The Census Microdata Wealth Index: An Application to Predict Education Outcomes in Developing Countries

Rodrigo Lovaton Davila [a], Dorothy Gondwe [b], Aine Seitz McCarthy [a], Phatta Kirdruang [a] and Uttam Sharma [a]

Last updated: May 2010

Abstract

This research aims to develop a valid and consistent measure for socioeconomic status at the household level using census microdata available from the Integrated Public Use Microdata Series- International (IPUMS-I), the world's largest census database. First, we use principal component analysis to compute a wealth index based on housing characteristics and asset ownership. The validation strategies include comparing our proposed index with the widely used Demographic and Health Survey (DHS) wealth indices and then verifying the predictive power of our index on education enrollment and attendance. Moreover, we attempt to identify general conditions necessary to produce an internally consistent asset index based on census microdata. Our results show a consistently positive effect of the wealth index on education outcomes across four census samples (Peru 1993, South Africa 1996, Brazil 2000, and Colombia 2005). Furthermore, graphical analysis of kernel distributions suggests our measure is comparable to that of the DHS. Finally, through a stepwise elimination procedure, we find evidence supporting the internal consistency of the census asset index. As an important practical implication of results, we are able to propose a methodology to determine which assets are more important in determining household socioeconomic status.

a University of Minnesota, Department of Applied Economics and Minnesota Population Center
b University of Minnesota, School of Public Health and Minnesota Population Center

We would like to thank Matt Sobek, Lara Cleveland, Deborah Levison and Ragui Assaad for their input and guidance on the IPUMS-I project. We would also like to thank Joe Ritter for his valuable feedback and suggestions.

1. Introduction

A measure of socioeconomic status of a household is an important element in most economic and demographic analyses. This measure is useful, not only in terms of estimating poverty and inequality within a society, but it can also be used as a control variable in finding the effects of other variables associated with wealth (Filmer and Pritchett, 2001). Based on theoretical grounds, household income and expenditure levels are often used as measures of household socioeconomic status. However, collecting data on income and expenditures can be costly. As a result, most demographic and household surveys that contain income and expenditures data tend to have small sample sizes.

In contrast, large-scale data collection on population and housing such as census surveys can overcome problems of small sample sizes and underrepresentation of certain population groups in smaller geographical units. Although the main feature of the census microdata is the enumeration of individuals and households in the country at a particular point in time, it has advantages over other household surveys for at least three reasons. First, census microdata are more commonly available than nationally representative household surveys[1]. Second, due to the larger scale, census data are more comprehensive when compared to household surveys in representing all population groups accurately[2]. Third, the larger number of observations in census microdata can provide more precise estimates for statistical purposes. Given all of these reasons, census data are a promising source for conducting social and economic research.

To date, the Integrated Public Use Microdata Series (IPUMS) - International, at Minnesota Population Center, University of Minnesota, has collected the one of the world's largest archive of free and publicly available (though restricted) census samples. Currently, the database includes 158 census samples taken during 1960 to present from 55 countries around the world. Furthermore, IPUMS-International data are composed of microdata at individual and household levels. The data includes information on household characteristics as well as a wide range of population characteristics, such as basic demographic, fertility, education, occupation, migration, and others, which are coded and documented in systematically across countries and years.

Nevertheless, despite the availability of census data and its comprehensiveness, most of these census samples, particularly from developing countries, do not collect information on income or expenditures, which are used widely as a measure of socioeconomic status. The lack of this measure

[1] For example, IPUMS International has available three censuses for Israel (1972, 1983, and 1995) and one for Palestine (2007), but neither country has microdata from DHS or the Living Standards Measurement Surveys (LSMS).

[2] For example, ethnicities that are a small proportion of the population might not be well represented in a survey.

limits the ability of researchers to perform analyses using census data. Thus, it is essential to develop a measure of household socioeconomic status based on the other information usually available in censuses. This proposed measure will not only improve the use of census data in social and economic research, but will also give some insights about the relative socioeconomic status of households in a particular country during a specific year.

The asset-based approach to determine socioeconomic status has been widely used in previous studies as an appropriate measure of household wealth (Montgomery et al, 2000; Filmer and Pritchett, 2001; Sahn and Stiefel, 2000 and 2003; McKenzie, 2005). Even though census data are widely available and collects information on assets, there are no large-scale efforts to date to develop an asset-based measure of relative household wealth. Given the advantages of census data and the lack of socioeconomic measures in most censuses, the goal of this paper is to develop a valid and consistent measure for socioeconomic status at the household level using census microdata available from IPUMS-International. More specifically, we attempt to use non-monetary indicators including asset ownership, utilities, dwelling characteristics, appliances, and other amenities that are generally available in censuses to compute an asset index. To validate the asset index calculation from census microdata, we attempt to evaluate the validity of the proposed index through an application on education and also to suggest some conditions (or criteria) for consistency of the index when using different census samples. These justifications will be illustrated by using selected samples from IPUMS-International.

The paper is organized as follows: section two provides a review of the literature of asset-based wealth indices and their outcomes, section three covers the methodologies we used, section four provides an overview of the data used, section five is a discussion of our results and section six provides the conclusion and extension of future research. There are five appendices with figures and tables to support our results.

2. Literature review

To measure socioeconomic status of households, we need long-term and stable indicators. There are advantages of using the asset index as a proxy for consumption expenditure. The asset index overcomes the limitations of utilizing consumption expenditure and income in measuring wealth levels of households. In addition to errors in collecting income data (particularly in a developing country), income and consumption expenditure data are vulnerable to seasonal fluctuations due to economic cycles, recall bias, poor quality of price deflators, and are notoriously lacking in nationally representative surveys in developing countries. (Assaad, Levison, and Zibani, 2010; McKenzie, 2005; Sahn and Stiefel, 2000;

Montgomery et al, 2000). In contrast, the asset-based wealth index is more readily available and is a more stable and long-term measurement of household socioeconomic status.

Further, previous research has provided empirical evidence of the asset-based wealth index appropriateness in measuring household's socioeconomic status. A common approach has been to compare the asset index household classification with that from consumption expenditure or income. This type of comparison assumes consumption expenditure or income is a good proxy of household wealth. The results so far have indicated the asset index is comparable to the household expenditure in predicting household wealth (Filmer and Pritchett, 2001; Sahn and Stiefel, 2003; Booysen et al, 2008; Filmer and Scott, 2008). Filmer and Pritchett (2001) demonstrated the empirical validity and reliability of the asset based wealth index using data sets from India, Indonesia, Nepal and Pakistan. They examined classification differences when using household expenditures versus an asset index, and analyzing the effect of each of them on selected educational outcomes. Filmer and Pritchett claim that the asset index works as well, or better, than traditional expenditure based measurements in predicting education enrollment status. Sahn and Stiefel (2003) find mixed results when conducting direct comparisons of the asset index and predicted outcomes with the distribution of the reported per capita consumption expenditure but show that the asset index is a valid predictor of child nutrition outcomes and is comparable or better to predicted or reported expenditures. Booysen et al (2008) compared an asset index constructed using the DHS survey to per capita expenditures in Ghana and found relatively high positive correlations (between 0.42 and 0.49).

In addition, other studies aim to assess the effectiveness of the asset index to identify inequalities or predict outcomes associated with household socioeconomic status. In particular, the distribution of specific outcomes across different strata of wealth levels is used to assess validity of the asset index. That is, we expect that people classified at the lowest wealth levels will have the worst outcomes as compared to those classified at the highest wealth levels. Several studies have explored empirically validity of the asset-based approach for education outcomes (Filmer and Pritchett, 1999 and 2001; Minujin and Bang, 2002; McKenzie, 2005; Filmer and Scott, 2008), fertility (Bollen et al, 2002, Filmer and Scott, 2008), nutrition (Sahn and Stiefel, 2003), as well as morbidity and mortality (Houweling et al, 2003; Filmer and Scott, 2008). Even though the evidence of the relative performance of the asset-based measures with respect to other socioeconomic status indicators is mixed, the overall conclusion seems to point to the validity of the asset index approach.

In practice, the main limitation of the asset index that has surfaced from previous research is the problem of clumping and truncation. This is due to broad categories of asset ownership or household characteristics, disregarding the quality of assets or public delivery of household services all of which may result in assigning households that may not have the same wealth level the same asset score. (McKenzie, 2005; Wall & Johnston, 2008) The use of census samples which are in general much larger than most nationally representative surveys, may inherently provide more variance in asset distribution but care should be taken where broad categories of asset ownership or household characteristics are used to define component indicators of the asset index.

Even though census data are widely available and collects similar information on asset ownership, household characteristics, and access to utilities, there are no large scale efforts to develop measures of relative household socioeconomic status. We apply the asset index approach to IPUMS census samples, a large collection of datasets which are free and publicly available, and explore conditions for the validity and internal consistency of the constructed asset index.

3. Methodology

In applying the asset index approach, we focus on two separate but interrelated questions. First, we aim to verify internal consistency of the index, taking into account that the number and type of data available vary widely across censuses. Second, we test the validity of the index in measuring household socioeconomic status for census microdata through an application on education outcomes.

Calculation of the asset index is performed through Principal Component Analysis (PCA), a data reduction technique, which creates orthogonal linear combinations from a set of variables, and orders them according to their contribution to the overall variability of the variables analyzed. In order to apply PCA to census microdata, all variables are transformed into a dichotomous version, including the categorical variables for housing characteristics (e.g. material of walls or floor) or access to utilities (e.g. type of water or sewage service).

The first research question is focused on general conditions necessary to produce an internally consistent index based on census microdata. The underlying issue is the variable availability across censuses, which could have any number of assets listed or discrepancies on how data were collected. Even though the general recommendation has been to use the most variables available, as long as those are related to unobserved wealth (Rutstein and Johnson, 2004; McKenzie, 2005), it remains unclear which types of assets have larger contributions to the constructed measure and what the minimum number of necessary variables is. Further, two data problems could arise and restrict the power of the asset index: (i) clumping, if a limited number of values are produced; and (ii) truncation, if there are no indicators available to explain differences at the tails of the wealth distribution (McKenzie, 2005; Minujin & Bang, 2002).

In order to set a standard for input requirements for the index, we perform stepwise elimination of the variables with the smallest PCA scoring factor (in absolute value) and recalculate the index with the remaining variables. In each step, we verify the level of agreement of rankings through Spearman rank correlations and the internal consistency of the indices using the Cronbach's alpha. The Cronbach's alpha measure of reliability will generally increase as the inter-correlations among variables increases (Cortina, 1993). The coefficient is calculated using the following formula:

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K} \sigma_{Y_i}^2}{\sigma_X^2}\right)$$

Where K is the number of variables, σ_X^2 is the variance of the observed total test, and $\sigma_{Y_i}^2$ is the variance of the component of component I for the current sample.

Furthermore, we examined the properties of the resulting asset index distributions and verified the agreement level of results using census microdata with comparable DHS datasets. In particular, we identified possible clumping and truncation problems graphically in the resulting kernel distributions. For purposes of comparing results with DHS datasets, we selected censuses coinciding in time with DHS data collection for specific countries. Since both DHS and IPUMS-I data are nationally representative, we would expect similar distributions of the asset index. In order to verify the agreement between the two, we calculated statistics representing the distribution of each standardized index (percentiles, skewness, and kurtosis) and compared indices graphically using kernel density estimation methods.

The second research question refers to the validity of an asset based index to measure household socioeconomic status applied to census microdata. The question of validity is examined through an application on education outcomes, which are expected to be highly dependent on a household relative standing in the socioeconomic status distribution. We first compare distributions of education enrollment and attainment by quintiles using the census and DHS wealth index. Then, we estimated a probit regression for school enrollment using the wealth index, controlling for other individual and household-level variables. In particular, the model takes the following general form: $\Pr(y = 1 | X) = \Phi(X'\beta)$, where Pr(Y=1|X) is the probability of being enrolled in school given the wealth index and a variety of other independent variables. In addition, as part of the stepwise procedure, we estimated school enrollment regressions using the asset index produced at each step, in order to verify changes in the accuracy of results (measured by the standard error on the wealth index) and explanatory power (measured by the R-squared).

4. Data

In this study, we used the following IPUMS census samples: 1993 Peru (10%, unweighted) with 564,765 households; 1996 South Africa[3] (10%, weighted) with 993,801; 2000 Brazil (6%, weighted) with 2, 652,352 households; 2002 Senegal (10%, unweighted) with 107,999 households; and 2005 Colombia (10%, weighted) with 1,054,901 households. The data include information on a broad range of population and household variables, including household's asset ownership, access to utilities, and dwelling characteristics. A detailed description of the variables available for the asset index is included in Annex 1.

For purposes of comparing results, part of the analysis was performed also on similar microdata from the Demographic and Health Survey (DHS). The DHS typically collects information on a broad range of population characteristics, health conditions, health indicators such as fertility, maternal and child mortality, family planning methods, access to health services, and achievement of specific health policy objectives. Surveys are nationally representative and frequently sample households with specific population groups. However, one important difference between census microdata and DHS is that most DHS samples are based on an eligible population of women of reproductive age, 15 to 49 years, and can sometimes include men of reproductive age, 15 to 59 years. In addition, while DHS includes a set of assets that is generally similar across countries, we observe more differences in asset availability in census microdata. We used DHS data from five countries: Senegal, South Africa, Brazil, Colombia, and Peru. The Demographic and Health Survey of Senegal 2005 (EDS-IV) is a nationally representative survey of 7,412 households with women aged 15 to 49 years and men 15 to 59 years. The survey was based on the urban-rural stratified national sample of 8,000 households from the 2002 census. The 1998 South African Demographic and Health Survey (SADHS) is a representative probability sample of the population living in 12,860 private households and containing women aged 15 to 49. The sampling frame for the SADHS was the 1996 census. The National Demographic and Health Survey of Brazil (PNDS 1996) is a nationally and regionally representative survey of 13,283 households with women aged 15 to 49 years. The sampling frame for the PNDS 1996 was the 1991 population census. The National Survey of Demographic and Health of Colombia (ENDS 2005) is nationally representative survey of 37,211 households containing women of childbearing age (13 to 49 years old). The sampling frame was the 1993 National Population Census of Colombia. The 1992 Demographic and Family Health Survey of Peru (ENDES 1991-1992) is a nationally representative survey of 13,479 private households containing women between 15 and 49 years. The sampling frame was the 1984 National Survey on Nutritional and Health of Peru.

[3] In one of the IPUMS samples, 1996 South Africa, 19 districts in Eastern Cape are not organized into households thus individuals were treated as separate households, if they reported household characteristics.

5. Results

5.1. Internal consistency of the asset index

The number and type of assets included in census microdata vary considerably across countries. We performed a stepwise elimination of variables to determine what assets contribute the most to the final wealth distribution. In each step, the variable with the lowest loading coefficient in absolute value was eliminated since it was contributing the least to the calculation of the index. Then, Cronbach's alpha was calculated to analyze internal consistency of the remaining variables and Spearman rank correlations to examine changes in the ordering of households given by the asset index distribution. We expected increasing internal consistency and high rank correlations as we eliminate meaningless variables, but possibly decreasing consistency and relatively smaller rank correlations as we eliminate variables that are more important in defining the wealth index.

The stepwise procedure was performed for three samples: Peru 1993, Colombia 2005, and South Africa 1996. The former two samples have relatively more asset variables available (about 60 indicators in each case), while the latter has limited variables available (21 indicators) and asset types (only fuel for cooking, water source, toilet type, and household members per room). Detailed graphs showing results are included in Appendix 5. We observe that internal consistency is slightly increasing during the early variable eliminations. This is consistent with the hypothesis that by eliminating variables that have a low contribution to the definition of wealth we are able to achieve higher internal consistency. For example, the third variable to be dropped for the Peru data was 'tricycle', which intuitively should not be important in determining wealth and actually only a small proportion of households own one. Furthermore, for all samples we observe that after eliminating about two thirds of the available variables, both internal consistency and the rank correlations begin decreasing considerably. Even though the subset of assets in the last third of available variables is different for each sample, we observe that frequently only one or two categories are left for housing materials and utilities, and that most of the variables left are durable goods. For example, in the case of Colombia, the last twenty-three variables include ten durable goods and thirteen housing and utility variables. Most of these remaining housing and utility variables include only two categories of the original (flooring and walls materials, water source, toilet type, and fuel used for cooking), while only one has three categories left (main method for trash disposal).

The second check for internal consistency of the wealth index was to graphically compare the distribution of the asset index based on census and DHS data, with all variables available in each database. To begin, we recreate the DHS wealth index using the original DHS household data (calling this replica DHS wealth index) to verify that we are using the same methodology in creating wealth indices. While we cannot verify that the DHS wealth index is in fact measuring wealth, we would expect to see the census index showing a comparable distribution, implying that both indices are at least measuring the same unobserved phenomenon. It is important to note, however, that despite being on the same scale, 'distance' is not well defined in these distributions and it is possible that both indices are valid without their graphical distributions coinciding. The kernel densities for the asset index were estimated both for census and DHS data for six countries: Senegal, Brazil, South Africa, Colombia, Peru, and Egypt. In Appendix 2, we show summary statistics for the standardized asset indices and in Appendix 4 we include kernel density estimations for the distribution of the asset index for DHS and Census.

Results show that for all countries we obtain comparable distributions of the wealth index, except for the case of South Africa. In particular, the shape of the asset index distributions almost coincides for all countries, with small areas of discrepancy, which could be explained by the fact that the set of variables available in each dataset is not exactly the same. Furthermore, we do not observe considerable problems of clumping or truncation in any of the countries analyzed, with the exception again of South Africa. The cutoff points for percentiles, skewness, and kurtosis almost coincide for all samples, except for South Africa. For example, in the case of Colombia, the 25th, 50th, and 75th percentiles are almost the same, while we have comparable skewness (-0.05 as compared to -0.44) and kurtosis (1.94 as compared to 2.75). In the specific case of South Africa, DHS shows a smoother distribution and the census asset index has some clumping problems. Clearly, this is due to the fact that the number of variables available in census microdata for South Africa is smaller and comprises a more limited set of assets.

5.2. Validity of the asset index

The question of validity is concerned with the asset index actually measuring wealth and not any other phenomenon that could be associated with ownership of durable goods, housing characteristics, or access to utilities. The validity of the asset indices is checked by examining education outcomes, which should be highly dependent on household wealth. First, we calculated differences in school enrollment and educational attainment by quintiles of the asset index. We would expect considerable differences between the top and bottom quintiles if the asset index is correctly measuring wealth. The analysis was performed both for census and DHS data, in order to compare the relative performance of the asset indices defined in each case. Figures 1A and 1B show the proportion of children 6-14 years old that enrolled in school by asset index quintile for Brazil, South Africa, Peru, Senegal, and Colombia, both for census and DHS data.

The figures on school enrollment by quintile using census microdata show considerable differences between the top and bottom quintile, which range between 14 percentage points for South Africa to 42 percentage points for Senegal (Figure 1B). Moreover, we are able to identify a strictly increasing enrollment pattern as we move from the bottom to the top quintile for all samples analyzed.

When we compare census results with DHS data (Figure 1A) we observe that the differences between the top and bottom quintiles are almost the same. Furthermore, the increasing pattern seems to coincide in each case, with South Africa showing slight increases and Senegal sharp increases moving from the bottom to the top of the wealth distribution. As we would expect, this same pattern is reflected in the comparison of primary and secondary school completion by quintile between census and DHS data. These education inequality measurements can be seen in Appendix 3 (Figure 2A-3B).

Figure 1A: School enrollment by DHS wealth index quintiles

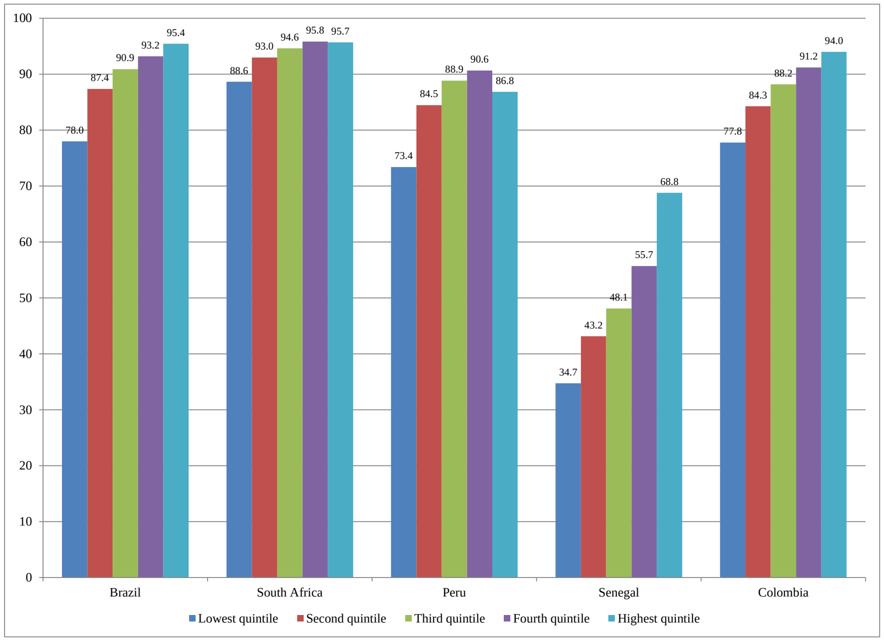
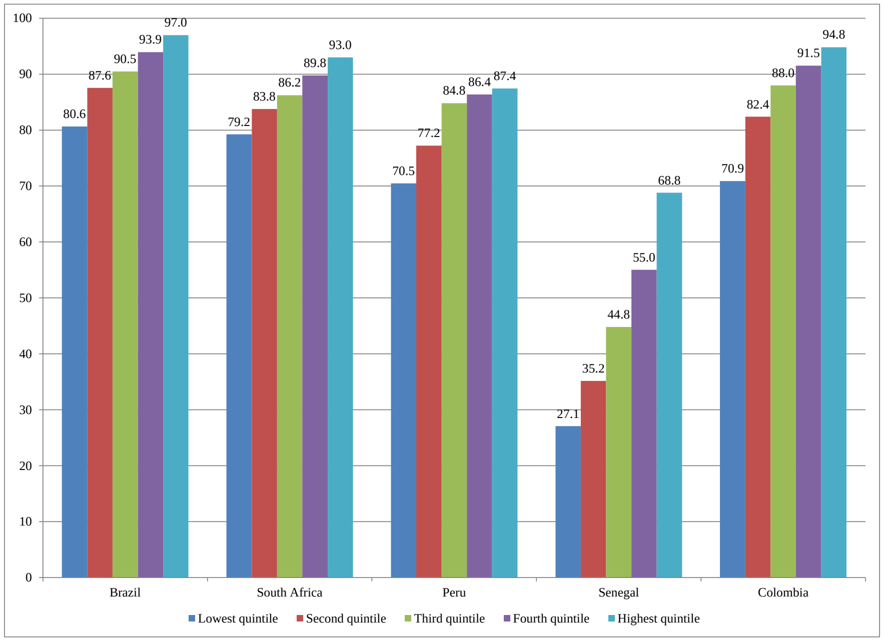


Figure 1B: School enrollment by census wealth index quintiles



The validity of the asset index was also explored through probit regressions for school enrollment conditional on the wealth index and other individual and household-level variables[4]. Regressions were estimated for children ages 6 to 14 for the following census samples: Brazil 2000, Colombia 2005, Peru 1993, and South Africa 1996. Results are shown in Table 1. The coefficient column shows the values and standard errors of the wealth index coefficient in each sample's regression. This coefficient is positive and significant in all cases, as expected. This indicates that the measurement of wealth, as represented by the census microdata wealth index, has a positive impact on child school enrollment. For example, we can see by the marginal effects column that a one percent increase in the wealth index is represented by a 0.0325 increase in school enrollment for Brazil. While the value of the coefficients and marginal effects are not comparable across samples, given that wealth is measured differently in each country, the fact that each is positive and significant in predicting education enrollment is further evidence of a valid measure of household wealth.

Table 1: School enrollment regression results

Sample	Coefficient	Z-statistic	P-value	Marginal effect
Brazil 2000	**0.3169 (0.0025)**	**125.61**	**<.001**	**0.0325 (0.0003)**
Colombia 2005	**0.1225 (0.0010)**	**113.41**	**<.001**	**0.0190 (0.0002)**
Peru 1993	**0.1674 (0.0038)**	**36.73**	**<.001**	**0.0344 (0.0008)**
South Africa 1996	**0.2405 (0.0033)**	**70.9**	**<.001**	**0.0461 (0.0006)**

Further, we ran this same regression after performing the stepwise elimination of the least impactful variables and recorded the marginal effects and R-squared from each regression (Figures 8A, 9A, and 10A). These figures show a relatively constant R-squared value for the initial period of variable elimination and then a small period of instability before it begins to drop significantly. Likewise, the marginal effects value of the coefficient on the wealth index shows stability over the elimination of approximately the first third of variables eliminated, but becomes much less stable when approximating the marginal effects of the wealth index with far fewer variables. Furthermore, we lose precision in the estimates as we eliminate more variables, effect that is translated into increases in the standard errors for

[4] Besides the wealth index, we control for sex, age, and age squared of the child, sex, age, and age squared of the household head, educational attainment dummies for the household head, and an indicator of urban residence.

the wealth index coefficient. In particular, the 95% confidence interval for the marginal effects shown in Figures 8A, 9A, and 10A is wider as we drop variables, even though this is difficult to observe since we have a large number of observations which allow for small-sized standard errors.

One thing to note here is that since initial number of household variables is quite different across samples, this has a major impact on the results of the marginal effects of the wealth index. As previously mentioned, the wealth index for South Africa was created using only 21 household variables. This fact is reflected in the way the marginal effects graph shows much less detail in the instability and the way the R-squared does not decrease as dramatically when the number of included variables drops. In general, we argue that the South Africa results may be less reliable due to this small number of household variables included in the creation of the wealth index.

6. Conclusions

This paper seeks to display that the census microdata wealth index is both internally consistent and valid in its representation of household socio-economic status for all samples examined here. The evidence provided by the stepwise procedure and graphical analysis of kernel density distributions showed that we are consistently measuring the unobserved socioeconomic status at the household level and that we achieve relatively similar performance to the DHS measure. Furthermore, we indeed observe differences in school enrollment and educational attainment across the asset index quintiles, showing consistently that households at the top of the distribution have better outcomes than those at the bottom. The probit regression gives consistently positive and significant marginal effects of an increased household wealth index on a child's education. Moreover, as we perform individual variable removal and re-run the regression, we see this marginal effect being consistently positive (even though we lose precision of the estimates), while predictive power is generally constant until the wealth index is comprised of too few household variables.

An important practical implication arises from our results. The methodology of the stepwise elimination provides a starting point to determine which (and possibly how many) household variables have a more important contribution to household socioeconomic status and, thus, are necessary to obtain a valid asset index. If, for example, during the stepwise procedure, the remaining ten core variables across samples all include a categorical variable on the type water supply, while none of these include an asset such as boat, tricycle, or sewing machine, then we may be able to more confidently conclude that water supply has a more essential role in classifying households by socioeconomic status in developing countries and that boat, tricycle or sewing machine play a less essential role. As we observe, the Spearman rank correlations show that we obtain almost the same ordering of households by socioeconomic status for all samples for nearly the first third of variables eliminated, which suggests that a subset of assets achieves similar results to an all-variable asset index. In the case of Colombia, for example, we obtain similar results by using all the 60 variables available or a subset based on only 40. When this process is applied to more samples, we may be able to say more accurately which of these remaining core variables are necessary for a valid index. This stepwise procedure makes headway in giving a consistent methodology to determine which household variables are more important to be included in the census microdata wealth index.

Further steps of research will provide additional evidence to develop an asset index methodology to be widely applied to census microdata. First, the analysis will be applied to a greater number of the widely-available IPUMS-I census samples. Second, alternative weighting procedures will be explored. In particular, some comparison of the wealth index to predicted expenditures or income data (if available) may provide further confirmation of its validity. The inclusion of the census microdata wealth index in the IPUMS-I dataset will enhance social science research by giving a reference point to represent socio-economic status. This paper provides evidence of a valid census microdata wealth index and a potential new methodology in evaluating which household variables to include in this index.

Appendix 1: Variable Availability, Census Microdata

	Colombia 2005	Peru 1993	South Africa 1996	Brazil 2000	Senegal 2002
Durable assets					
Telephone	X	X	X	X	X
Television	X	X		8	X
Refrigerator	X	X		X	X
Blender	X				
Stereo	X	X			
Radio		X		X	X
VCR/DVD player					X
Video Camera		X			
Washing machine	X	X		X	
Vacuum		X			
Fan	X				
Computer	X	X			
Oven (gas or electric)	X				X
Microwave	X			X	
Shower	X				
Hot water heater	X				
Air conditioning	X			X	X
Bicycle	X	X			X
Tricycle		X			
Motorcycle or scooter	X	X			X
Car or truck	X	X		9	X
Boat	X				X
Cart for transportation					X
Draft animals					X
Plough					X
Knitting Machine		X			
Sewing Machine		X			

	Colombia 2005	Peru 1993	South Africa 1996	Brazil 2000	Senegal 2002
Utilities					
Water source	8	9	7	6	8
Waste water				6	
Sewage		5			
Type of toilet	7	3	4		6
Electricity		X		X	X
Dwelling characteristics					
Floor material	5	7			4
Wall material	7	8			
Roof material		7			
Kitchen		3			
Fuel used for cooking	7		7		5
Waste disposal method	6			7	
Members per sleeping room	X	X	X	X	
Number of bathrooms				X	
Other					
Dwelling ownership	4			6	
Total	62	59	20	57	38

Note: An ‘X’ indicates that the sample had this household variable; the numbers indicate the number of categories for categorical variables.

Appendix 2: Summary details of Wealth Indices, census and DHS

	Brazil Census 2000	Brazil DHS 1996	South Africa Census 1996	South Africa DHS 1998	Peru Census 1993	Peru DHS 1992	Colombia Census 2005	Colombia DHS 2005
Percentiles								
1%	-2.29	-2.94	-1.42	-1.59	-1.45	-1.64	-1.70	-2.35
5%	-2.06	-2.01	-1.26	-1.42	-1.44	-1.46	-1.58	-1.95
10%	-1.65	-1.53	-1.12	-1.29	-1.28	-1.32	-1.44	-1.51
25%	-0.61	-0.64	-0.88	-0.86	-0.90	-0.93	-0.89	-0.61
50%	0.22	0.09	-0.25	-0.13	0.07	-0.06	0.11	0.16
75%	0.71	0.73	0.88	0.89	0.95	0.98	0.79	0.70
90%	1.18	1.15	1.44	1.50	1.38	1.33	1.28	1.15
95%	1.40	1.29	1.74	1.62	1.56	1.54	1.56	1.45
99%	1.64	1.38	2.14	1.83	1.81	1.72	1.96	2.02
Obervations	1740553	56755	107999	12247	2181359	73042	982934	37211
Mean	0.00	-0.06	0.00	0.00	0.04	0.00	0.00	0.00
Std. Dev.	1.00	1.02	1.00	1.00	1.01	1.01	1.00	1.00
Variance	1.00	1.04	1.00	1.00	1.01	1.01	1.00	1.00
Skewness	-0.63	-0.78	0.45	0.22	0.05	0.06	-0.05	-0.45
Kurtosis	2.65	3.19	1.97	1.82	1.66	1.67	1.94	2.76

Appendix 3: Education Attainment

Figure 2A: Percent Primary School Completion by DHS wealth index quintiles

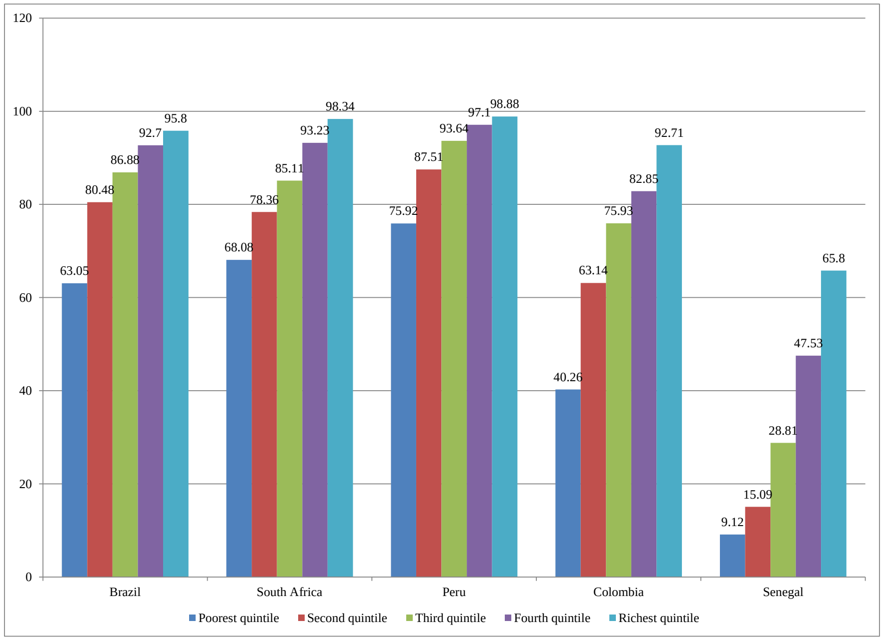


Figure 2B: Percent Primary School Completion by census wealth index quintiles

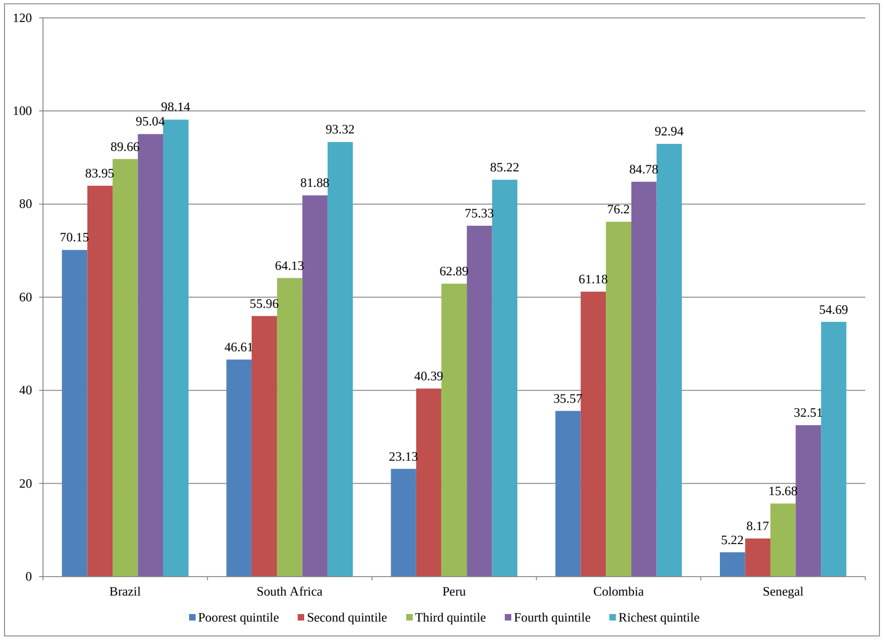


Figure 3A: Percent Secondary School Completion by DHS wealth index quintiles

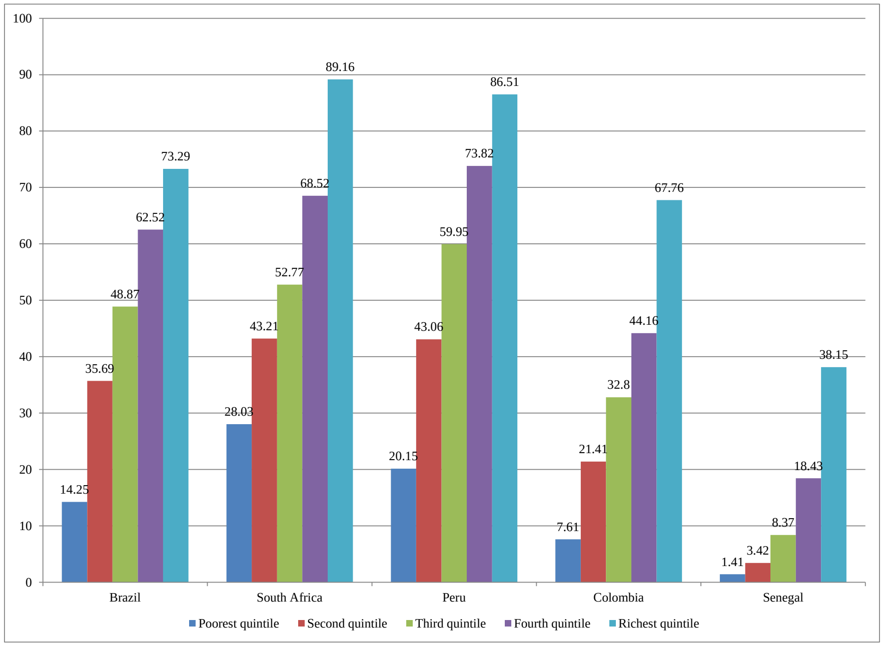
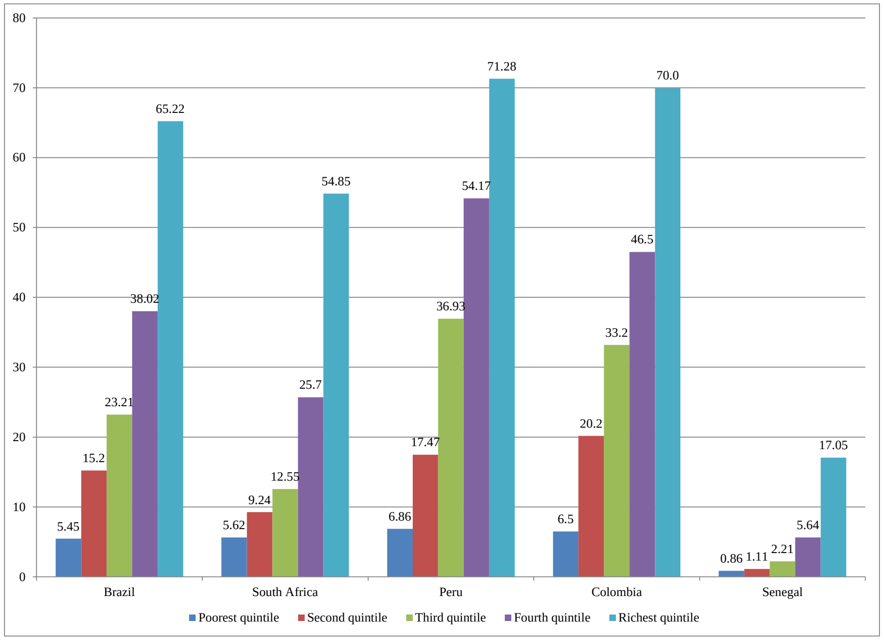


Figure 3B: Percent Secondary School Completion by census wealth index quintiles



Appendix 4: Kernel Density Distributions for DHS and Census Asset Indices

Figure 2: Senegal Census 2002 and DHS 2005

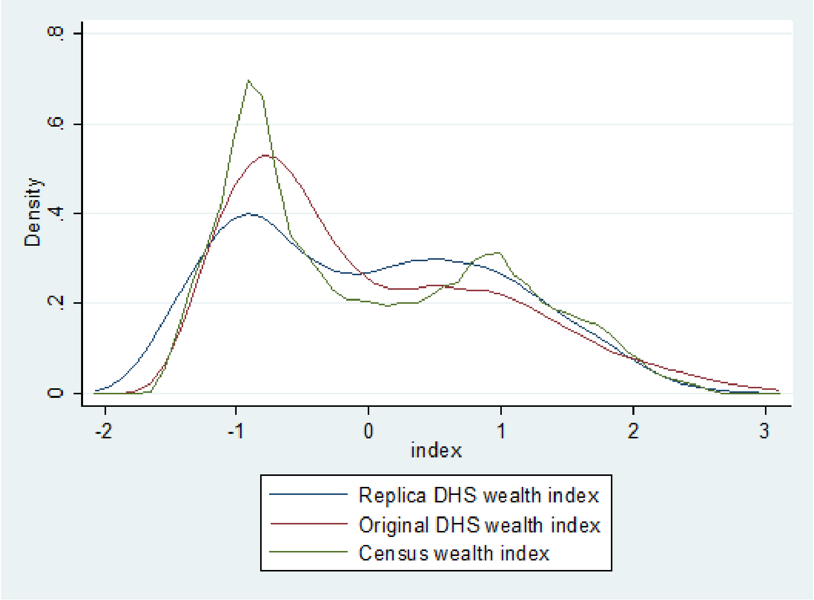


Figure 3: Brazil Census 2000 and DHS 1996

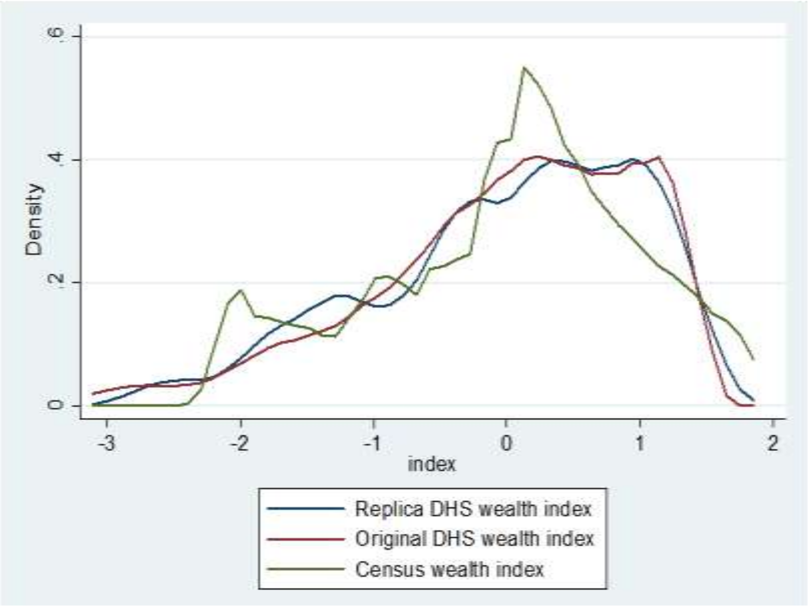


Figure 4: South Africa Census 1996 and DHS 1998

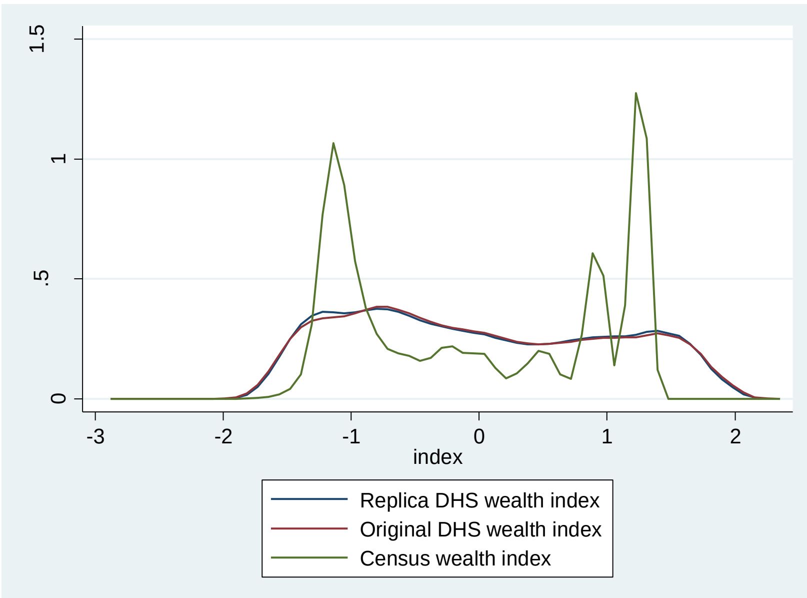


Figure 5: Colombia Census and DHS 2005

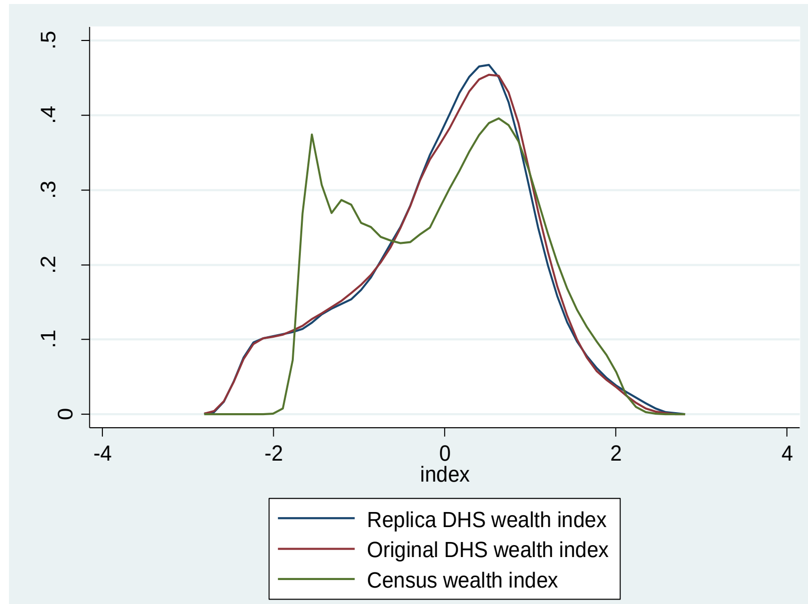
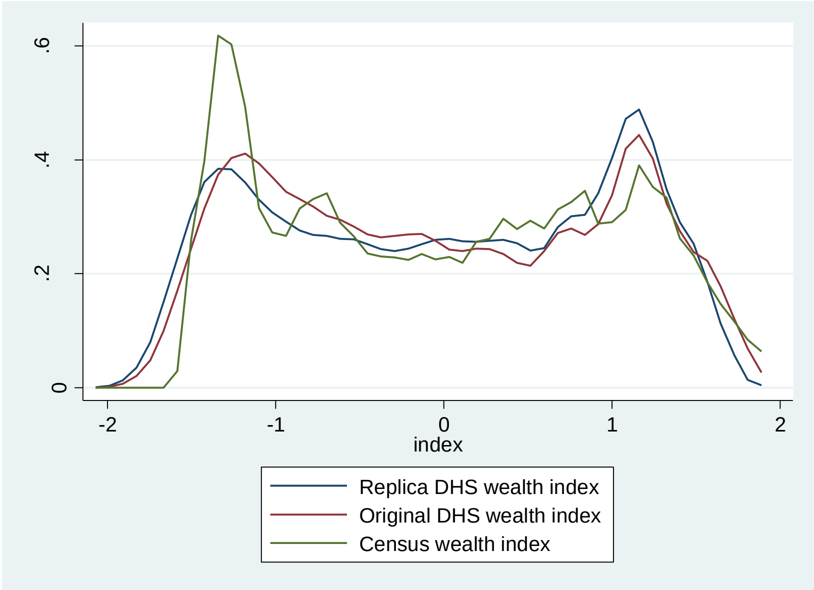


Figure 6: Peru Census 1993 and DHS 1992



Appendix 5: Stepwise Procedure Results

Figure 8A: Peru Census 1993, School attendance regression results

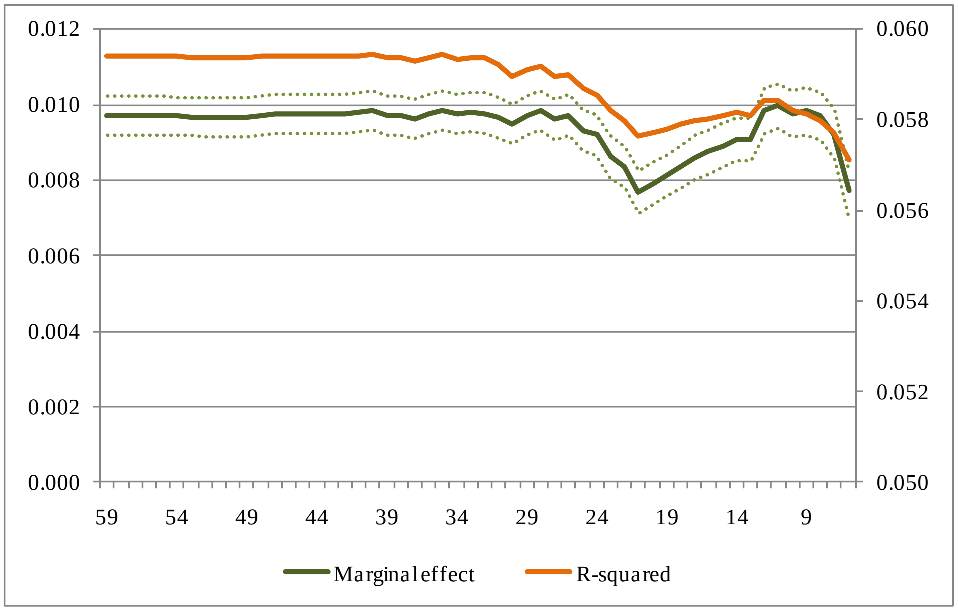


Figure 8B: Peru Census 1993, Cronbach alpha and Spearman rank correlations

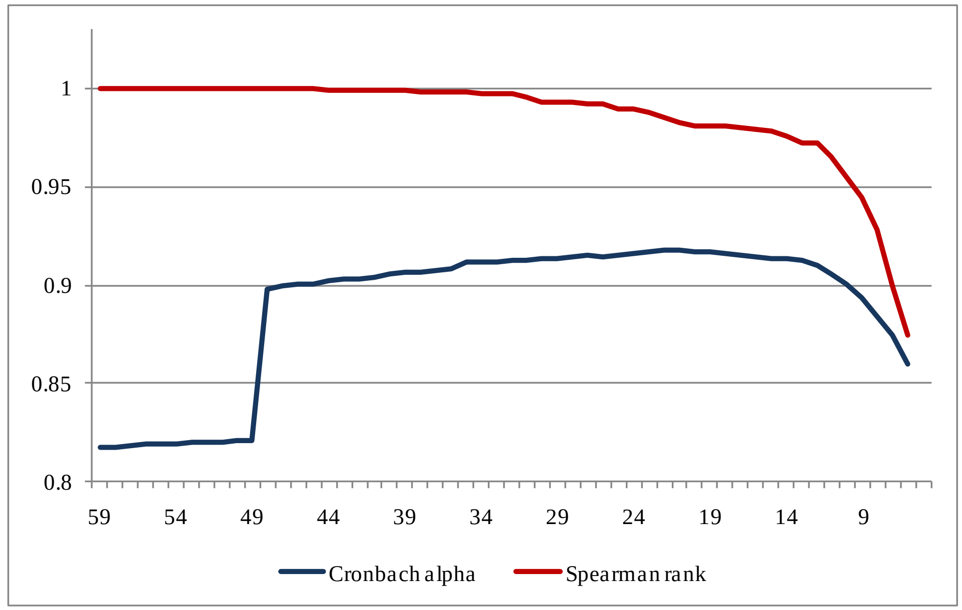


Figure 9A: South Africa Census 1996, School attendance regression results

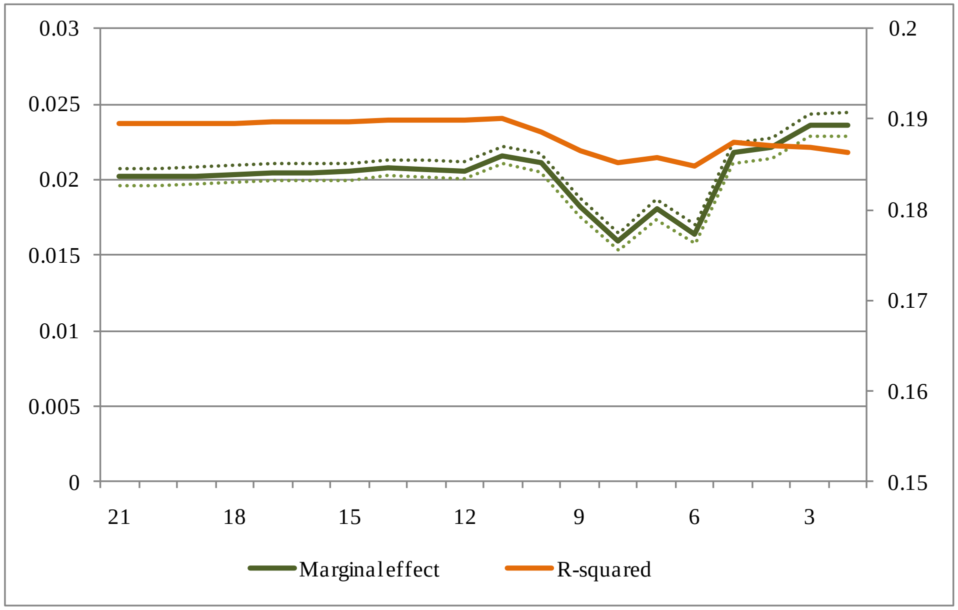


Figure 9B: South Africa Census 1996, Cronbach alpha and Spearman rank correlations

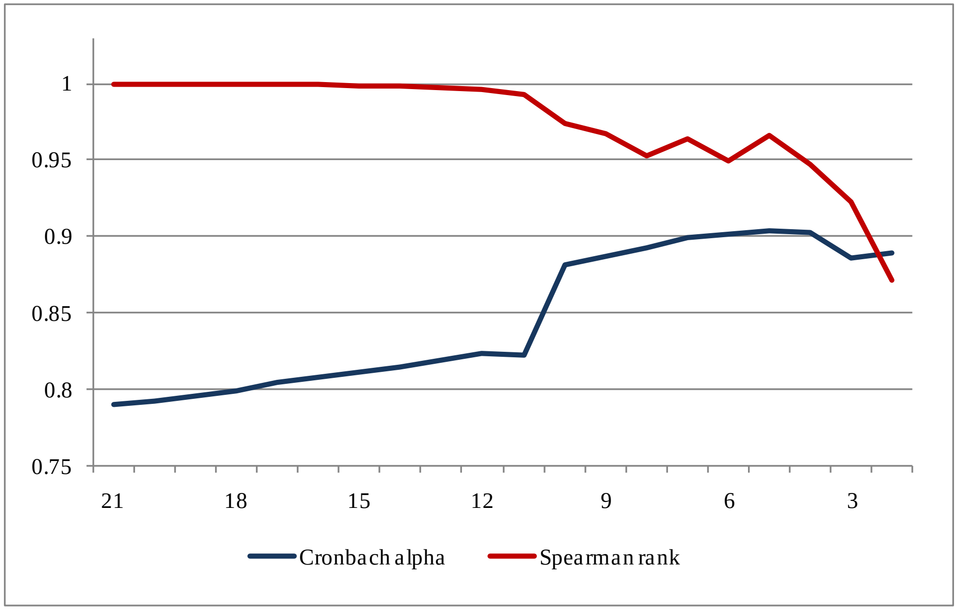


Figure 10A: Colombia Census 2005, School attendance regression results

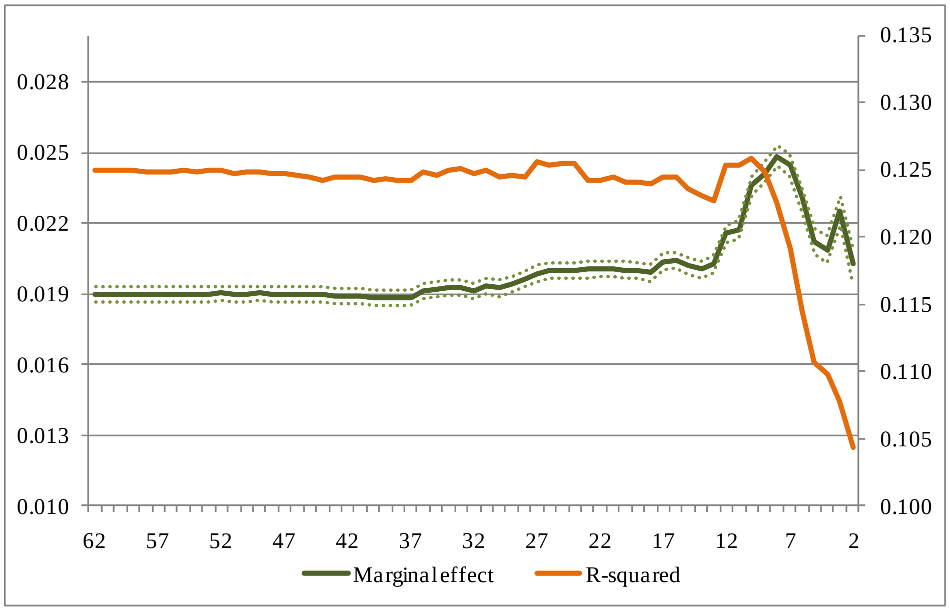


Figure 10B: Colombia Census 2005, Cronbach alpha and Spearman rank correlations

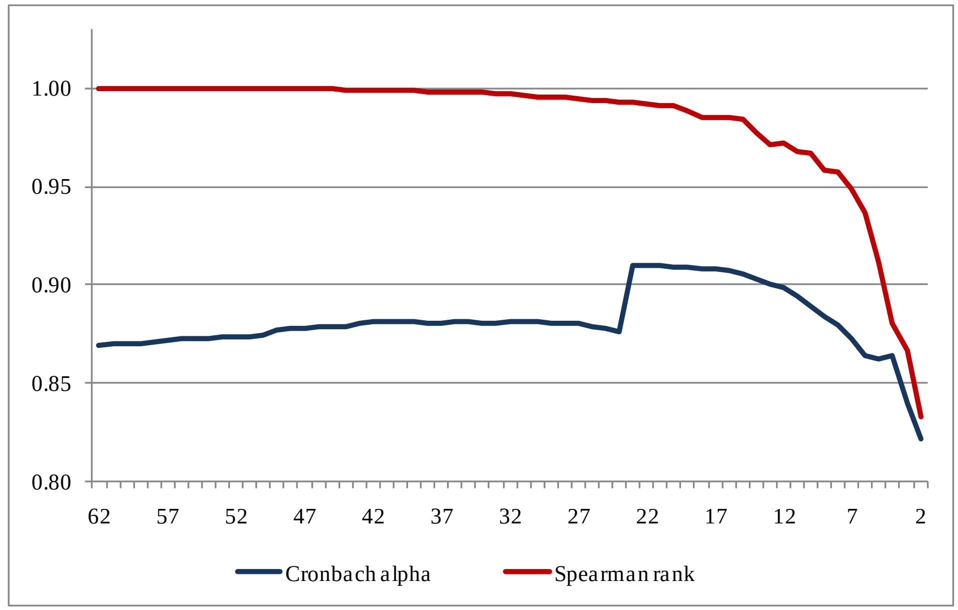


Figure 11A: Brazil Census 2000, School attendance regression results

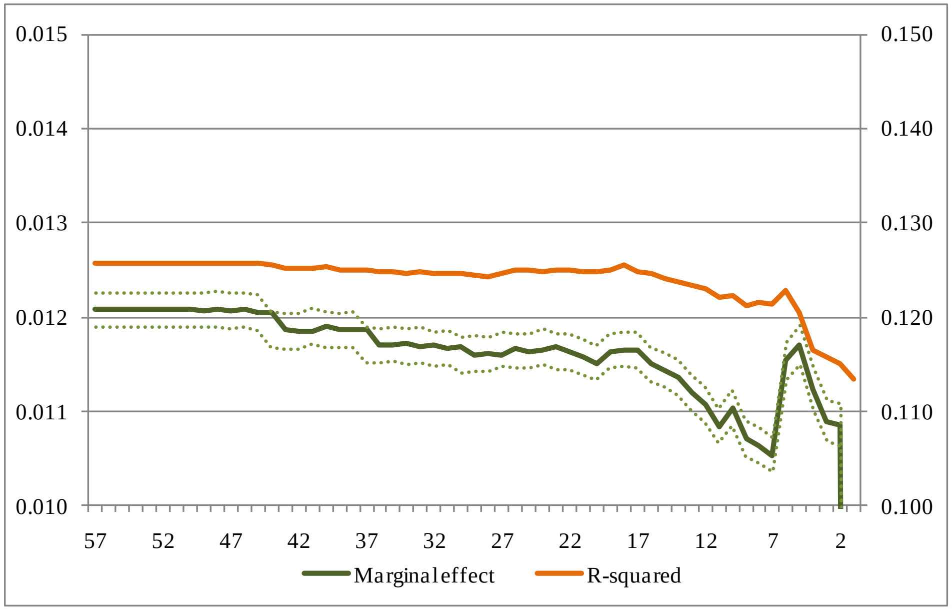
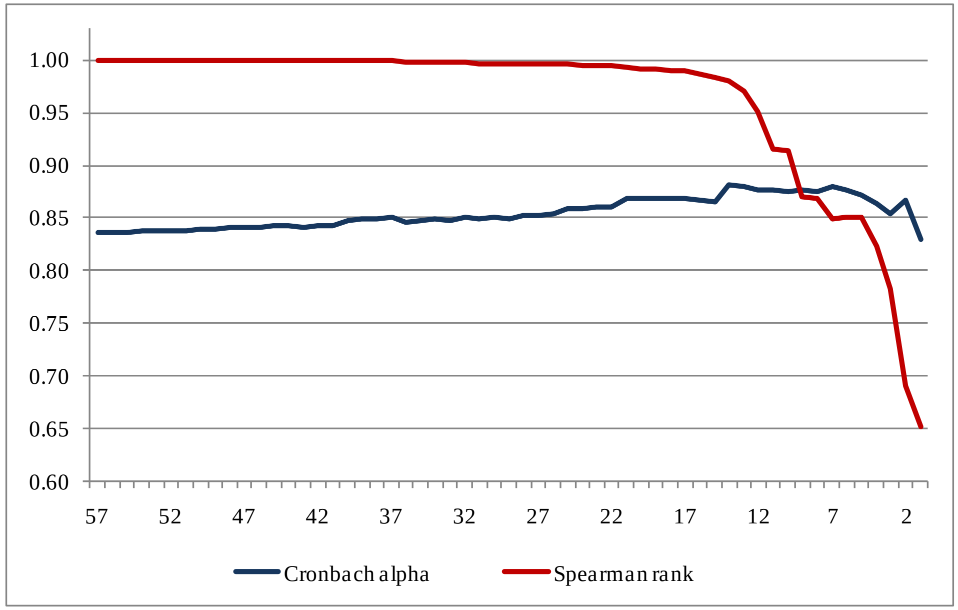


Figure 11B: Brazil Census 2000, Cronbach alpha and Spearman rank correlations



References

Assaad, Ragui; Levison, Deborah; and Nadia Zibani (2010). "The effect of domestic work on girls' schooling: Evidence from Egypt." In: *Feminist Economics*, 16(1), 79.

Booysen, Frikkie; Van Der Berg , Ser Vaas; Burger, Ronelle; Von Maltitz, Michael; Gideon Du Rand (2008). "Using an asset index to assess trends in poverty in seven sub-Saharan African countries." In: *World Development*, 36(6), 1133.

Cortina, Jose (1993). "What is coefficient alpha? An examination of theory and applications." In: *Journal of Applied Psychology*, 78(1), 98.

Ferguson, B.D.; Tnadon, A.; Gakidou, E.; and C.J.L. Murray (2003). "Estimating permanent income using indicator variables". Geneva: World Health Organization, Evidence and Information Policy Cluster Discussion Paper No. 44.

Filmer, Deon, and Lant Pritchett (1999). "The effect of household wealth on educational attainment: Evidence from 35 countries". In: *Population and Development Review*, 25(1), 85.

Filmer, Deon, and Lant Pritchett (2001). "Estimating wealth effects without expenditure data or tears: an application to educational enrollments in states of India." In: *Demography, 38*(1), 115.

Filmer, Deon, and Kinnon Scott (2008). "Assessing asset indices." Washington D.C.: World Bank, Policy Research Working Paper No. 4605.

Houweling, Tanja; Kunst, Anton; and Johan Mackenbach (2003). "Measuring health inequality among children in developing countries: does the choice of the indicator of economic status matter?" In: *International Journal for Equity and Health*, 2(8).

Kolenikov, Stanislav, and Gustavo Angeles (2009). "Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer?" In: *Review of Income and Wealth*, 55(1), 128.

McKenzie, David (2005). "Measuring inequality with asset indicators." In: *Journal of Population Economics, 18*(2), 229.

Minujin, Alberto and Joon Hee Bang (2002). "Indicadores de inequidad social. Acerca del uso del 'índice de bienes' para la distribución de los hogares." In: *Desarrollo Económico*, 42 (165), 129.

Montgomery, Mark; Gragnolati, Michele; Burke, Kathleen; and Edmundo Paredes (2000). "Measuring living standards with proxy variables." In: *Demography*, ***37***(2)**,**155.

Rutstein, Shea, and Kiersten Johnson (2004). "The DHS Wealth Index." Maryland: ORC Macro, DHS Comparative Reports No. 6.

Sahn, David, and David Stifel (2000). "Poverty comparisons over time and across countries in Africa." In: *World Development*, *28*(12), 2123.

Sahn, David, and David Stiefel (2003). "Exploring alternative measures of welfare in the absence of expenditure data". In: *Review of Income and Wealth*, 49 (4), 463.

Wall, Martin, and Deborah Johnston (2008). "Counting heads or counting televisions: Can asset-based measures of welfare assist policy-makers in Russia?" In: *Journal of Human Development*, 9(1), 131.