Economics Letters 121 (2013) 257–262

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Using an asset index to simulate household income

Kenneth Harttgen[a], Sebastian Vollmer[b,c,*]

[a] ETH Zurich, Department of Humanities, Social and Political Sciences, NADEL. Clausiusstrasse 37, 8092 Zurich, Switzerland
[b] University of Göttingen, Department of Economics, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany
[c] Harvard School of Public Health, Department of Global Health and Population, United States

HIGHLIGHTS

- Survey data for developing countries often do not include income or expenditure data.
- We propose a method to simulate household income using DHS and macroeconomic data.
- We illustrate our approach for Bolivia, Indonesia and Zambia.
- We calculate Gini and Atkinson inequality measures for the simulated household income.
- We perform an inequality decomposition by education, sex, household size and rural/urban.

ARTICLE INFO

Article history:
Received 7 December 2012
Received in revised form
8 August 2013
Accepted 9 August 2013
Available online 19 August 2013

JEL classification:
D31
I31
I32
O1

Keywords:
Asset index
Household income
Developing countries

ABSTRACT

Commonly available survey data for developing countries often do not include income or expenditure data. This data limitation puts severe constraints on standard poverty and inequality analyses. We provide a simple approach to simulate household income based on publicly available Demographic and Health Surveys (DHS) and macroeconomic data. We illustrate our approach with DHS data for Bolivia, Indonesia and Zambia.

1. Introduction

Household income or expenditure data are often used to measure current and long-term welfare of households and within-country inequality (see for example Deaton, 1997). The availability of household survey data has increased the understanding of within-country inequality and its determinants. Large-scale national representative household survey data has become more and more available in recent years. However, commonly available survey data for developing countries – such as the Demographic Health Surveys (DHS) – often do not include income or expenditure data.

Filmer and Pritchett (2001) and Sahn and Stifel (2001) have proposed a one-dimensional index based on household assets and other household characteristics as a proxy of long-term material welfare to overcome the problem of missing income and expenditure data. The so-called 'asset index' is often used in empirical literature on poverty and inequality analysis as a proxy variable for household income. There is a large body of literature that uses an asset index to explain inequalities in educational outcomes (e.g. Ainsworth and Filmer, 2006; Bicego et al., 2003), health outcomes (e.g. Bollen et al., 2002; Schellenberg et al., 2003), child malnutrition (e.g. Sahn and Stifel, 2003; Tarozzi and Mahajan, 2005), or child mortality (e.g. Sastry, 2004) when data on income or expenditure is not available. In addition, asset indices are used to analyze changes and determinants of poverty (Harttgen and Misselhorn, 2006 (wie in Bibliografie); Stifel and Christiaensen, 2007; World Bank, 2006). It has also been examined whether an asset index can serve as a proxy for income or expenditure (e.g. McKenzie, 2005;

* Corresponding author at: University of Goettingen, Department of Economics, Platz der Goettinger Sieben 3, 37073 Goettingen, Germany. Tel.: +49 551 39 8170.
E-mail addresses: harttgen@nadel.ethz.ch (K. Harttgen), svollmer@uni-goettingen.de (S. Vollmer).

0165-1765/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.econlet.2013.08.014

Howe et al., 2009). Filmer and Scott (2008) validate the use of various asset index methods by comparing how asset index outcomes match with results using per capita expenditures.

In this paper, we provide a simple approach to simulate household income[1] from an asset index using commonly available DHS and macroeconomic data. We illustrate our approach with an inequality decomposition analysis for Bolivia, Indonesia and Zambia. There are many other possible applications for the simulated household income data due to its link to the comprehensive source of socio-economic and health indicators from the DHS. The DHS are relatively frequent and comparable across countries and over time. For example, one could use the imputed income data to study the determinants of income poverty, effects of public spending or poverty traps.

2. Methodology

First, we use principal component analyses to construct an asset index for each household. Second, we estimate the national income distribution using the Gini coefficient and average per capita income. To match the asset index with the income distribution, we make the assumption that the ranking of households within the asset distribution is the same as the ranking within the income distribution. We assign the level of income at the p-quantile of the national income distribution to the household at the p-quantile of the asset distribution.

We follow the approach of Filmer and Pritchett (2001) and Sahn and Stifel (2003) to construct an asset index. The main idea of this approach is to construct an aggregated one-dimensional index over the range of different dichotomous variables of household assets capturing housing durables and information on the housing quality that indicate the material status (welfare) of the household:

$$A_i = b_1 a_{i1} + b_2 a_{i2} + \cdots + b_k a_{ik} \quad (1)$$

$$a_{ik} = \beta_k c_i + u_{ik} \quad (2)$$

for $i = 1, \ldots, N$ households and $k = 1, \ldots, K$ household assets. A_i is the asset index, the a_{ik} refers to the respective asset of the household i recorded as dichotomous variables in the DHS data sets, and the b_k are the weights for each asset that are used to aggregate the indicators to a one-dimensional index. In the model, the ownership of an asset k of household i, identified by a_{ik}, is a linear function of an unobserved factor, which in our case is material welfare c_i. The relationship between the asset k in c_i is given by β_k plus a noise component u_{ik}, where both terms have to be estimated (Sahn and Stifel, 2001).[2]

For the estimation of the weights and for the aggregation of the index, we use a principal component analysis as proposed by Filmer and Pritchett (2001). The first principal component is the asset index.[3] The principal component analysis is structured by a set of equations where the asset variable is related to a set of latent factors:

$$\begin{aligned} \tilde{a}_{1i} &= v_{11}A_{1i} + v_{12}A_{2i} + \cdots + v_{1k}A_{ki} \\ &\cdots \\ \tilde{a}_{ki} &= v_{k1}A_{1i} + v_{k2}A_{2i} + \cdots + v_{kk}A_{ki}, \end{aligned} \quad (3)$$

where the $\tilde{a}$ are the k asset indicators (the a's in Eq. (1)) normalized by their mean and their standard deviations; A are the k principal components and v are the weights that relate the principal components to the ownership of the asset (Filmer and Scott, 2008). After the weights v have been estimated, the inversion of the equation system (3) yields the following set of equations:

$$\begin{aligned} A_{1i} &= b_{11}\tilde{a}_{1i} + b_{21}\tilde{a}_{2i} + \cdots + b_{k1}\tilde{a}_{ki} \\ &\cdots \\ A_{ki} &= b_{1k}\tilde{a}_{1i} + b_{2k}\tilde{a}_{2i} + \cdots + b_{kk}\tilde{a}_{ki}. \end{aligned} \quad (4)$$

The equation for the first principal component is the equation with the highest variance. The weights that are used to aggregate the asset variables into a one-dimensional index are given by the set $(b_{11}, b_{21}, \ldots, b_{k1})$. The asset index is calculated for each individual, weighted by the household size.

We assume that national income distributions follow a log-normal distribution. Formally, the log-normal distribution LN (μ, σ) is defined as the distribution of the random variable $Y = \exp(X)$, where X has a normal distribution with mean μ and standard deviation σ. The Gini coefficient G of LN (μ, σ) is given by $G = 2\Phi(\sigma/\sqrt{2}) - 1$, where Φ is the distribution function of the standard normal distribution. Therefore, the parameters μ and σ of LN (μ, σ) can be determined from the average income $E(Y)$ and the Gini coefficient G as follows:

$$\sigma = \sqrt{2\phi^{-1}}\left(\frac{G+1}{2}\right), \qquad \mu = \log\left(E(Y)\right) - \sigma^2/2.$$

The crucial assumption for mapping the asset index to household income is that the ranks within both distributions are equal, i.e. the household at the p-quantile of the asset distribution will also be at the p-quantile of the income distribution. The corresponding household income is then given by the p-quantile $x(p)$ of the log-normal distribution of income:

$$x(p) = e^{\mu+u}\,(p) \cdot \sigma$$

where $u(p)$ is the p-quantile of the standard normal distribution and μ and σ are the parameters of the log-normal distribution of income.

The described approach also allows the analysis of changes in poverty and/or inequality over time if multiple survey rounds are available.

3. Empirical illustration

To illustrate our approach we use Demographic and Health Survey (DHS) data for three countries, thereby capturing different regions of the developing world: Bolivia, Indonesia, and Zambia (see Table 1). For each country, we illustrate our approach for two periods, thus allowing us to analyze inequality across countries as well as over time. For this, the asset index is calculated separately for

Table 1
Data sources and macroeconomic statistics.

Data source	Demographic and Health Survey (DHS)	PovcalNet		
Country	Survey year	Year	Means $	Gini (%)
Bolivia	2003	2002	185.99	60.24
Bolivia	2008	2008	214.33	56.29
Indonesia	2003	2002	61.83	30.39
Indonesia	2007	2005	72.22	34.11
Zambia	2002	2002	41.07	42.08
Zambia	2007	2006	42.4	57.49

Note: Mean income refers to monthly income per capita PPP (Purchasing power parity). PPPs come from International Comparison Program (ICP).

[1] Throughout this paper 'household income' refers to household income per capita. This means that although we provide data on an individual basis, each household member within a household has the same household income. Any poverty and inequality analysis is then weighted by the household size.

[2] The model is based on the following assumptions: (i): households are distributed *i.i.d.*; (ii): $E(u_i|c_i) = 0$; (iii): $V(u_i) = \text{Diag}\left\{\sigma_1^2, \ldots, \sigma_K^2\right\}$.

[3] An alternative way to estimate the weights for the assets to derive the aggregated index would be to employ a factor analysis, for example, by Sahn and Stifel (2001). However, the two estimation methods show very similar results.