International Journal of Economics and Financial Issues

ISSN: 2146-4138

available at http: www.econjournals.com

International Journal of Economics and Financial Issues, 2023, 13(3), 126-140.

Measuring Distribution of Wealth in Zambia Using Census Micro Data: An Application of Principal Component Analysis

Floyd Mwansa*

University of Lusaka, Zambia. *Email: floyd.mwansa@unilus.ac.zm

Received: 02 February 2023 **Accepted:** 10 April 2023 **DOI:** https://doi.org/10.32479/ijefi.14301

ABSTRACT

Census data for Zambia was used to estimate the distribution of wealth in Zambia by constructing the Wealth Index as a measure of socioeconomic status using Principal Component Analysis. The reliability of the index is observed from three fronts; coherence, robustness and validity in representing household socioeconomic status. Classifying the households across all quartiles is highly consistent and robust. The index's performance in predicting the welfare distribution is analogous to established and most widely used methods from Demographic Health Surveys, as evidenced by similarities in the statistical distributions. Unlike other survey estimates, the index has been produced at the subnational level, such as district, enabling the classification of Zambia's districts according to their socioeconomic status. The index can be used to predict other socioeconomic outcomes, such as education and health, via Small Area Estimation techniques and determine district-level resource allocation by the central government.

Keywords: Principal Component Analysis, Wealth Index, Socioeconomic Status
JEL Classification: C1

1. INTRODUCTION

Measuring wealth is an essential and highly sought-after topic in economics; it dates back to Adam Smith's 1776 book "An Enquiry into the Nature and Causes of the Wealth of Nations" This was a classic work that advanced the understanding of how wealth is created, distributed, and measured in society.

Most, if not all, growth models in economics are based on the creation of wealth by utilizing factors of production because any welfare improvement or poverty reduction cannot occur without society's welfare improvement. This paper focuses on measuring wealth distribution in society as an indicator of socioeconomic status (SES). As noted by Cassiers and Thiry 2014, the pursuit of new ways of measuring socioeconomic improvement is critical since it relates to society's welfare, and hence scholars and policymakers need to look beyond the already established measures such as Gross Domestic Product (GDP).

In many countries, including Zambia, estimates of socioeconomic indicators are from surveys such as the Living Conditions Monitoring Survey (LCMS), which collects household information on income, consumption, asset ownership, and socio-demographics (LCMS, 2015)[1]. All these estimates are used as proxy measures of household wealth, but surveys of this nature only produce regional or province-level estimates without indicating at a sub-national level, such as districts, hence the need to have more valid estimates for the district level.

Household prosperity vis-a-vis improvements in socioeconomic status is mainly measured through the income earned or expenditure incurred in a specified period without considering the broader characteristics such as asset ownership, access to dwelling facilities such as water, sanitation and source of energy. One would argue that such factors are highly correlated with either

1. At the time of conducting this research, the nost latest Living Condition Monitoring Survey was done in 2015

income or expenditure at the household level. Within available economic literature, there seems to be some ambiguity regarding over-reliance upon measuring socioeconomic status using income or expenditure survey-based methods relying heavily on reported statistics which have been argued to be prone to some statistical bias. Hence measures based on observed characteristics such as asset ownership and access to facilities can help improve the reliability of the measurement of household well-being.

As an alternative to the measurement of household SES using household asset ownership and dwelling characteristics, we employ census data in building the Wealth Index (WI) as a measure of socioeconomic well-being at a sub-national level. WI, a composite estimate based on asset ownership and other dwelling characteristics, represents a household's long-run SES as opposed to income and expenditure estimates, which are argued to provide a short-run scenario (Mathieu et al., 2019).

Measuring SES using the WI as advanced in this research has more advantages because a common problem with household income and consumption expenditures is their unpredictability. Income and expenditure are unstable in less developed countries and prone to seasonal effects. Despite considerable literature on the measurement of household well-being, many scholars, such as Rodrigo Lovaton Davila et al. (2022), have questioned the use of household income or expenditure collected through surveys as a perfect measure of household SES. Additionally, previous research has revealed that measures of SES which relies on observable characteristics than on reported data, such as income or expenditure, have been found to pose less bias and are more reliable; Filmer and Scott (2012), Tarozzi and Deaton (2009) and Christiaensen et al. (2012). As Howe et al. (2008) argued, the difficulties associated with using household expenditures or income as surrogates for SES suggest that the WI can be considered a superior alternative in measuring wealth distribution.

Researchers and policymakers acknowledged a significant knowledge gap in measuring SES at the household level. Hence, this study contributes to the economic literature on the measurement of household wealth using census microdata to show wealth distribution at a sub-national level. This research has both academic and policy implications. From an academic standpoint, this study has advanced more rigorous tools by utilizing readily available observational country-level census data in constructing the WI. Whereas from a policy standpoint, the study has provided policymakers, program managers, and others with clear evidence regarding the distribution of household wealth at subnational levels, such as provinces and districts, which are usually absent from available surveys.

Additionally, the motivation for this research was shaped by the need to understand the distribution of wealth so that it can be used as a predictor variable in determining socioeconomic outcomes such as education and health and as an input in Small Area Estimation (SAE) techniques such as Area Level models.

2. LITERATURE REVIEW

2.1. Principle Components and Wealth Index Model

Principal Component Analysis (PCA) is a statistical technique within the multivariate analysis. The earliest users of PCA were Pearson (1901) and Hotelling (1933) and more recently by Mariolis and Tsoulfidis (2018), Lovaton Davila et al. (2022) and Tsoulfidis and Athanasiadis (2022). The intricate idea of PCA is to try to define the variation in the variables in a set of multivariate data as sparingly as possible using a set of derived uncorrelated variables, utilizing the particular linear combination of variables in the original data (Rabe-Hesketh and Everitt, 2007). PCA reduces the dimensionality of large datasets such as Census microdata while maximizing interpretability and minimizing data loss (Jolliffe and Cadima, 2016). PCA creates orthogonal linear combinations of uncorrelated indices called the Principal Components (PCs) from a set of variables by assigning weights to each particular variable according to their contribution to the overall variability.

As a descriptive tool, PCA requires no distributional assumptions and, as such, is primarily an adaptive exploratory method that can be in use on various data types. Although it is commonly assumed that the dataset has a multivariate normal (Gaussian) distribution for inferential purposes, PCA as a descriptive tool does not require any distributional assumptions (Jolliffe and Cadima, 2016).

Hence, generating principal components from variable X_1 through to X_n will take the general form;

$$PC_m = a_{m1}X_1 + a_{m2}X_2 + \ldots\ldots + a_{mn}Xn \quad (1)$$

Where a represents the weight for the *mth* principal component and the *nth* variable

In this paper, we used PCA to construct a WI following the method advanced by Lovaton Davila et al. (2022), as shown in equation (2);

$$WI_i = w_1 A_{1i} + w_2 A_{2i} + \ldots + w_k A_{ki} \quad (2)$$

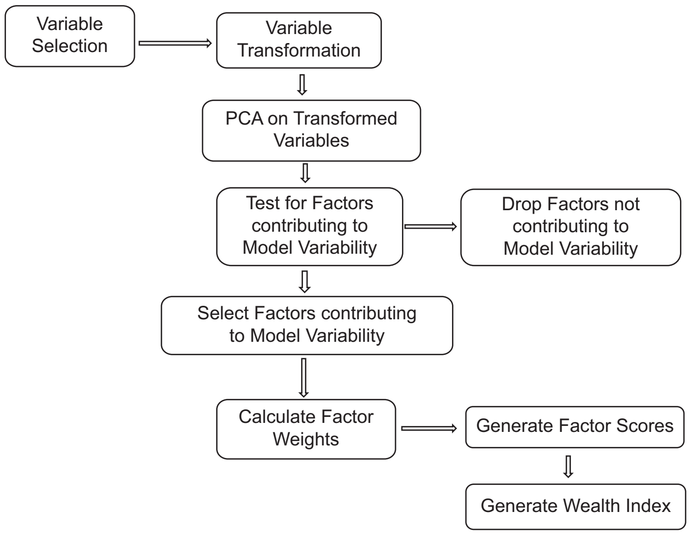
Where; WI_i is a calculated index for the ith household, A_{ji} is an indicator of the presence of a variable of interest for ith household and w_j is the weight associated with the jth asset (j = 1,2……k).

Based on equation (2), PCA is a data reduction technique that creates an orthogonal linear combination of variables (A_{Jl}) and, in turn, assign an associated weight based on each variable's contribution to the overall variability (Jolliffe and Cadima, 2016). Section 3.2 and Figure 1 present a detailed estimation of using census microdata to derive WI using PCA

2.2. Application of PCA in Measuring Socioeconomic Status

There are several methods in economics and statistics for combining multiple variables or indicators into a single univariate index. Several studies, including Howe et al. (2008 & 2009), Filmer and Scott (2012), and Ngo and Christiaensen (2018), claim that the construction method used has a significant impact on index performance. However, it is unknown whether this is correct. Other techniques, such as the inverse frequency index (INV) developed by Morris et al. (2000) and the

Figure 1: Wealth index algorithm



Source: Author's construction based on reviewed literature

dichotomous hierarchical ordered probit (DHP) method developed by Ferguson et al. (2003), are used in addition to PCA-based indices. PCA remains the most widely used technique due to its mathematical foundation and simplicity (Tsoulfidis and Athanasiadis, 2022).

The most common application of PCA is the WI constructed by the Demographic and Health Survey (DHS) program. The DHS program constructs the wealth index using PCA-based SES measures based on household characteristics such as asset ownership and dwelling rather than income or expenditure. The constructed WI in the DHS is used to explain and predict the distribution of health outcomes across various SES groups. Aside from the DHS-advanced application of PCA in the health sector, PCA has widespread use in economic and regional geography (Vom Hofe and Bhatta, 2007; Ramos and Moreno, 2013; Lovaton Davila et al., 2015 and 2022), finance (Plerou et al., 2002; Farné and Vouldis, 2021), and environment and climate change modelling (Christian Borja-Vega and Alejandro de la Fuente, 2013).

Filmer and Pritchett (2001) conducted the first and most notable research on using PCA to construct the WI, using household asset indicators to replace income or consumption. The basis of their research was the analysis of household assets for India; they went on to validate their results using data on household assets and consumption for other countries similar to India in terms of socioeconomic make-up, such as; Indonesia, Pakistan and Nepal. Their research found that using PCA to estimate the WI provides a reasonable and robust proxy for household SES. They also argued that asset-based measures depict an individual's or a household's long-run economic status rather than short-term measures derived from income or expenditure measures, immune to short-term fluctuations in measuring economic well-being and economic shocks (see also the argument advanced by Howe et al., 2008). Filmer and Pritchett also acknowledged that WI is likely to be correlated with either income or expenditure, as indicated by previous research by Gasparini et al. (2008) and Lora (2008). However, the two socioeconomic measures are likely to depict different levels of economic well-being because there may be disparities in household wealth rankings based on asset indices versus those based on consumption expenditure, Lovaton Davila et al. (2022), and this is the area that should be researched further to tease out and understand the source of such disparities. Other studies, such as Ngo and Christiaensen (2018), argued that the asset-based approach to estimating SES is consistent with other measures, and the asset-based WI could be used as a reliable indicator of SES.

Further research on the use of census microdata to construct WI was done by Lovaton Davila et al. (2022), who used PCA to develop a valid and consistent measure of SES at the household level. The computed WI was based on asset ownership, utilities, and dwelling characteristics. Validation strategies include comparing the proposed index to widely used DHS WI and confirming socioeconomic gradients on school enrollment and educational attainment. Additional statistical tests, such as kernel distribution analysis, were also performed, revealing that the measure was reliable. In addition, their findings revealed that the wealth index consistently positively affected educational outcomes. Furthermore, the methodology employed by Lovaton Davila et al. (2022) suggested which assets are more significant for defining household SES.

For survey-based SES indices, Mathieu et al. (2019) evaluated alternative methodologies to a DHS-based WI, such as; count measures, item response theory, Mokken scale analysis, multiple correspondence analysis, polychoric PCA and predicted income. They observed that statistical validity, ease of calculation, consistency of results and empirical plausibility should be the primary determinant in the choice of methodology when constructing WI.

According to the literature reviewed in this article, even though there are a variety of variables to use in measuring SES, including income, expenditure, consumption, asset ownership, and dwelling characteristics. Focusing on data availability, completeness, and validity is critical, even at the most basic level, such as a municipality. With most surveys conducted worldwide designed to produce valid estimates at a national or regional level, deriving PCA-based WI using census microdata at the district level is highly valuable.

On the other hand, one must be cautious about how far PCA can be used in the construction of an SES index because other researchers regard PCA as only reflecting artificially constructed indices, branding the practice as highly arbitrary, particularly when it comes to deciding how many components contribute to model variability and which variables to include in the analysis. This study used well-established and well-tested procedures to determine which components to include in the final WI computation to ensure that we produce an empirically plausible index following our literature review. Any use of a PCA-constructed WI must be accompanied by testing the reliability in terms of coherence, robustness, and validity, which is what this study did. Doing so will help refute criticisms levelled against methodologies based on PCA.

3. METHODOLOGY

3.1. Data Source and Variable Selection

In this study, we used the Zambian Census 2010[2] 10% sample, which included 279,271 households with a

2 2010 census dataset was the most latest at the time of this research as the 2020 was only conducted in 2020 are the full results with the accompying data are yet to be disseminated by the Zambia Statistical Agency.

comprehensive national representation, Annexe 1. The Census data contains all the information on the broad spectrum of the population and household characteristics. The census data contains variables of interest such as; household asset ownership, access to water and sanitation services and housing characteristics. Annexe 2 contains all the variables used in calculating the WI.

3.2. Model Estimation Procedure

Specifically, our PCA used 27 variables from census microdata to extract WI. We group the variables into three thematic areas; asset ownership, dwelling characteristics, and the presence of utilities. Following the procedure outlined by Filmer and Pritchett (2001), all variables, including categorical variables, were dichotomized after transformation as mentioned above (Annexe 2).

We took the following steps to simplify our methodology:

1. Identify selection variables from the census data set that are likely to contribute to overall variability during the construction of WI;
2. Variable transformation into binary choice variables following the procedures by Filmer and Pritchett (2001);
3. Perform PCA on all transformed variables;
4. Select variables contributing more to model variability;
5. Calculate factor weights for all selected variables from 4;
6. Generate factor scores for all variables contributing to model variability; and
7. Generate non-standardized and standardized WI.

When conducting PCA, it is essential to remember that there is no established statistical procedure or assumption on choosing variables to include in the analysis. The availability and applicability of variables and any guidance gleaned from prior research are the primary factors that tend to impact the choice of variables.

In measuring WI, various studies have used different variables. The absence of best practices in variable selection in computing indices related to SES is observed by Montgomery et al. (2000). The model used in this study is from DHS, where variables which relate to; durable asset ownership, access to utilities and infrastructure (e.g. sanitation facility and source of water), and housing characteristics where selected. After variable selection, descriptive analyses were performed for all variables to inform decisions on data management issues, such as variable categorization (Annexe 2).

Another pre-diagnostic procedure done before PCA was the standardization of variables through variable transformation, where dummy variables for all categorical variables were created since PCA is sensitive to differences in the units of measurement of variables (Bolch and Huang, 1974).

4. PRINCIPAL COMPONENT ANALYSIS RESULTS

After converting the initial categorical variables, we used 103 discrete variables to run our PCA. The PCA results are shown in Annex 3, along with the component results and Eigenvalues. Most statisticians have used two specific methods in determining which components to include in the analysis of results: Kaiser's criterion, or the eigenvalue rule, and Catell scree plot techniques. With Kaiser's criterion, only those factors with an eigenvalue (the variances extracted by the factors) of 1.0 or more are retained.

Following Kaiser's criterion, we retained 43 components (Annexe 3), which Catell's scree test confirmed (Figure 2). The scree Plot depicts the plots of each component's eigenvalues and only considers the components above the marked line.

The PCA results (Annexe 3) show that 43 components accounted for 79.05% of the total variance in the data set. A look at the factor loadings (Annexe 4) revealed the following:

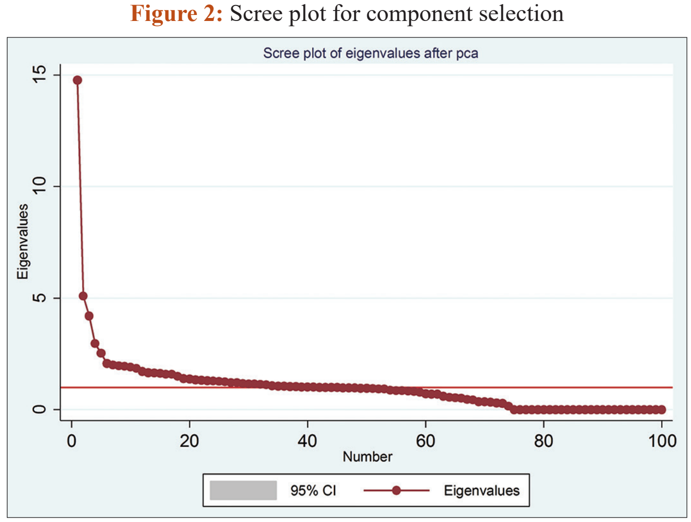
- The presence of utility variables such as; lighting, cooking and heating with electricity are associated with positive loadings. Improved sources of drinking water and sanitation also showed positive loadings, which is also the case for having an improved way of managing solid waste;
- Improved dwelling characteristics, such as; living in a house with walls and floors made with concrete and a roof made out of iron sheets, are all associated with positive loadings; and
- Asset ownership of all forms was also associated with positive loadings

In the PCA framework, loadings represent correlations and a correlation between a component and variable estimates. Indication of positive loadings in variables related to the presence of utilities, dwelling characteristics and asset ownership indicates that they are positively related to wealth.

5. HOUSEHOLD WELFARE INDEX

When calculating the Wealth Index (WI), determining the factor score coefficients or component score values was the first step. Since PCA is an estimation command like regression, we predicted the factor scores after PCA. A determination was made regarding the number of factors needed in our analysis, capturing our model's more comprehensive explanation power and 43 scores

Figure 2: Scree plot for component selection



Source: Author's construction from census data for Zambia were predicted according to the number of components with an Eigenvalue greater than 1. In our analysis, we used STATA statistical software.

Following the calculation of factor scores, appropriate weights for each factor were determined because the variations, as captured by the size of the Eigenvalue, are always different. The calculation of weights was accomplished by dividing the PCA respective component proportion by the component 43 cumulative proportion. The first component weight, for example, was calculated by dividing 14.34% by 79.05%. The computed weights were multiplied by the predicted factor score to obtain the weighted factor score of each observation in the data set, which was then employed to aggregate all of the scores into a single index to reflect a measure of the wealth of a Non-Standardized Wealth Index (NWI) as shown in equation (3).

$$PC43\,NWI = \left(\frac{PC1pr}{Cpr\ for\ PC43}\right) * PC1 + \left(\frac{PC2pr}{Cpr\ for\ PC43}\right) * PC2 + \ldots + \left(\frac{PC43pr}{Cpr\ for\ PC43}\right) * PC43 \quad (3)$$

Where;
NWI = Non Standardized wealth Index,
PC1pr = Percentage proportion for PC1
PC2pr = Percentage proportion for PC2
PC1 and PC2 = Respective predicted scores for both Component 1 and 2

Cpr for PC43 = Cumulative percentage proportion at PC43

The resulting NWI measures the wealth status of one household relative to the other on a linear scale. NWI can be positive or negative, creating difficulties in terms of interpretation. Hence, the need to develop a Standardized Wealth Index (SWI) with values ranging from 0 to 100, using the formula:

$$SWI = \left(\frac{NWI - minNWI}{maxNWI - MinNWI}\right) * 100 \quad (4)$$

Where;
SWI = Standardized Wealth Index
NWI = Non-Standardized Wealth Index
minNWI= Minimum value of Non-Standardized Wealth Index
maxNWI= Maximum value of Non-Standardized Wealth Index

The standardization procedure used in this research is based on the work of other researchers, such as; Antony and Rao (2007), Hightower (1978), and Sekhar et al. (1991). The procedure makes interpretation easier because the natural interpretation is that the higher the value, the higher the wealth status of a particular household or district, and so on.

5.1. Household Wealth Distribution in Zambia

Household wealth distribution in Zambia is not uniform across various groups, according to our calculated WI. Figure 3 shows that our index skewed to the left in highly urbanized set-ups and the right in rural areas. Our provincial analysis also reveals that the most urbanized provinces, such as Lusaka and Copperbelt provinces have a higher ranking than their rural counterparts (Annex 5)

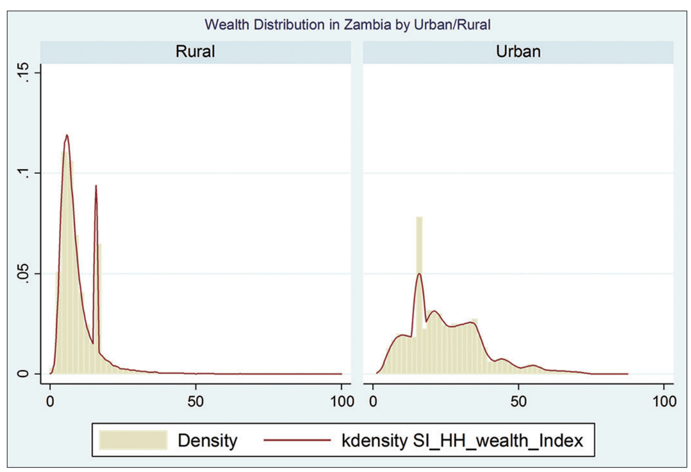
Furthermore, we produce five groups (quantile) ranging from the least advantaged (1st quantile) to the most advantaged (5th quantile). Table 1 shows how these quantiles are transformed into a poverty scale.

To prove variability across the five groups created, we conducted Bartlett's test of equal variances and Levene's test of homogeneity of variances to test whether the variance is the same for each group. Bartlett and Levene's homogeneity tests assume that the variances in the populations from different samples are equal. For example, if the resulting P-value of both Bartlett and Levene's test is less than the critical value, then differences in sample variances are unlikely to have happened unintended. The test results showed a significant level of 0.000 for both, a value enough to reject the hypothesis (the probability should be <0.05 to reject the null at a 95% confidence interval). Hence, our Wealth Index demonstrated considerable variability across groups.

5.2. District-Level Distribution of Wealth in Zambia

The WI calculated in the research allowed the ranking of regions based on their mean score—two rankings were performed; provincial and district. At the provincial level, Lusaka ranks the highest, followed by Copperbelt, with the last two being Muchinga and Northern. Since the WI represents SES, its performance was compared with Zambia's Incidence of poverty from the LCMS

Figure 3: Regional comparison of wealth distribution in Zambia



Source: Constructed from the 2010 census data

Table 1: Wealth index and poverty ranking

WI quintile	Poverty rank 3	Mean	SD
1	Poorest	4.308	1.024
2	Poorer	7.258	0.911
3	Middle	13.264	2.462
4	Rich	20.031	2.441
5	Richer	36.253	10.038

Bartlett's test for equal variances: χ^2 (4)=4.205, P>χ^2=0.000; Levene's test for homogeneity: W_0=47817.385, df (4, 279266) *P*>F = 0. SD: Standard deviation

(Central Statistics Office (CSO)[3], 2015). Our index is almost consistent as the top four wealthiest provinces show the lowest incidence of poverty based on the 2015 LCMS, as seen in the Table 2.

The performance of the WI at the district level is better in urban-based districts than in rural ones. This is predictable because of the aspects of the index in terms of its constituencies; the presence of utilities, asset ownership and housing characteristics are much more expected in the most urbanized set-up, as shown in Figure 4. Our index was compared with the poverty levels across Zambia districts extracted from the World Bank 2015 report "*Mapping Subnational Poverty in Zambia*". Districts with the lowest poverty level showed a relatively better index score than districts with high poverty levels. The spatial district-level distribution of wealth in Zambia is shown in Figure 4 and Annexe 6.

3 The Zambia Central Statistics Office has now called Zambia Statistics Agency (https://www.zamstats.gov.zm/)

6. RELIABILITY OF THE WEALTH INDEX

Since the WI is argued to be a measure of socioeconomic status, any research must produce an index that is reliable in classifying various grouping according to their distribution of wealth. We test our index from three fronts; coherence, robustness and reasonableness.

6.1. Internal Coherence

Table 3 shows wealth classification regarding dwelling characteristics, asset ownership and utilities access. We note

Table 2: Comparative measure of wealth index and poverty measurement

Province	Mean WI score (%)	Provincial rank based on wealth index	Incidence of poverty 4 (%)	Provincial rank based on poverty incidence 5
Lusaka	27.48	1	20.2	1
Copperbelt	22.56	2	30.8	2
Southern	14.46	3	57.6	4
Central	12.99	4	56.2	3
Western	11.21	5	82.2	10
North West	10.68	6	66.4	5
Eastern	10.37	7	70.0	7
Luapula	9.84	8	81.1	9
Muchinga	9.30	9	69.3	6
Northern	8.73	10	79.7	8

Source: Author's calculations. WI: Wealth index

Figure 4: Wealth distribution in Zambia

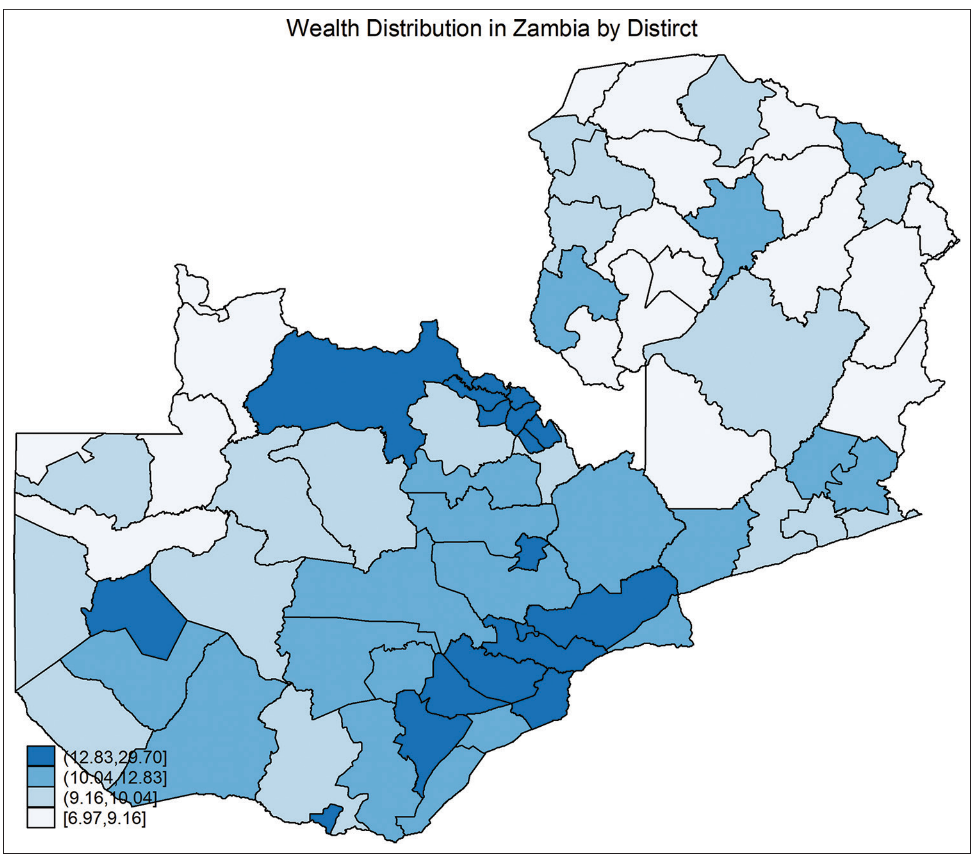


Table 3: Wealth classification and household characteristics

Variable type	Wealth classification (%)				
	Poorer	Poorest	Middle	Richer	Richest
Dwelling characteristics					
Housing type - traditional	29.45	29.49	32.61	6.04	2.42
Housing type - conventional	0.09	0.01	11.31	28.85	59.75
Roof materials - thatch/palm leaf	34.84	39.28	23.95	1.61	0.32
Roof materials - metal/iron sheets	4.61	0.52	29.38	24.79	40.71
Building materials - low quality	28.55	29.04	30.90	6.74	4.76
Building materials - high quality	1.07	0.10	18.45	26.95	53.45
Asset ownership					
Own radio	18.22	15.89	18.31	14.98	32.59
Own television	3.03	0.2	11.18	18.32	67.27
Own refrigerator	0.06	0	0.85	2.72	96.37
Own bicycle	23.25	33.2	19.06	10.35	14.15
Own motor vehicle	0	0	0.94	4.82	94.25
Access to utilities					
Regularly collected solid waste	5.16	0.59	6.44	13.69	74.13
Flush toilet	0.54	0.06	1.41	11.03	86.96
Pit latrine	20.16	29.64	18.29	17.69	14.22
Access to safe drinking water	15.84	11.52	17.06	20.29	35.29
Access to unsafe drinking water	32.65	39.68	20.53	5.41	1.73
The main source of lighting (electricity)	0.19	0.02	0.9	9.21	89.68
The main source of cooking (electricity)	0.01	0	0.1	3.66	96.23
The main source of heating (electricity)	0.01	0	0.12	2.74	97.13

Source: Author's computations from zambia 2010 census data

huge variations in wealth classifications (poorest, poorer, middle, richer and richest). 60% of the households classified as richest live in conventional and improved houses compared to the poorest with less than a percent. Similarly, regarding the quality of building materials, the richest tend to live in dwellings built with high-quality materials (54%) compared with the poor (<2%).

Our Wealth Index is consistent in classifying the rich versus poor households with community-level variables such as access to utilities and dwelling characteristics or household-specific variables such as asset ownership. With our calculated index, we are comforted by the clear distinction across poorer, poorest, middle, richer and richest asset ownership variables (housing) not related to dwelling characteristics (infrastructure), such as "Improved conventional housing type" (<1% of the poor (poorer and poorest) versus 89% of the rich (richer and richest).

6.2. Accessing the Robustness of Wealth Index

The PCA constructing the WI works like an estimation command in standard statistical procedures. Hence the robust check will examine how the index estimate behaves when the PCA specifications are modified by removing or adding variables. Table 4 reports household classifications based on different wealth classes from two different models; Model 1 included all variables, whereas Model 2 only included variables on dwelling characteristics and the presence of utilities. Table 4 clearly shows that there is not much overlap between the classifications of the wealth classes based on these two models.

The proportion of households classified to be poor in both models is similar (40%); this is also true for the richest category. This moderately small overlap between the wealth rankings under these two models highlights that our wealth index is robust enough to classify SES.

Table 4: Comparison of the classification of Zambian households on alternative wealth index measures

Wealth classes	Households (%)	
	Model 1 (all)	Model 2 (dwelling characteristics and utility presence)
Poorer	55,854 (20.0)	55,987 (20.1)
Poorest	55,855 (20.0)	55,895 (20.0)
Middle	75,241 (26.9)	69,298 (24.8)
Richer	36,471 (13.1)	42,686 (15.3)
Richest	55,850 (20.0)	55,405 (19.8)
Total	279,271 (100.0)	279,271 (100.0)

Source: Author's PCA analysis of Zambia 2010 Census data. PCA: Principal component analysis

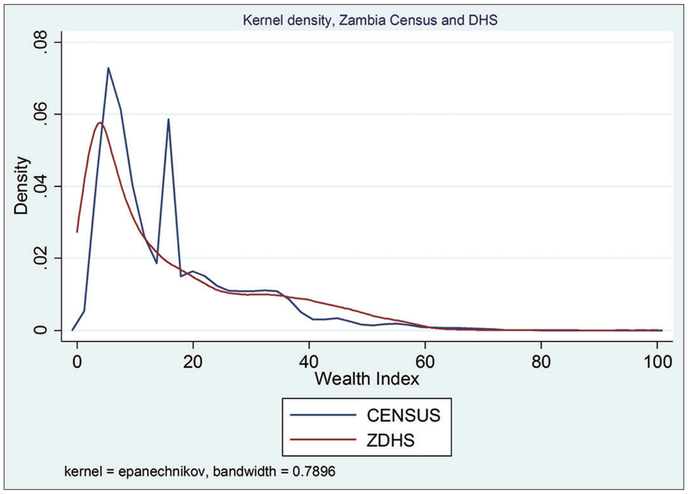
6.3. Comparison of Wealth Index with Alternative Measures of Socioeconomic Status

In Zambia, the established measures of household SES are conducted in two surveys; the WI from ZDHS and the poverty estimates from the Living Conditions Monitoring Survey (LCMS). To prove the reliability of our WI, we compared the estimates for both the socioeconomic measures from ZDHS (CSO et al, 2014) and LCMS with our calculated index. Further district-level estimates are compared with the World Bank 2015 Report "*Mapping Subnational Poverty in Zambia*".

6.4. Comparison with Demographic Health Survey

We used a graphical comparison (Figure 5) of the distribution of wealth based on our index (census) and the index derived from the ZDHS data to evaluate the performance. This comparison helps further in the question of validity for our index by verifying that it measures wealth and not some other phenomenon associated with asset ownership, housing characteristics, and access to utilities.

Figure 5: Wealth index scores, 2010 census and ZDHS



Source: Constructed from the 2010 census and 2013 demographic health survey data for Zambia

Table 5: Summary statistics for wealth index 2010 census and 2013 Zambia demographic and health survey

Statistical measure	Census	ZDHS
Mean	15.7526	15.88575
Median	12.07062	10.25589
25th percentile	6.441969	4.287017
75th percentile	20.97644	24.2142
SD	12.31534	14.73292
Skewness	1.504454	1.125589
Kurtosis	5.360189	3.421562

Source: Author's calculations from both 2010 census household data and 2013 ZDHS. ZDHS: Zambia demographic and health survey, SD: Standard deviation

Furthermore, executing the comparisons, the two indices were normalized for them to be measured on the same scale; this was done by subtracting the minimum value and dividing it into the difference between maximum and minimum.

The kernel densities for the two indices are depicted in Figure 5. The shapes of the kernel densities for the two indices are comparable, with only a tiny amount of variation between them. This variation may be explained by the fact that the sets of variables used for each dataset are not the same (and also due to dissimilarities in data gathering). As shown in Table 5, the summary statistics, such as the mean, median, skewness, kurtosis and percentiles, are highly comparable.

6.5. Comparison with Living Conditions Monitoring Survey

According to the 2015 LCMS, the national poverty rate stood at 54.4%; with our Wealth Index, poverty levels were estimated to be 40% combining the poorest and the poor categories. The difference is because the LCMS poverty is calculated from the expenditure side while our index estimated wealth based on; asset ownership, utility presence and characteristics of housing conditions. Using the poverty mapping at the district level as provided by the World Bank, we estimated the relationship between district poverty levels and each district's WI mean score. The rank correlation is −0.84 ($P < 0.001$, $n = 74$), meaning that our index is negatively related to poverty levels, as can be seen in Annexe 6, where the districts with a high mean score of wealth index have low levels of poverty.

7. CONCLUSION

This paper presented the use of census microdata in constructing a wealth index to represent household SES using PCA procedures. In recent times, PCA has become a popular method for computing indices representing socioeconomic standing over the long run. Our estimated WI is very reliable in coherence, robustness and validity. The index is consistent in classifying households across all quartiles in terms of the relationship between the poor and the rich and the variables at the community level. We examined the index's robustness by fitting it into two different models, and since the results did not show a greater degree of overlap between the estimated indexes, we can conclude that our index is robust and internally consistent. As demonstrated by the kernel density function and the comparison of other summary statistics such as the mean, variance median and kurtosis, the performance of our index in predicting the welfare distribution is comparable to that of other well-established methods that are utilized by the vast majority of the time. These methods come from both the DHS and the LCMS.

Our index is highly reliable at sub-national levels, such as the district level, because census microdata was used with sufficient sample sizes, in contrast to ZDHS and LCMS, whose estimates are only valid at the provincial level due to sampling issues. The index has also made it possible to classify districts according to their socioeconomic standing, which is very important for future studies involving the application of the index to the prediction of other socioeconomic outcomes, such as education and health, through the use of SAE techniques.

REFERENCES

Antony, G. M. & Rao, K. V. (2007). A composite index to explain variations in poverty, health, nutritional status and standard of living: Use of multivariate statistical methods. Public Health, 121, 578-587.

Bolch, B.W., Huang, C.J. (1974), Multivariate Statistical Methods for Business and Economics. Hoboken: Prentice-Hall.

Borja-Vega, C., de la Fuente, A. (2013), Municipal Vulnerability to Climate Change and Climate-Related Events in Mexico, Policy Research Working Paper Series 6417. United States: The World Bank.

Cassiers, I., Thiry, G. (2014), A high-Stakes Shift: Turning the Tide from GDP to New Prosperity Indicators. London: Routledge.

Central Statistical Office (CSO), Ministry of Health (MOH), Tropical Diseases Research Centre (TDRC), University of Zambia, and Macro International Inc. (2014), Zambia Demographic and Health Survey 2013-14. Calverton, Maryland, USA: Central Statistical Office and Macro International Inc.

Central Statistical Office. (2015), Living Conditions Monitoring Survey 2015. Lusaka: Living Conditions Monitoring Branch.

Christiaensen, L., Lanjouw, P., Luoto, J., Stifel, D. (2012), Small area estimation-based prediction methods to track poverty: Validation and applications. Journal of Economic Inequality, 10(2), 267-297.

De La Fuente, A., Murr, A.E., Ramirez, R., Ericka, G. (2015), Mapping

Subnational Poverty in Zambia (English). Washington, DC: World Bank Group. Available from: https://www.documents.worldbank.org/curated/en/766931468137977527/mapping-subnational-poverty-in-Zambia

Farné, M., Vouldis, A. (2021), Banks' business models in the euro area: A cluster analysis in high dimensions. Annals of Operations Research, 305(1), 23-57.

Ferguson, B., Tandon, A., Gakidou, E., Murray, C. (2003), Estimating Permanent Income Using Indicator Variables. Geneva: World Health Organization.

Filmer, D., Pritchett, L.H. (2001), Estimating wealth effects without expenditure data-or tears: An application to educational enrollments in states of India. Demography, 38(1), 115-132.

Filmer, D., Scott, K. (2012), Assessing asset indices. Demography, 49(1), 359-392.

Gasparini, L., Escudero, W.S., Marchionni, M., Olivieri, S. (2008), Income, Deprivation, and Perceptions in Latin America and the Caribbean: New Evidence from the Gallup World Poll. La Plata, Argentina: Latin American Research Network, Inter-American Development Bank, and Center for the Study of Distribution, Labor and Social Affairs (CEDLAS).

Hotelling, H. (1933), Analysis of complex statistical variables into principal components. Journal of Educational Psychology, 24, 498-520.

Hightower, W. L. (1978). Development of an index of health utilizing factor analysis. Medical Care, 16, 245-55.

Howe, L.D., Hargreaves, J.R., Gabrysch, S., Huttly, S.R. (2009), Is the wealth index a proxy for consumption expenditure? A systematic review. Journal of Epidemiology and Community Health, 63, 871-877.

Howe, L.D., Hargreaves, J.R., Huttly, S.R. (2008), Issues in the construction of wealth indices for the measurement of socioeconomic position in low-income countries. Emerging Themes in Epidemiology, 5(1), 3.

Jolliffe, I.T., Cadima, J. (2016), Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A, 374, 20150202.

Lora, E., editor. (2008), Beyond Facts: Understanding Quality of Life. Washington, DC: Inter-American Development Bank.

Lovaton Davila, R. (2015), Testing Alternative Aggregation Methods Using Ordinal Data for a Census Asset-Based Wealth Index. San Diego, CA: Population Association of America.

Lovaton Davila, R., McCarthy, A.S., Gondwe, D., Kirdruang, P., Sharma, U. (2022), Water, walls, and bicycles: Wealth index composition using census microdata. Journal of Demographic Economics, 88, 79-120.

Mariolis, T., Tsoulfidis, L. (2018), Less is more: Capital theory and almost irregular uncontrollable actual economies. Contributions to Political Economy, 37(1), 65-88.

Mathieu, J.P.P., Grépin, K.A., Grignon, M. (2019), Approaches and alternatives to the wealth index to measure socioeconomic status using survey data: A critical interpretive synthesis. Social Indicators Research, 148, 1-46.

Montgomery, M.R., Gragnolati, M., Burke, K.A., Paredes, E. (2000), Measuring living standards with proxy variables. Demography, 37(2), 155-174.

Morris, S., Carletto, C., Hoddinott, J., Christiaensen, L. (2000), Validity of rapid estimates of household wealth and income for health surveys in rural Africa. Journal of Epidemiology and Community Health, 54, 381-387.

Ngo, D., Christiaensen, L. (2018), The Performance of a Consumption Augmented Asset Index in Ranking Households and Identifying the Poor, Policy Research Working Paper. Washington, DC: World Bank.

Pearson, K. (1901), On lines and planes of closest fit to points in space. Philosophical Magazine, 2, 559-572.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L., Guhr, T., Stanley, E. (2002), Random matrix approach to cross-correlations in financial data. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 65(6), 1-18.

Rabe-Hesketh, S., Everitt, B.S. (2007), A Handbook of Statistical Analyses Using Stata. 4th ed. Boca Raton, FL: Chapman Hall/CRC.

Ramos, C., Moreno, B. (2013), Characterization of Spanish economic sectors from an economic and environmental perspective: Evolution and forecast of greenhouse gas emissions. Regional and Sectoral Economic Studies, 13(2), 117-134.

Sekhar, C. C., Indrayan, A., & Gupta, S. M. (1991). International Journal of Epidemiology, 20(1), 246-250.

Tarozzi, A., Deaton, A. (2009), Using census and survey data to estimate poverty and inequality for small areas. Review of Economics and Statistics, 91(4), 773-792.

Tsoulfidis, L. (2022), Capital theory debates: New developments and direction. Investigación Económica, 81(318), 3-36.

Tsoulfidis, L., Athanasiadis, I. (2022), A new method of identifying key industries: A principal component analysis. Journal of Economic Structures, 11, 1-23.

Vom Hofe, R., Bhatta, S.D. (2007), Method for identifying local and domestic industrial clusters using interregional commodity trade data. The Industrial Geographer, 4(2), 1-27.

ANNEXURES

Annexe 1: Sample distribution by province and district

Province	District	Households, n (%)
Central	Chibombo	6073 (2.17)
	Kabwe	4512 (1.62)
	Kapiri Mposhi	5196 (1.86)
	Mkushi	3317 (1.19)
	Mumbwa	4431 (1.59)
	Serenje	3413 (1.22)
	Subtotal	26,942 (9.65)
Copperbelt	Chililabombwe	1902 (0.68)
	Chingola	4582 (1.64)
	Kalulushi	2217 (0.79)
	Kitwe	10,762 (3.85)
	Luanshya	3404 (1.22)
	Lufwanyama	1885 (0.67)
	Masaiti	2300 (0.82)
	Mpongwe	2072 (0.74)
	Mufulira	3376 (1.21)
	Ndola	9696 (3.47)
	Subtotal	42,196 (15.11)
Eastern	Chadiza	2191 (0.78)
	Chipata	9867 (3.53)
	Katete	5146 (1.84)
	Lundazi	6886 (2.47)
	Mambwe	1512 (0.54)
	Nyimba	1817 (0.65)
	Petauke	6541 (2.34)
	Subtotal	33,960 (12.16)
Luapula	Chienge	3142 (1.13)
	Kawambwa	3066 (1.10)
	Mansa	5300 (1.90)
	Milenge	929 (0.33)
	Mwense	3008 (1.08)
	Nchelenge	3765 (1.35)
	Samfya	4944 (1.77)
	Subtotal	24154 (8.65)
Lusaka	Chongwe	4229 (1.51)
	Kafue	5236 (1.87)
	Luangwa	542 (0.19)
	Lusaka	39,180 (14.03)
	Subtotal	49,187 (17.61)
Muchinga	Chama	2237 (0.80)
	Chinsali	3086 (1.11)
	Isoka	1683 (0.60)

(*Contd*...)

Annexe 1: (*Continued*)

Province	District	Households, n (%)
	Mafinga	1536 (0.55)
	Mpika	4238 (1.52)
	Nakonde	2527 (0.90)
	Subtotal	15307 (5.48)
Northern	Chilubi	1850 (0.66)
	Kaputa	2612 (0.94)
	Kasama	4821 (1.73)
	Luwingu	2639 (0.94)
	Mbala	4215 (1.51)
	Mporokoso	2114 (0.76)
	Mpulungu	2402 (0.86)
	Mungwi	3233 (1.16)
	Subtotal	23,886 (8.55)
North Western	Chavuma	699 (0.25)
	Ikelenge	605 (0.22)
	Kabompo	1923 (0.69)
	Kasempa	1447 (0.52)
	Mufumbwe	1064 (0.38)
	Mwinilunga	1899 (0.68)
	Solwezi	4923 (1.76)
	Zambezi	1571 (0.56)
	Subtotal	14,131 (5.06)
Southern	Choma	4602 (1.65)
	Gwembe	1037 (0.37)
	Itezhi Tezhi	1299 (0.47)
	Kalomo	4622 (1.66)
	Kazungula	2080 (0.74)
	Livingstone	3280 (1.17)
	Mazabuka	4517 (1.62)
	Monze	3455 (1.24)
	Namwala	1737 (0.62)
	Siavonga	1883 (0.67)
	Sinazongwe	2094 (0.75)
	Subtotal	30,606 (10.96)
Western	Kalabo	2780 (1.00)
	Kaoma	3758 (1.35)
	Lukulu	1719 (0.62)
	Mongu	3803 (1.36)
	Senanga	2624 (0.94)
	Sesheke	2182 (0.78)
	Shang'ombo	2036 (0.73)
	Subtotal	18,902 (6.77)
Totals		279,271 (100.00)

Annexe 2: List of variables used in principal component analysis

Variable	Categories	Frequency (%)
Assets ownership		
Radio	Yes	58.32
Television	Yes	30.59
Refrigerator	Yes	15.82
Telephone	Yes	1.53
Bicycle	Yes	38.21
Motor vehicle	Yes	4.77
Internet facility	Yes	1.32
Computer/laptop	Yes	3.65
Motorcycle	Yes	0.66
Plough	Yes	9.34
Boat/canoe	Yes	3.81
Scotch cart	Yes	3.31
Donkey	Yes	0.29
Mobile phone	Yes	52.16
Oxen	Yes	7.42
Wheelbarrow	Yes	9.08
Dwelling characteristics		
Housing type	Traditional	67.82
	Conventional house	30.41
	Other	1.78
Roof	Thatch/palm leaf	50.16
	Metal/iron sheets	48.57
	Other	1.27
Walls	Burnt bricks	24.87
	Mud bricks	46.04
	Compressed cement/bricks	26.39
	Other	2.7
Floor	Concrete	43.24
	Mud	55.55
	Other	1.21
Occupancy	Single household	81.25
	One house in several housing units	7.58
	Shared	0.96
	Vacant	4.09
	Noncontact	1.82
	Nonresidential	4.3
Presence of utilities		
Lighting	Electricity	21.94
	Gas	0.13
	Wood	1.66
	Candle	27.69
	Paraffin	20.21
	Solar	2.9
	Bio fuel	0.11
	Diesel	2.15

(*Contd*...)

Annexe 2: (*Continued*)

Variable	Categories	Frequency (%)
	None	1.2
	Other	22
Cooking	Electricity	16.77
	Gas	0.15
	Wood	53.46
	Paraffin	0.07
	Cow dung	0.11
	Charcoal	29.12
	Coal	0.06
	Solar	0.02
	Bio fuel	0.03
	Diesel	0.01
	None	0.04
	Other	0.18
Heating	Electricity	11.67
	Gas	0.07
	Wood	33.73
	Paraffin	0.13
	Cow dung	0.1
	Charcoal	32.82
	Coal	0.09
	Solar	0.22
	Bio fuel	0.02
	Diesel	0.05
	None	18.02
	Other	3.07
Solid waste disposal	Regular collected	7.19
	Irregularly collected	2.39
	Burnt	8.02
	Roadside dumping	6.47
	Another dumping	16.23
	Burying/pit	55.81
	Other	3.88
Toilet type	Flush private connected to water sewer	10.09
	Flush private connected to stand-alone	2.3
	Flush communal	0.65
	Pit latrine	63.88
	VIP	1.83
	Bucket	0.03
	Other	0.55
	No toilet	20.67
Drinking water	Piped water inside the housing unit	60.81
	Unprotected well	38.49
	Other	0.7

VIP: Ventilated improved pit

Annexe 3: Principal component analysis results

Component	Eigenvalue	Difference	Proportion (%)	Cumulative (%)
1	14.770200	9.67910000	14.34	14.34
2	5.091070	0.88890100	4.94	19.28
3	4.202160	1.23869000	4.08	23.36
4	2.963480	0.43004200	2.88	26.24
5	2.533440	0.46021700	2.46	28.70
6	2.073220	0.07148230	2.01	30.71
7	2.001740	0.04938790	1.94	32.66
8	1.952350	0.02002100	1.90	34.55
9	1.932330	0.03383320	1.88	36.43
10	1.898490	0.04526110	1.84	38.27
11	1.853230	0.14373900	1.80	40.07
12	1.709490	0.06077930	1.66	41.73
13	1.648710	0.00292772	1.60	43.33
14	1.645790	0.02974130	1.60	44.93
15	1.616050	0.03251540	1.57	46.50
16	1.583530	0.00437362	1.54	48.03
17	1.579160	0.09186770	1.53	49.57
18	1.487290	0.09314690	1.44	51.01
19	1.394140	0.02242380	1.35	52.36
20	1.371720	0.04374840	1.33	53.70
21	1.327970	0.01944030	1.29	54.99
22	1.308530	0.00527163	1.27	56.26
23	1.303260	0.02661320	1.27	57.52
24	1.276640	0.01731170	1.24	58.76
25	1.259330	0.01360630	1.22	59.98
26	1.245730	0.03620340	1.21	61.19
27	1.209520	0.01085900	1.17	62.37
28	1.198660	0.02972520	1.16	63.53
29	1.168940	0.01830340	1.13	64.67
30	1.150640	0.00643703	1.12	65.78
31	1.144200	0.02136590	1.11	66.89
32	1.122830	0.02055850	1.09	67.98
33	1.102270	0.02339700	1.07	69.05
34	1.078880	0.02286420	1.05	70.10
35	1.056010	0.00105258	1.03	71.13
36	1.054960	0.01337390	1.02	72.15

(*Contd*...)

Annexe 3: (*Continued*)

Component	Eigenvalue	Difference	Proportion (%)	Cumulative (%)
37	1.041590	0.01523070	1.01	73.16
38	1.026360	0.01255970	1.00	74.16
39	1.013800	0.00109317	0.98	75.14
40	1.012700	0.00291177	0.98	76.13
41	1.009790	0.00578588	0.98	77.11
42	1.004010	0.00283569	0.97	78.08
43	1.001170	0.00667394	0.97	79.05
44	0.994496	0.00221179	0.97	80.02
45	0.992284	0.00453178	0.96	80.98
46	0.987752	0.01000800	0.96	81.94
47	0.977744	0.00142807	0.95	82.89
48	0.976316	0.01075330	0.95	83.84
49	0.965563	0.00207575	0.94	84.78
50	0.963487	0.02073590	0.94	85.71
51	0.942751	0.01790330	0.92	86.63
52	0.924848	0.00499825	0.90	87.52
53	0.919850	0.05998640	0.89	88.42
54	0.859863	0.00757509	0.83	89.25
55	0.852288	0.00898608	0.83	90.08
56	0.843302	0.01361750	0.82	90.90
57	0.829685	0.01359090	0.81	91.70
58	0.816094	0.05154320	0.79	92.50
59	0.764551	0.04835220	0.74	93.24
60	0.716198	0.01524640	0.70	93.93
61	0.700952	0.00486658	0.68	94.61
62	0.696085	0.09900980	0.68	95.29
63	0.597076	0.05335290	0.58	95.87
64	0.543723	0.02164710	0.53	96.40
65	0.522076	0.01720650	0.51	96.91
66	0.504869	0.04782380	0.49	97.40
67	0.457045	0.03428960	0.44	97.84
68	0.422756	0.06618020	0.41	98.25
69	0.356576	0.00040274	0.35	98.60
70	0.356173	0.01433210	0.35	98.94
71	0.341841	0.03562800	0.33	99.27
72	0.306213	0.02206310	0.30	99.57
73	0.284150	0.12608400	0.28	99.85
74	0.158065	0.15806500	0.15	100.00

Annexe 4: Variables components loadings

Variable	Loadings for principal components				
	1	2	3	4	Average
Housing (traditional)	−0.2176	0.0445	0.0319	−0.0962	−0.0085
Housing (conventional)	0.2169	−0.0432	−0.0318	0.0926	0.0079
Housing (others)	0.0111	−0.0096	−0.0018	0.0265	0.0039
Roof (thatch or palm leaf)	−0.2046	0.0691	0.1336	−0.0321	−0.0058
Roof (metal or iron sheet)	0.2079	−0.0690	−0.1354	0.0240	0.0044
Roof (others)	−0.0142	−0.0006	0.0077	0.0362	0.0061
Walls (burnt bricks)	−0.0062	0.0430	−0.1193	−0.1312	−0.0036
Walls (mud bricks)	−0.1632	0.0097	0.1004	−0.0240	−0.0097
Walls (compressed cement bricks)	0.2012	−0.0577	−0.0132	0.1164	0.0057
Walls (others)	−0.0280	0.0133	0.0425	0.1053	0.0241
Floor (concrete)	0.2187	−0.0658	−0.1081	0.0304	0.0052
Floor (mud)	−0.2150	0.0648	0.1037	−0.0476	−0.0100
Floor (others)	−0.0134	0.0037	0.0186	0.0788	0.0221
Single household	0.0086	−0.0388	0.0755	−0.0811	−0.0024
One house in several units	−0.0173	0.0466	−0.0716	0.0669	0.0043
Shared	0.0221	−0.0152	−0.0217	0.0505	−0.0047
Drinking (improved source)	0.1500	−0.0652	−0.0864	0.0442	−0.0219
Drinking (unimproved source)	−0.1493	0.0662	0.0856	−0.0460	0.0152
Drinking (others)	−0.0071	−0.0044	0.0061	0.0100	0.0390
Lighting (electricity)	0.2110	0.0707	0.0828	0.0904	0.0000
Lighting (gas)	−0.0039	0.0002	−0.0001	−0.0046	0.0161
Lighting (wood)	−0.0293	0.0107	0.0518	0.0494	0.0154
Lighting (candle)	−0.0014	−0.1729	−0.1264	−0.0198	0.0021
Lighting (paraffin)	−0.0878	0.0082	0.0420	−0.1000	−0.0214
Lighting (solar)	−0.0041	0.0428	−0.0979	−0.0725	0.0102
Lighting (biofuel)	−0.0027	0.0012	−0.0037	−0.0011	−0.0006
Lighting (diesel)	−0.0282	0.0321	−0.0039	0.0197	0.0253
Lighting (none)	−0.0240	0.0117	0.0358	0.0734	0.0117
Lighting (others)	−0.0967	0.0731	0.0289	0.0165	−0.0037
Cooking (electricity)	0.1937	0.1060	0.1164	0.0978	−0.0019
Cooking (gas)	0.0000	0.0115	0.0065	−0.0059	0.0140
Cooking (wood)	−0.1988	0.1489	0.0484	0.0336	−0.0109
Cooking (paraffin)	−0.0032	−0.0008	0.0004	−0.0021	0.0280
Cooking (cow dung)	−0.0024	−0.0011	0.0029	0.0187	0.0310
Cooking (charcoal)	0.0594	−0.2512	−0.1501	−0.1186	0.0053
Cooking (coal)	0.0017	−0.0038	−0.0023	−0.0035	0.0000
Cooking (solar)	0.0014	0.0048	−0.0046	−0.0121	0.0340
Cooking (biofuel)	−0.0017	−0.0001	−0.0002	−0.0015	0.0153
Cooking (diesel)	0.0004	0.0019	0.0012	−0.0006	0.0129
Cooking (none)	−0.0002	−0.0052	0.0036	0.0097	0.0169
Cooking (others)	−0.0013	−0.0007	0.0055	0.0084	0.0072
Heating (electricity)	0.1645	0.1140	0.1226	0.0763	−0.0040
Heating (gas)	−0.0027	0.0050	0.0009	0.0007	0.0129
Heating (wood)	−0.1395	0.1227	0.0478	0.0424	−0.0178
Heating (paraffin)	−0.0035	−0.0027	−0.0015	−0.0029	0.0259
Heating (cow dung)	−0.0015	−0.0011	0.0032	0.0148	0.0298
Heating (charcoal)	0.0360	−0.2000	−0.1366	−0.1533	−0.0030
Heating (coal)	0.0003	−0.0015	−0.0023	−0.0067	0.0000
Heating (solar)	0.0011	0.0015	−0.0056	−0.0176	0.0316
Heating (biofuel)	0.0001	0.0004	0.0003	0.0020	0.0155
Heating (diesel)	0.0015	−0.0024	0.0004	0.0021	0.0141
Heating (none)	−0.0057	−0.0014	0.0041	0.0677	0.0137
Heating (other)	−0.0083	−0.0006	0.0048	0.0116	0.0093
Radio (yes)	0.0956	0.0967	−0.1207	−0.1868	−0.0033
Radio (no)	−0.0956	−0.0967	0.1207	0.1868	0.0033
Television (yes)	0.2006	0.0622	−0.0341	−0.0102	0.0056
Television (no)	−0.2006	−0.0622	0.0341	0.0102	−0.0056
Refrigerator (yes)	0.1913	0.1233	0.1108	0.0328	0.0062
Refrigerator (no)	−0.1913	−0.1233	−0.1108	−0.0328	−0.0062
Telephone (yes)	0.0591	0.1348	0.1137	−0.1062	−0.0050
Telephone (no)	−0.0591	−0.1348	−0.1137	0.1062	0.0050
Bicycle (yes)	−0.0434	0.1381	−0.1374	−0.3090	−0.0057
Bicycle (no)	0.0434	−0.1381	0.1374	0.3090	0.0057
Vehicle (yes)	0.1132	0.1904	0.1319	−0.0886	0.0294

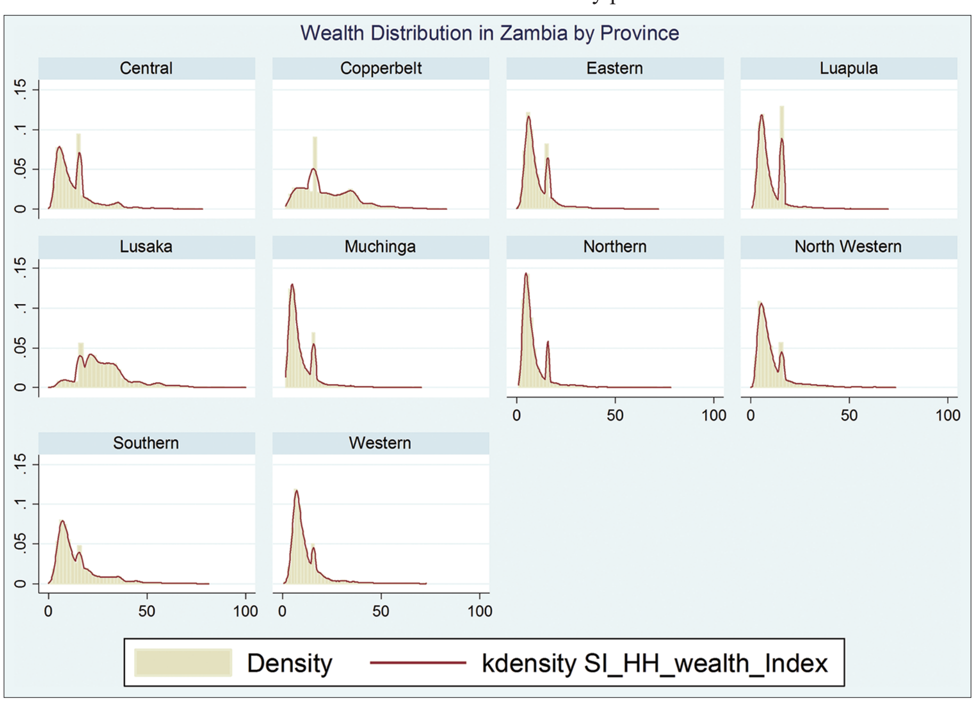
(*Contd...*)

Annexe 4: (*Continued*)

Variable	Loadings for principal components				
	1	2	3	4	Average
Vehicle (no)	−0.1132	−0.1904	−0.1319	0.0886	−0.0294
Internet (yes)	0.0700	0.1532	0.1412	−0.1175	0.0071
Internet (no)	−0.0700	−0.1532	−0.1412	0.1175	−0.0071
Computer (yes)	0.1128	0.1856	0.1764	−0.0890	0.0178
Computer (no)	−0.1128	−0.1856	−0.1764	0.0890	−0.0178
Motorcycle (yes)	0.0222	0.0759	−0.0219	−0.0952	−0.0102
Motorcycle (no)	−0.0222	−0.0759	0.0219	0.0952	0.0102
Plough (yes)	−0.0537	0.2450	−0.2680	0.1281	−0.0022
Plough (no)	0.0537	−0.2450	0.2680	−0.1281	0.0022
Boat/canoe (yes)	−0.0418	0.0517	0.0234	−0.0092	0.0035
Boat/canoe (no)	0.0418	−0.0517	−0.0234	0.0092	−0.0035
Scotch cart (yes)	−0.0273	0.2030	−0.2605	0.0949	0.0055
Scotch cart (no)	0.0273	−0.2030	0.2605	−0.0949	−0.0055
Donkey (yes)	−0.0065	0.0690	−0.0675	0.0525	0.0015
Donkey (no)	0.0065	−0.0690	0.0675	−0.0525	−0.0015
Mobile phone (yes)	0.1611	0.0069	−0.1576	−0.0924	0.0035
Mobile phone (no)	−0.1611	−0.0069	0.1576	0.0924	−0.0035
Oxen (yes)	−0.0476	0.2268	−0.2506	0.1518	−0.0049
Oxen (no)	0.0476	−0.2268	0.2506	−0.1518	0.0049
Wheelbarrow (yes)	0.0823	0.0985	−0.0491	−0.1731	0.0061
Wheelbarrow (no)	−0.0823	−0.0985	0.0491	0.1731	−0.0061
Solid waste (regularly collected)	0.0944	0.0327	0.0577	0.0785	0.0113
Solid waste (irregular collected)	0.0256	−0.0162	−0.0013	0.0585	0.0086
Solid waste (burnt)	−0.0106	−0.0001	−0.0097	0.0308	−0.0154
Solid waste (roadside dumping)	0.0040	−0.0272	−0.0186	0.0801	0.0126
Solid waste (another dumping)	−0.0538	0.0336	0.0142	0.1401	−0.0042
Solid waste (burying pit)	0.0009	−0.0338	−0.0376	−0.2508	−0.0239
Solid waste (other)	−0.0365	0.0268	0.0306	0.0808	0.0531
Toilet type (flush connected)	0.1464	0.0924	0.1099	0.0692	0.0051
Toilet type (flush not connected)	0.0677	0.0572	0.0487	−0.0051	0.0071
Toilet type (flush communal)	0.0248	−0.0121	−0.0013	0.0431	0.0076
Toilet type (pit latrine)	−0.0376	−0.1506	−0.1336	−0.2609	−0.0193
Toilet type (VIP pit latrine)	0.0214	−0.0016	−0.0271	0.0075	−0.0115
Toilet type (bucket)	0.0022	−0.0013	−0.0002	0.0106	0.0164
Toilet type (others)	0.0022	−0.0013	−0.0002	0.0106	0.0164
Toilet type (no toilet)	−0.0993	0.0899	0.0662	0.2433	0.0122
House owner (no)	−0.1451	0.1282	0.0320	−0.1248	0.0104
House owner (yes)	0.1451	−0.1282	−0.0320	0.1248	−0.0104

VIP: Ventilated improved pit

Annexe 5: Wealth distribution by province



Annexe 6: District-level wealth index and poverty headcount estimate

District	Mean WI score (%)	District rank based on WI	Poverty head count	District rank based on PHC
Lusaka	29.7	1	0.18	1
Livingston	28.49	2	0.28	2
Kitwe	25.77	3	0.29	4
Chililabombwe	25.64	4	0.3	7
Mufulira	25.18	5	0.3	5
Chingola	24.79	6	0.32	9
Ndola	24.4	7	0.31	8
Luanshya	23.51	8	0.33	10
Kabwe	22.19	9	0.33	11
Kafue	22.07	10	0.4	12
Kalulushi	21.42	11	0.3	6
Mazabuka	16.95	12	0.63	18
Mongu	15.7	13	0.71	26
Chongwe	15.49	14	0.61	17
Siavonga	14.41	15	0.72	29
Solwezi	13.9	16	0.5	13
Monze	13.29	17	0.75	36
Choma	12.91	18	0.72	32
Sinazongwe	12.83	19	0.77	38
Luangwa	12.7	20	0.7	23
Chipata	12.43	21	0.72	33
Chibombo	11.79	22	0.73	34
Kasama	11.67	23	0.51	15
Itezhi tezhi	11.65	24	0.7	24
Kapiri Mposhi	11.6	25	0.68	22
Sesheke	11.59	26	0.85	59
Mansa	11.44	27	0.65	20
Mkushi	11.24	28	0.71	27
Mumbwa	11.24	29	0.64	19
Senanga	11.23	30	0.87	65
Nakonde	11.02	31	0.72	31
Mpongwe	10.93	32	0.71	25
Namwala	10.87	33	0.72	30
Mambwe	10.78	34	0.81	46
Gwembe	10.47	35	0.82	56
Kalomo	10.46	36	0.75	37

(*Contd*...)

Annexe 6: (*Continued*)

District	Mean WI score (%)	District rank based on WI	Poverty head count	District rank based on PHC
Nyimba	10.15	37	0.78	41
Katete	9.93	38	0.82	54
Nchelenge	9.91	39	0.77	39
Mpulungu	9.89	40	0.81	45
Masaiti	9.85	41	0.51	14
Kalabo	9.8	42	0.88	70
Isoka	9.77	43	0.81	48
Kasempa	9.76	44	0.81	47
Mpika	9.76	45	0.74	35
Mwense	9.74	46	0.79	42
Kawambwa	9.7	47	0.82	53
Shang'ombo	9.64	48	0.95	74
Kazungula	9.58	49	0.68	21
Petauke	9.5	50	0.82	50
Mufumbwe	9.48	51	0.87	66
Lufwanyama	9.35	52	0.8	44
Zambezi	9.33	53	0.87	64
Kaoma	9.26	54	0.82	55
Chadiza	9.21	55	0.81	49
Lukulu	9.16	56	0.86	63
Samfya	9.14	57	0.91	72
Kabompo	9.08	58	0.9	71
Serenje	9.01	59	0.78	40
Lundazi	8.9	60	0.84	58
Chienge	8.88	61	0.82	57
Chilubi	8.69	62	0.87	67
Chinsali	8.63	63	0.85	60
Mwinilunga	8.39	64	0.29	3
Chavuma	8.31	65	0.87	68
Chama	8.29	66	0.71	28
Milenge	8.25	67	0.88	69
Mbala	8.01	68	0.82	52
Kaputa	7.89	69	0.79	43
Luwingu	7.58	70	0.86	62
Mafinga	7.52	71	0.91	73
Mporokoso	7.38	72	0.82	51
Ikelenge	7.29	73	0.59	16
Mungwi	6.97	74	0.86	61

WI: Wealth index, PHC: Primary health centre