

Supplementary Appendix

Table of Contents

<i>Supplementary Appendix</i>	<i>1</i>
1. Methods	2
1.1 Classical-PCA Wealth Index.....	2
1.2 Sparse-PCA Wealth Index	2
2. Sensitive analysis of threshold selection in feature-selection-PCA	2
3. Internal coherence of wealth indices for feature-selection-PCA, sparse-PCA, and robust-PCA .	4
4. Cross-method agreement of wealth indices	11
5. External validation of the wealth indices – Consistency with household income classification .	11
6. External validation of the wealth indices – Analysis for household with zero income	12
7. Reference	13

1. Methods

1.1 Classical-PCA Wealth Index

We followed the steps used in constructing the DHS Wealth Index¹ to calculate the classical-PCA wealth index. The principal component analysis (PCA) is carried out on the correlation matrix of the standardized asset ownership binary data. PCA creates new orthogonal components by maximizing the variance of the data and the first principal component accounts for the largest possible variance across the variables.² The weights of each household asset are drawn from the PCA weights in the first principal component, assuming that wealth is the component that accounts for the greatest amount of variation in the assets and characteristics of households. Therefore, the wealth index for each household is a linear combination of all assets with PCA weights as corresponding coefficients. The formula is shown below:

$$WI_i = a_1 x_{i1}^* + a_2 x_{i2}^* + a_3 x_{i3}^* + \dots + a_p x_{ip}^*$$

where $a_k, k = 1, \dots, p$ are PCA weights for p assets, x_{ik}^* are the standardized values of asset k of household i . If asset k represents a numerical variable such as the number of constructions, $x_{ik}^* = (x_{ik} - \bar{x}_k)/s_k$; If asset k represents a dummy variable where 0 means not owning and 1 means owning the asset,

$$x_{ik}^* = \begin{cases} \frac{1 - \bar{x}_k}{s_k}, & \text{if household } i \text{ owns asset } k, \\ \frac{0 - \bar{x}_k}{s_k}, & \text{if household } i \text{ does not own asset } k, \end{cases}$$

where \bar{x}_k and s_k are the mean and standard deviation of asset k for all households.

1.2 Sparse-PCA Wealth Index

The lasso (elastic net)³ is a penalized least squares approach that imposes additional constraints on the regression coefficients. Hence, the lasso (elastic net) estimated $\hat{\beta}$ is obtained by minimizing

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p |\beta_j|^2$$

where λ_1, λ_2 are non-negative values. The lasso is a special case of the elastic net with $\lambda_2 = 0$.

Zou and Hastie (2006)⁴ proposed a “self-contained” regression-type criterion for driving PCs by including the lasso (elastic net) penalty into the PCA objective function in order to produce sparse weights. Let α and β be $p \times k$ matrices, X_i denote the i -th row vector of the matrix $X_{n \times p}$. For any $\lambda > 0$ and $\lambda_{1,j} \geq 0$, let

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{\alpha, \beta} \sum_{i=1}^n |X_i - \alpha \beta^T X_i|^2 + \lambda \sum_{j=1}^k |\beta_j|^2 + \sum_{j=1}^k \lambda_{1,j} |\beta_j|, \quad \text{subject to } \alpha^T \alpha = I_k.$$

Then each column of $\hat{\beta}$ is proportional to the sparse PCA weights of each principal component. The PCA weights of the first PC is obtained by normalization of $\hat{\beta}_1$, i.e., $\hat{A} = c(\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p) = \frac{\hat{\beta}_1}{|\hat{\beta}_1|}$.

2. Sensitive analysis of threshold selection in feature-selection-PCA

To validate the selection of the threshold for eliminating assets with small weights in feature-selection-PCA method, a sensitivity analysis is conducted. The analysis involved examining the Spearman's rank correlation between feature-selection-PCA and classical-PCA under varying thresholds as illustrated in Figure 1. The results indicated that the correlation of wealth quintiles exhibited a gradual decline across various thresholds, while Spearman's correlation remained above 0.9 when more than 30% of asset indicators were retained in the PCA analysis. Conversely, when less than 30% of indicators were retained, the correlation tended to decrease significantly and was less than 0.9.

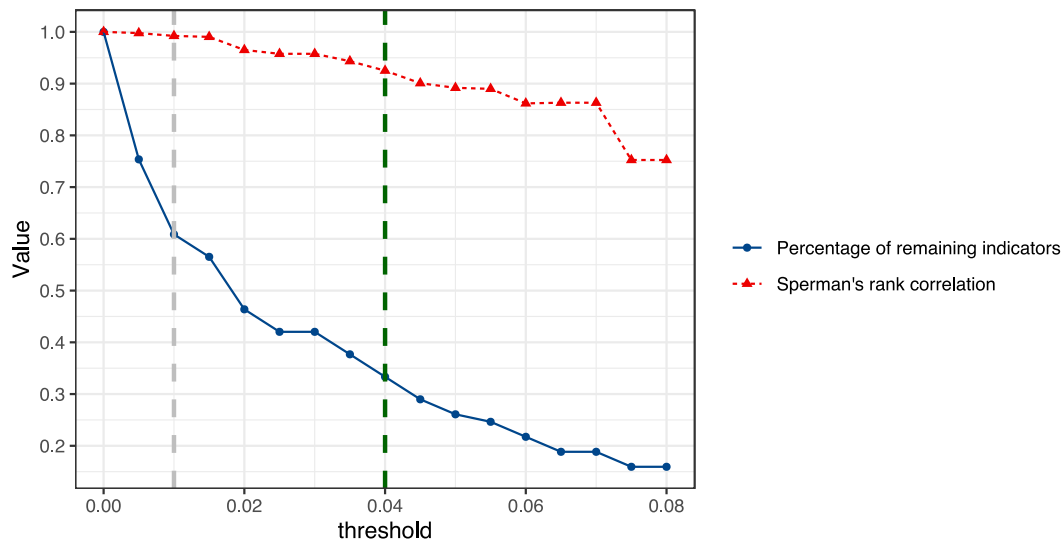


Figure S1. Sensitive analysis of spearman's rank correlation between feature-selection-PCA and classical-PCA under varying thresholds

In addition, the average classification accuracy, sensitivity, and precision are calculated for five wealth quintiles with varied thresholds (Figure 2). The results demonstrate that the average accuracy exhibits a slight decline, whereas the average sensitivity and precision exhibit a notable drop. Note that when the threshold is set above 0.04 which preserves over 30% of indications for study, the feature-selection-PCA shows a high level of performance in terms of average accuracy (greater than 0.9).

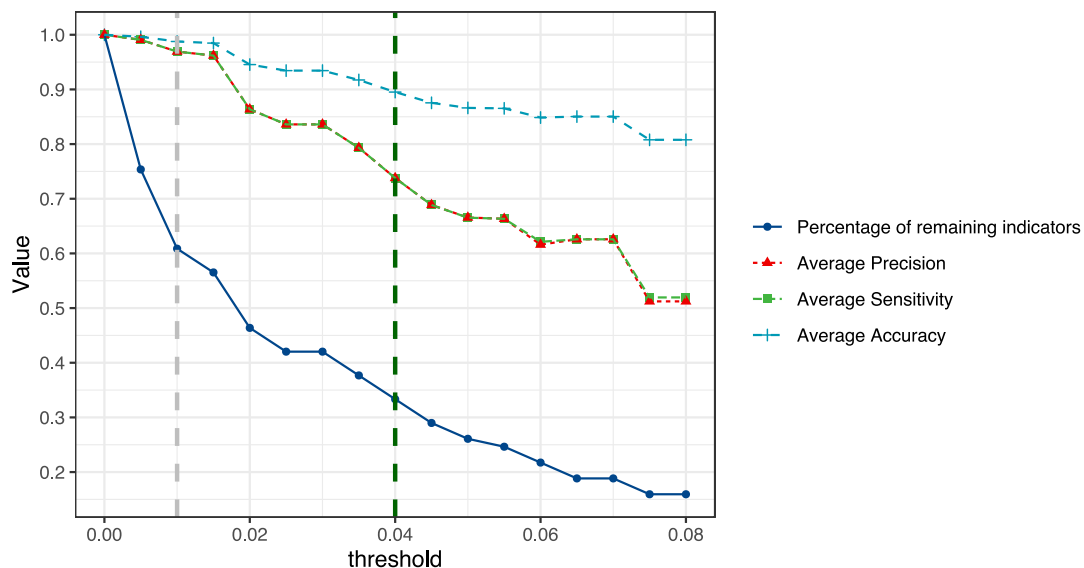


Figure S2. Sensitivity analysis of feature-Selection-PCA classification results compared to classical-PCA: average accuracy, average sensitivity, and average precision.

The findings of the sensitivity analysis suggest that the choice of the threshold may have a notable impact on the correlation and accuracy of wealth quintiles, particularly when the percentage of retained indicators is less than

30%. It is evident that there exists a trade-off between the removal of indicators and the capability of the feature-selection-PCA wealth index. Therefore, researchers should carefully consider the threshold selection to ensure a sufficient number of indicators are retained to maintain the robustness of the wealth quintiles.

3. Internal coherence of wealth indices for feature-selection-PCA, sparse-PCA, and robust-PCA

Table S1. Percentage (or mean number) of households owning each asset indicator across wealth quintiles, and the regression results between each asset indicator and feature-selection-PCA rank.

	Mean (SE) or % N=25550	Wealth Quintiles					Regression ¹
		Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	β
Number of constructions (++++ ²)	1.565 (0.890)	1.320 (0.636)	1.405 (0.736)	1.551 (0.866)	1.721 (1.015)	1.829 (1.027)	0.133***
Number of members per sleeping room (++++)	4.030 (2.013)	3.350 (1.721)	3.767 (1.795)	4.115 (1.924)	4.372 (2.094)	4.585 (2.243)	0.308***
Number of bed nets (++++)	2.149 (1.288)	1.304 (0.954)	1.808 (1.064)	2.196 (1.108)	2.595 (1.269)	2.844 (1.377)	0.387***
Ownership of cattle or pigs (--- ³)	7.86%	11.90%	7.80%	7.20%	7.00%	5.40%	-0.194***
Number of cows (> 1 year of age)							
1-4	0.20%	0.10%	0.20%	0.20%	0.20%	0.30%	0.159
5-9	0.06%	0.00%	0.00%	0.00%	0.10%	0.10%	0.432*
10 or more	0.03%	0.00%	0.00%	0.00%	0.00%	0.10%	0.566
Number of cows (< 1 year of age)							
1-4	0.27%	0.30%	0.30%	0.20%	0.30%	0.30%	0.000
5-9	0.07%	0.10%	0.00%	0.10%	0.10%	0.10%	-0.106
10 or more	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.774
Number of pigs (> 6 weeks of age)							
1-4 (---)	4.90%	8.30%	4.90%	4.30%	4.10%	2.90%	-0.253***
5-9	0.97%	1.10%	1.00%	1.00%	1.00%	0.80%	-0.043
10 or more	0.28%	0.30%	0.20%	0.30%	0.30%	0.30%	0.097
Number of pigs (< 6 weeks of age)							
1-4 (--)	2.77%	4.60%	2.70%	2.60%	2.20%	1.90%	-0.226***
5-9	0.92%	1.00%	1.00%	0.80%	0.90%	0.80%	-0.065

¹ The β 's are the slope coefficients of the logistic regression for all binary indicators, or the slope coefficients of the linear regression for Number of constructions, Number of members per sleeping room, Number of nets. ***: $P < 0.001$; **: $P < 0.01$; *: $P < 0.05$; :: $P < 0.1$

² ++++: Positive PCA weight larger than 0.01; +++: Positive PCA weight smaller than 0.01 and larger than 0.005; ++: Positive PCA weight smaller than 0.005 and larger than 0.001; +: Positive PCA weight smaller than 0.001

³ ----: Negative PCA weight less than -0.01; ---: Negative PCA weight larger than -0.01 and smaller than -0.005; --: Negative PCA weight larger than -0.005 and smaller than -0.001; -: Negative PCA weight larger than -0.001

10 or more	0.20%	0.30%	0.20%	0.10%	0.20%	0.20%	-0.097
Main housing building type							
Apartment	0.19%	0.00%	0.10%	0.00%	0.40%	0.40%	0.664***
Conventional house (++++)	13.17%	0.00%	0.00%	0.00%	0.60%	65.20%	5.734***
Flat	0.09%	0.10%	0.10%	0.10%	0.10%	0.10%	0.000
Hut (----)	29.01%	41.10%	38.10%	33.80%	26.10%	5.90%	-0.422***
Precarious (---)	19.53%	13.80%	19.50%	23.50%	28.10%	12.70%	0.041***
Traditional mud house (----)	36.91%	44.80%	41.70%	41.70%	42.50%	13.90%	-0.268***
Other (++)	1.11%	0.20%	0.50%	0.90%	2.10%	1.80%	0.486***
Wall material in the main house							
Adobe block (----)	54.55%	67.60%	67.50%	65.70%	54.90%	17.20%	-0.488***
Bamboo (----)	26.02%	52.30%	31.30%	23.50%	18.90%	4.00%	-0.635***
Bark (----)	4.17%	11.30%	4.40%	2.80%	1.70%	0.60%	-0.724***
Brick block (++++)	15.67%	0.00%	0.00%	0.90%	13.10%	64.30%	2.545***
Cardboard	0.09%	0.10%	0.10%	0.10%	0.00%	0.10%	-0.065
Cement blocks (++++)	3.55%	0.00%	0.00%	0.00%	0.30%	17.50%	4.369***
Palm tree (----)	9.38%	20.10%	10.00%	8.00%	6.60%	2.20%	-0.504***
Paper	0.12%	0.20%	0.10%	0.10%	0.10%	0.10%	-0.147
Plastic bags (---)	1.62%	5.60%	1.20%	0.50%	0.50%	0.40%	-0.929***
Reed (---)	3.80%	8.40%	3.90%	3.30%	2.90%	0.50%	-0.503***
Tin	4.74%	2.10%	3.40%	5.70%	8.60%	3.90%	0.199***
Tinned wood	4.63%	2.00%	3.40%	5.60%	8.40%	3.70%	0.197***
Wood (+++)	9.77%	4.00%	7.10%	10.80%	14.80%	12.20%	0.281***
Zinc (++++)	12.13%	0.00%	0.10%	0.20%	7.30%	53.00%	2.736***
Other (----)	26.05%	40.20%	32.90%	28.30%	22.40%	6.30%	-0.430***
Main water sources for consumption							
Fountain (+++)	10.74%	4.00%	6.70%	10.70%	17.00%	15.20%	-0.356***
Hole protected with hand pump outside backyard (+++)	50.58%	33.20%	49.00%	54.20%	57.10%	59.50%	0.247***
Hole with manual pump inside house	0.49%	0.20%	0.40%	0.60%	0.50%	0.80%	0.258***
Mineral bottled water	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	15.185
Piped water in neighbor house (++++)	1.39%	0.00%	0.00%	0.20%	1.40%	5.30%	1.580***
Piped water inside house (++++)	0.38%	0.00%	0.00%	0.00%	0.00%	1.90%	4.614***
Piped water within compound (++++)	1.48%	0.00%	0.00%	0.00%	0.10%	7.30%	4.633***
Protected inside backyard	0.46%	0.20%	0.10%	0.50%	0.70%	0.70%	0.338***
Protected outside backyard (--)	4.81%	7.00%	5.70%	5.10%	4.30%	2.00%	-0.255***
Rainwater	0.08%	0.00%	0.10%	0.20%	0.10%	0.00%	-0.152
Surface (River, Lake, Lagoon) (----)	11.04%	21.60%	14.10%	10.70%	7.30%	1.50%	-0.526***
Unprotected inside household (--)	1.78%	2.30%	2.20%	2.20%	1.70%	0.50%	-0.233***
Unprotected outside household (----)	16.56%	31.40%	21.50%	15.40%	9.60%	4.90%	-0.512***

Water from tank truck	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.000
Other	0.05%	0.00%	0.00%	0.10%	0.10%	0.00%	0.00%	0.084
Time for taking main water sources								
Under 10 min (++++)	31.84%	6.10%	18.60%	37.00%	47.20%	50.20%	0.594***	
Between 10-30 min (----)	45.63%	61.50%	54.10%	40.90%	35.30%	36.40%	-0.285***	
Between 30-60 min (----)	17.75%	25.10%	21.40%	17.30%	14.20%	10.60%	-0.254***	
More than one hour (---)	4.76%	7.30%	5.80%	4.80%	3.30%	2.70%	-0.267***	
Main energy source for lighting								
Batteries (----)	68.79%	84.60%	81.30%	79.00%	62.90%	36.30%	-0.591***	
Candles	0.78%	0.00%	0.40%	1.10%	2.10%	0.30%	0.294***	
Electricity (++++)	11.65%	0.00%	0.00%	0.00%	7.10%	51.10%	2.774***	
Firewood (----)	11.96%	15.00%	16.30%	15.90%	10.80%	1.90%	-0.311***	
Gas	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.169	
Generator	0.04%	0.00%	0.00%	0.00%	0.10%	0.10%	1.375**	
Oil	0.15%	0.00%	0.10%	0.10%	0.50%	0.10%	0.409***	
Solar panel (++++)	4.66%	0.10%	0.40%	1.90%	12.10%	8.80%	0.814***	
Other	1.55%	0.30%	1.40%	1.80%	3.30%	1.00%	0.223***	
Ownership of radio (++++)	24.66%	8.00%	13.80%	22.80%	37.60%	41.10%	0.525***	
Ownership of television (++++)	7.47%	0.00%	0.00%	0.00%	0.50%	36.80%	4.455***	
Ownership of cell phone (++++)	38.47%	8.60%	20.80%	37.70%	58.30%	66.90%	0.752***	

Table S2. Percentage (or mean number) of households owning each asset indicator across wealth quintiles, and the regression results between each asset indicator and sparse-PCA rank.

	Mean (SE) or % N=25550	Wealth Quintiles					Regression
		Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	β
Number of constructions (++++)	1.565 (0.890)	1.320 (0.644)	1.409 (0.742)	1.545 (0.858)	1.723 (1.014)	1.828 (1.026)	0.133***
Number of members per sleeping room (++++)	4.030 (2.013)	3.352 (1.724)	3.782 (1.800)	4.111 (1.936)	4.365 (2.084)	4.578 (2.240)	0.304***
Number of bed nets (++++)	2.149 (1.288)	1.268 (0.941)	1.830 (1.045)	2.189 (1.111)	2.621 (1.268)	2.839 (1.376)	0.393***
Ownership of cattle or pigs (---)	7.86%	10.40%	7.90%	7.50%	8.00%	5.40%	-0.136***
Number of cows (> 1 year of age)							
1-4	0.20%	0.10%	0.20%	0.20%	0.20%	0.30%	0.200*
5-9	0.06%	0.00%	0.00%	0.10%	0.10%	0.10%	0.474*
10 or more	0.03%	0.00%	0.00%	0.00%	0.00%	0.10%	0.566
Number of cows (< 1 year of age)							
1-4	0.27%	0.30%	0.30%	0.20%	0.30%	0.30%	0.022

5-9	0.07%	0.10%	0.00%	0.10%	0.10%	0.10%	-0.079
10 or more	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.774
Number of pigs (> 6 weeks of age)							
1-4 (--)	4.90%	7.20%	5.00%	4.50%	4.80%	3.00%	-0.190***
5-9	0.97%	0.90%	1.00%	1.00%	1.10%	0.80%	-0.014
10 or more	0.28%	0.20%	0.30%	0.30%	0.30%	0.30%	0.092
Number of pigs (< 6 weeks of age)							
1-4 (--)	2.77%	3.80%	2.70%	2.70%	2.80%	1.90%	-0.135***
5-9	0.92%	1.00%	0.80%	1.00%	0.90%	0.80%	-0.028
10 or more	0.20%	0.20%	0.30%	0.10%	0.20%	0.20%	-0.039
Main housing building type							
Apartment (++)	0.19%	0.00%	0.10%	0.10%	0.40%	0.40%	0.699***
Conventional house (++++)	13.17%	0.00%	0.00%	0.00%	0.50%	65.30%	5.955***
Flat	0.09%	0.10%	0.10%	0.10%	0.10%	0.10%	0.131
Hut (--)	29.01%	39.90%	38.00%	34.20%	26.90%	6.00%	-0.401***
Precarious (++++)	19.53%	15.00%	20.40%	22.40%	27.40%	12.50%	0.013
Traditional mud house	36.91%	44.60%	40.90%	42.30%	42.70%	13.90%	-0.261***
Other (+++)	1.11%	0.30%	0.50%	1.00%	2.00%	1.80%	0.416***
Wall material in the main house							
Adobe block (----)	54.55%	68.60%	66.90%	66.00%	54.10%	17.30%	-0.498***
Bamboo (----)	26.02%	52.20%	32.00%	23.10%	18.80%	4.00%	-0.640***
Bark (----)	4.17%	10.90%	4.40%	3.00%	1.90%	0.60%	-0.679***
Brick block (++++)	15.67%	0.00%	0.00%	0.70%	13.30%	64.30%	2.566***
Cardboard	0.09%	0.10%	0.10%	0.10%	0.00%	0.10%	0.000
Cement blocks (++++)	3.55%	0.00%	0.00%	0.00%	0.30%	17.50%	4.369***
Palm tree (----)	9.38%	19.50%	10.00%	8.40%	6.90%	2.20%	-0.482***
Paper	0.12%	0.20%	0.10%	0.10%	0.10%	0.10%	-0.065
Plastic bags (---)	1.62%	5.10%	1.40%	0.70%	0.50%	0.40%	-0.813***
Reed (---)	3.80%	7.70%	4.00%	3.60%	3.20%	0.50%	-0.443***
Tin (++)	4.74%	1.40%	3.10%	5.80%	9.30%	4.00%	0.259***
Tinned wood	4.63%	1.40%	3.00%	5.70%	9.20%	3.80%	0.258***
Wood (+++)	9.77%	3.50%	6.50%	11.10%	15.50%	12.20%	0.311***
Zinc (++++)	12.13%	0.00%	0.00%	0.20%	7.40%	53.10%	2.754***
Other (----)	26.05%	40.00%	33.70%	27.50%	22.50%	6.40%	-0.430***
Main water sources for consumption							
Fountain (++)	10.74%	0.00%	0.10%	0.20%	0.10%	0.00%	0.330***
Hole protected with hand pump outside backyard	50.58%	0.10%	0.40%	0.60%	0.50%	0.80%	0.222***
Hole with manual pump inside house	0.49%	31.60%	21.00%	15.20%	10.00%	4.90%	0.303***
Mineral bottled water	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	15.185
Piped water in neighbor house (++++)	1.39%	0.00%	0.00%	0.20%	1.40%	5.30%	1.530***

Piped water inside house (++++)	0.38%	0.00%	0.00%	0.00%	0.00%	1.90%	4.614***
Piped water within compound (++++)	1.48%	0.00%	0.00%	0.00%	0.10%	7.30%	4.633***
Protected inside backyard	0.46%	4.30%	7.40%	10.50%	16.30%	15.10%	-0.388***
Protected outside backyard (--)	4.81%	0.20%	0.10%	0.60%	0.70%	0.70%	-0.211***
Rainwater	0.08%	20.10%	14.60%	10.80%	8.10%	1.60%	-0.050
Surface (River, Lake, Lagoon) (----)	11.04%	35.60%	48.60%	53.50%	55.70%	59.50%	-0.480***
Unprotected inside household (--)	1.78%	6.30%	5.40%	5.80%	4.50%	2.00%	-0.142***
Unprotected outside household (----)	16.56%	1.70%	2.20%	2.30%	2.10%	0.50%	0.506***
Water from tank truck	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.000
Other	0.05%	0.00%	0.00%	0.10%	0.10%	0.00%	0.169
Time for taking main water sources							
Under 10 min (++++)	31.84%	5.40%	17.70%	37.70%	48.30%	50.10%	0.613***
Between 10-30 min	45.63%	62.20%	54.30%	40.60%	34.60%	36.50%	-0.294***
Between 30-60 min	17.75%	25.30%	21.80%	17.20%	13.80%	10.70%	-0.260***
More than one hour	4.76%	7.10%	6.20%	4.50%	3.30%	2.60%	-0.270***
Main energy source for lighting							
Batteries (----)	68.79%	85.30%	81.30%	77.90%	63.00%	36.30%	-0.598***
Candles	0.78%	0.20%	0.50%	1.20%	1.80%	0.20%	0.190***
Electricity (++++)	11.65%	0.00%	0.00%	0.00%	6.90%	51.30%	2.812***
Firewood (--)	11.96%	13.60%	16.10%	16.60%	11.60%	2.00%	-0.272***
Gas	0.01%	0.00%	0.00%	0.10%	0.00%	0.00%	0.000
Generator (+)	0.04%	0.00%	0.00%	0.00%	0.00%	0.10%	1.702**
Oil	0.15%	0.00%	0.10%	0.10%	0.40%	0.10%	0.346**
Solar panel (++++)	4.66%	0.10%	0.40%	1.90%	12.30%	8.60%	0.814***
Other	1.55%	0.80%	1.40%	1.90%	2.80%	0.90%	0.109**
Ownership of radio (++++)	24.66%	7.50%	14.50%	22.30%	38.00%	41.00%	0.530***
Ownership of television (++++)	7.47%	0.00%	0.00%	0.00%	0.50%	36.80%	4.645***
Ownership of cell phone (++++)	38.47%	7.50%	20.70%	38.40%	59.00%	66.70%	0.770***

Table S3. Percentage (or mean number) of households owning each asset indicator across wealth quintiles, and the regression results between each asset indicator and robust-PCA rank.

	Mean (SE) or % N=25550	Wealth Quintiles					Regression
		Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	β
Number of constructions (++++)	1.565 (0.890)	1.326 (0.645)	1.407 (0.740)	1.554 (0.873)	1.711 (1.005)	1.827 (1.026)	0.131***
Number of members per sleeping room (++++)	4.030 (2.013)	3.343 (1.714)	3.771 (1.797)	4.109 (1.921)	4.38 (2.096)	4.586 (2.243)	0.309***

Number of bed nets (++++)	2.149 (1.288)	1.307 (0.958)	1.812 (1.058)	2.201 (1.122)	2.586 (1.268)	2.842 (1.375)	0.384***
Ownership of cattle or pigs (---)	7.86%	12.80%	7.50%	7.00%	6.70%	5.30%	-0.223***
Number of cows (> 1 year of age)							
1-4	0.20%	0.10%	0.20%	0.20%	0.20%	0.30%	0.179
5-9	0.06%	0.00%	0.00%	0.00%	0.10%	0.10%	0.351
10 or more	0.03%	0.00%	0.00%	0.00%	0.00%	0.10%	0.566
Number of cows (< 1 year of age)							
1-4	0.27%	0.30%	0.30%	0.20%	0.30%	0.30%	-0.007
5-9	0.07%	0.10%	0.00%	0.10%	0.10%	0.10%	-0.106
10 or more	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.774
Number of pigs (> 6 weeks of age)							
1-4 (---)	4.90%	8.60%	4.90%	4.00%	4.00%	2.90%	-0.277***
5-9 (--)	0.97%	1.40%	0.80%	1.00%	0.80%	0.80%	-0.114*
10 or more	0.28%	0.30%	0.20%	0.30%	0.30%	0.30%	-0.097
Number of pigs (< 6 weeks of age)							
1-4 (--)	2.77%	4.90%	2.50%	2.50%	2.10%	1.80%	-0.249***
5-9 (--)	0.92%	1.40%	0.90%	0.70%	0.70%	0.80%	-0.144**
10 or more	0.20%	0.30%	0.20%	0.10%	0.20%	0.20%	-0.097
Main housing building type							
Apartment	0.19%	0.00%	0.10%	0.00%	0.40%	0.40%	0.664***
Conventional house (++++)	13.17%	0.00%	0.00%	0.00%	0.60%	65.20%	5.702***
Flat	0.09%	0.10%	0.10%	0.10%	0.10%	0.10%	0.000
Hut (----)	29.01%	40.40%	38.50%	33.70%	26.40%	5.90%	-0.414***
Precarious (---)	19.53%	13.70%	19.40%	23.50%	28.30%	12.70%	0.045***
Traditional mud house (----)	36.91%	45.70%	41.30%	41.80%	42.00%	13.80%	-0.277***
Other (++)	1.11%	0.20%	0.50%	0.90%	2.10%	1.80%	0.486***
Wall material in the main house							
Adobe block (----)	54.55%	68.00%	67.80%	65.60%	54.30%	17.10%	-0.497***
Bamboo (----)	26.02%	52.00%	31.80%	23.60%	18.80%	4.00%	-0.635***
Bark (----)	4.17%	11.30%	4.30%	2.90%	1.70%	0.60%	-0.717***
Brick block (++++)	15.67%	0.00%	0.00%	0.90%	13.20%	64.20%	2.543***
Cardboard	0.09%	0.10%	0.10%	0.10%	0.00%	0.10%	0.022
Cement blocks (++++)	3.55%	0.00%	0.00%	0.00%	0.30%	17.50%	4.369***
Palm tree (----)	9.38%	20.10%	9.80%	8.30%	6.70%	2.20%	-0.500***
Paper	0.12%	0.20%	0.10%	0.10%	0.10%	0.10%	-0.097
Plastic bags (---)	1.62%	5.50%	1.20%	0.60%	0.50%	0.40%	-0.890***
Reed (---)	3.80%	8.20%	3.90%	3.40%	2.90%	0.50%	-0.488***
Tin (++)	4.74%	1.40%	2.90%	5.40%	9.80%	4.10%	0.284***
Tinned wood (++)	4.63%	1.30%	2.80%	5.40%	9.70%	3.90%	0.283***

Wood (+++)	9.77%	3.30%	6.60%	10.60%	16.00%	12.40%	0.326***
Zinc (++++)	12.13%	0.00%	0.10%	0.20%	7.30%	53.10%	2.736***
Other (----)	26.05%	40.50%	32.90%	28.40%	22.10%	6.30%	-0.435***
Main water sources for consumption							
Fountain (+++)	10.74%	4.10%	6.80%	10.70%	17.00%	15.10%	0.353***
Hole protected with hand pump outside backyard (+++)	50.58%	33.50%	48.80%	54.40%	56.80%	59.40%	0.244***
Hole with manual pump inside house (++)	0.49%	0.10%	0.30%	0.60%	0.70%	0.80%	0.383***
Mineral bottled water	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	15.185
Piped water in neighbor house (++++)	1.39%	0.00%	0.00%	0.20%	1.40%	5.40%	1.589***
Piped water inside house (++++)	0.38%	0.00%	0.00%	0.00%	0.00%	1.90%	4.614***
Piped water within compound (++++)	1.48%	0.00%	0.00%	0.00%	0.10%	7.30%	4.633***
Protected inside backyard (++)	0.46%	0.10%	0.20%	0.40%	0.80%	0.70%	-0.441***
Protected outside backyard (--)	4.81%	7.10%	5.70%	5.00%	4.20%	2.00%	-0.262***
Rainwater	0.08%	0.00%	0.10%	0.20%	0.10%	0.00%	-0.152
Surface (River, Lake, Lagoon) (----)	11.04%	21.20%	14.20%	10.80%	7.50%	1.50%	-0.514***
Unprotected inside household (--)	1.78%	2.30%	2.30%	2.10%	1.70%	0.50%	-0.236***
Unprotected outside household (----)	16.56%	31.60%	21.60%	15.40%	9.50%	4.90%	0.517***
Water from tank truck	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.000
Other	0.05%	0.00%	0.00%	0.10%	0.10%	0.00%	0.042
Time for taking main water sources							
Under 10 min (++++)	31.84%	6.20%	18.80%	36.90%	47.00%	50.20%	0.590***
Between 10-30 min (----)	45.63%	61.10%	54.00%	41.30%	35.40%	36.30%	-0.281***
Between 30-60 min (----)	17.75%	25.10%	21.40%	17.20%	14.30%	10.70%	-0.252***
More than one hour (---)	4.76%	7.60%	5.80%	4.60%	3.20%	2.70%	-0.280***
Main energy source for lighting							
Batteries (----)	68.79%	84.20%	81.30%	79.20%	63.00%	36.30%	-0.586***
Candles (-)	0.78%	0.10%	0.40%	1.30%	1.90%	0.30%	0.255***
Electricity (++++)	11.65%	0.00%	0.00%	0.00%	7.10%	51.10%	2.774***
Firewood (----)	11.96%	15.10%	16.20%	15.40%	11.20%	1.90%	-0.308***
Gas	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.169
Generator	0.04%	0.00%	0.00%	0.00%	0.10%	0.10%	1.375**
Oil	0.15%	0.00%	0.10%	0.10%	0.50%	0.10%	0.409***
Solar panel (++++)	4.66%	0.10%	0.50%	1.90%	12.00%	8.90%	0.810***
Other (-)	1.55%	0.40%	1.40%	1.90%	3.10%	0.90%	0.176***
Ownership of radio (++++)	24.66%	8.20%	13.80%	22.80%	37.50%	41.00%	0.522***
Ownership of television (++++)	7.47%	0.00%	0.00%	0.00%	0.50%	36.80%	4.455***
Ownership of cell phone (++++)	38.47%	8.80%	21.00%	38.00%	57.70%	66.80%	0.744***

4. Cross-method agreement of wealth indices

Table S4. Spearman's rank correlations and p-values among four wealth index ranks.

	Classical-PCA rank	Feature-selection-PCA rank	Sparse-PCA rank	Robust-PCA rank
Classical-PCA rank	1.000
Feature-selection-PCA rank	0.992 (p-value < 2.2e-16)	1.000
Sparse-PCA rank	0.985 (p-value < 2.2e-16)	0.985 (p-value < 2.2e-16)	1.000	..
Robust-PCA rank	0.996 (p-value < 2.2e-16)	0.994 (p-value < 2.2e-16)	0.984 (p-value < 2.2e-16)	1.000

5. External validation of the wealth indices – Consistency with household income classification

Table S5. Linear regression results and Spearman's rank correlations between average monthly income (2022 US\$) on log scale with four different wealth ventiles⁴ (N=138).

	Linear regression results		Spearman's rank correlation
	Intercept	Slope	
Classical-PCA wealth ventiles	9.684 (p-value < .001)	0.675 (p-value = .009)	0.257
Feature-selection-PCA wealth ventiles	9.680 (p-value < .001)	0.664 (p-value = .011)	0.258
Sparse-PCA wealth ventiles	9.685 (p-value < .001)	0.672 (p-value = .010)	0.260
Robust-PCA wealth ventiles	9.675 (p-value < .001)	0.663 (p-value = .011)	0.256

⁴ Wealth ventile is calculated by dividing all households into equal 5% quantiles based on the wealth index scores.

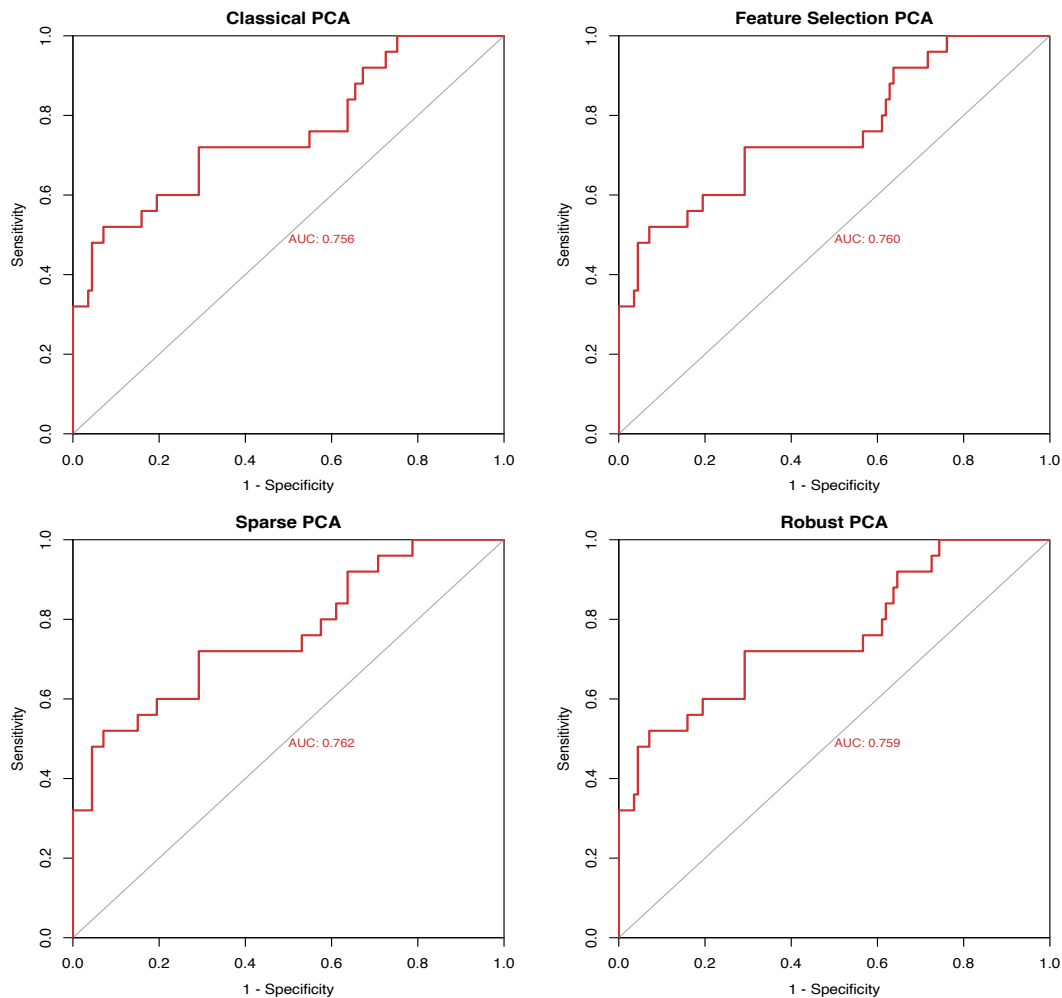


Figure S3. ROC curves and AUC values for classification results of four different wealth indices according to the income-based poverty status (N=138).

6. External validation of the wealth indices – Analysis for household with zero income

As implied in the paper, most of the households (399 out of 537) reported zero total income in the Health Economics Survey data. To provide a more comprehensive analysis of the issue, we investigated the relationship between the wealth index and households reporting zero income.

Figure S4 presents scatter plots of the wealth index quantiles for households with non-zero income and those reporting zero income. It is observed that households with non-zero income demonstrate a larger minimum wealth index, with the two groups exhibiting overlapping wealth index quantiles in the initial 50% quantile. However, in the latter 50% quantiles, households with non-zero income maintain a consistently higher wealth index in comparison to those with reported zero income.

This finding underscores the strong relationship between wealth indices and whether households have zero income. Although the study acknowledges the potential issue of respondent bias in self-reporting zero income, it nevertheless provides valuable insights into the issue at hand.

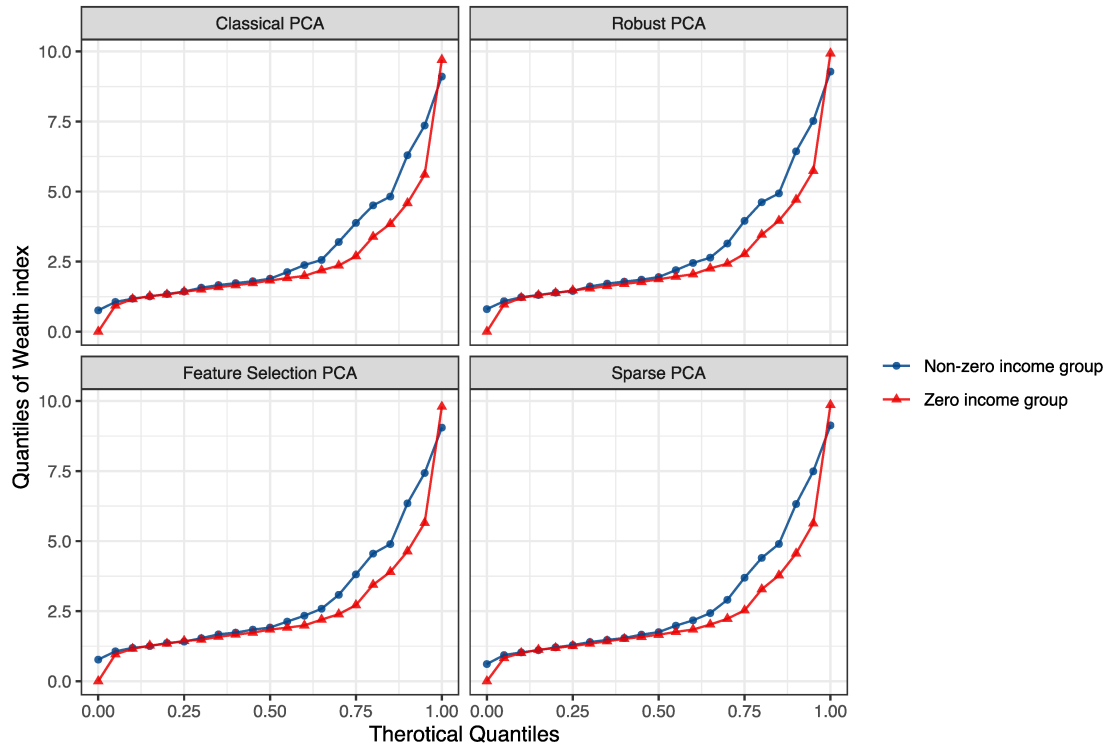


Figure S4. The QQ-plots for quantiles of wealth index of households.

7. Reference

- 1 Rutstein SO. Steps to constructing the new DHS Wealth Index. *Rockville, MD: ICF International* 2015.
- 2 Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2016; **374**: 20150202.
- 3 Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2005; **67**: 301–20.
- 4 Zou H, Hastie T, Tibshirani R. Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics* 2006; **15**: 265–86.