponds to a household and we cannot easily manipulate this to create an individual-level dataset, we can only compute a household income-based Gini index. Though this is unfortunate from the perspective of consistency, in terms of pedagogy and learning outcomes the main concern is for students to think through what is required to find a Gini index and what the unit of observation implies about what gets computed [↑](#footnote-ref-6)
7. As with poverty, several authors have used the NIDS to examine inequality in South Africa, such as Leibbrandt, Finn and Woolard (2012), Finn and Leibbrandt (2013a), and Finn, Leibbrandt, and Levinsohn (2014). Admittedly, the measures should be weighted by their survey weights, but such a discussion is beyond the ambit of the course I teach. [↑](#footnote-ref-7)