Identifying Poor Households in Nigeria

To: Nigeria Ministry of Economic Planning and Development (NEPD)
From: Jasmine Adams
Date: Monday, January 3, 2022

Research Objective and Summary of Findings

To help the NEPD identify which households to target for social protection programs, we analyze a sample of 5,000 observations from the World Bank LSMS Nigeria General Household Survey to better understand what individual, household, and environmental characteristics relate to household consumption levels. We then aim to create a model that accurately identifies poor households based on the most significant correlates to average monthly consumption. Our findings suggest that working primarily in agriculture, identifying with a political group, and most notably, residing in the northern region (excluding FCT Abuja) are significantly negatively associated with consumption. Conversely, working as an electrician or in healthcare, as well as having insurance, a bank account, or a community youth group were significantly positively associated with consumption. Surprisingly, household size, number of children, and urbanicity had positive, negative, and positive correlations respectively only for households in the southern region. Aside from a significant negative association with urbanicity, results were inconclusive for households in the north. When controlling for these factors, our model correctly identifies poor households with 75.6% efficacy.

Indicators of Poverty

Table 1.

Industry of Employment	Financial Resources	Household (HH)	Community Amenities	Geography
▪ Energy* (Electrician) ▪ Financial Services ▪ Healthcare* ▪ Agriculture*	▪ Insurance* ▪ Bank*	▪ HH Size* ▪ Number of Children* ▪ Head of HH Sex ▪ Connected to Electricity	▪ Political Group* ▪ Youth Group*	▪ Sector (Urban vs. Rural) ▪ Region* (North vs. South) ▪ States†

*p-value < 0.05 in final model
†p < 0.05 in final model for 18/36 states

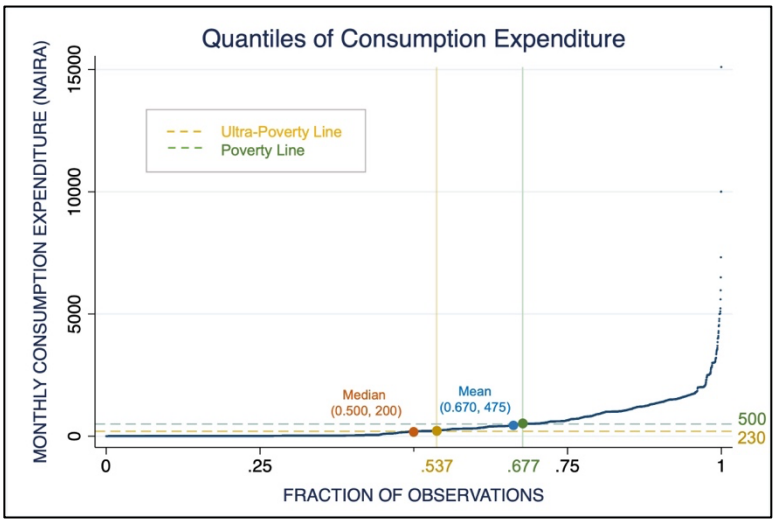
We analyzed a subset of 5,000 households from the panel section of the World Bank LSMS Nigeria General Household Survey, a larger national cross-sectional dataset surveying 20,000 households in 2010-2011 and again in 2012-2013. The panel section contains the average monthly

consumption expenditure (AMCE) of 4,991 households, approximately two-thirds of which are classified as poor as defined by having an AMCE below 500 Naira. Evidence from the sample indicates that region, and more specifically state, are strong predictors of respondents' income group. On average, households in the South (50% of sampled households) are significantly less likely to live below the poverty line compared to households in the North. More curiously, evidence suggests that region moderates the relationship between consumption expenditures and key independent variables including sector, household size, and number of children. The latter two are likely significant for households in the South only. Furthermore, about half of all households sampled report being connected to electricity, most of which cite the grid as their primary source of energy. However, despite 74.1% of unconnected households living below the poverty line, controlling for the significant variables in *Table 1* renders the relationship between AMCE and an electricity connection insignificant. Regardless, if approximately one-third of households are both poor and unconnected to electricity, then we surmise that a policy connecting households to the grid will notably increase welfare among the poor.

Poverty Measurements

We obtained a preliminary understanding of poverty in Nigeria by calculating the poverty headcount ratio. Approximately 67.02% and 53.74% of sampled households have an AMCE below the poverty line (500 Naira) and ultra-poverty line (230 Naira) respectively. Of those categorized as poor, only 19.82% consume above the ultra-poor poverty line, signaling that extreme poverty is rife among most of the populace. Given that a large majority of poor households are ultra-poor, the ultra-poverty line (230 Naira), which falls closer to the sample median (200 Naira), is a more appropriate marker of poverty than the standard poverty line (500 Naira). While the latter may be closer to the sample mean (475 Naira), the sample mean is heavily skewed by outliers and represents a consumption level far greater than that of the "typical" Nigerian household. Nevertheless, our upcoming model is based on a linear regression, which is biased toward predicting a normal (i.e., symmetrical) distribution around the mean rather than a skewed distribution around the median. Because of these limitations, we use poverty rather than ultra-poverty to assess the accuracy of our model.

Graph 1.



Complimentary to the headcount index, the Foster-Greer-Thorbecke poverty index offers two additional measurements poverty expressed as a ratio of the poverty line – poverty gap and

poverty severity. The former, FGT(1), measures the poor income group's economic shortfall per total capita. The latter, FGT(2), represents the magnitude of welfare inequality among the poor by giving more weight to those furthest below the poverty line. Our sample has a poverty gap of 0.53, meaning that the Nigerian government would need to increase consumption by 0.53 of the poverty line per total capita (1,322,615 Naira total) to eradicate poverty in Nigeria. The poverty severity is 0.47, meaning that the poorest households need about double the boost in consumption to meet the poverty threshold compared to the average poor household.

The headcount ratio, poverty gap, and poverty severity each provide limited but essential information regarding household welfare and the economic impact of targeted social programs. As a simple ratio of poverty, we can utilize the headcount index to estimate poverty prevalence and gauge how targeted households' increased consumption will shift the distribution of income groups within the country. Nonetheless, unlike the headcount ratio, which only measures poverty prevalence, poverty gap and poverty severity express the frequency, as well the nature of poverty. Specifically, we can use the poverty gap to estimate how much targeted interventions will likely mitigate the average economic discrepancy between poor households and the poverty threshold. We can then use poverty severity to assess poverty beyond the aggregate level and determine how well programs reach the most impoverished households. Taken together, these three poverty indices can provide a holistic estimate of targeted social programs' overall impact on welfare.

Predicting Household Consumption Levels

Given the challenges of collecting households' financial data, we sought to create a model that accurately identifies poor households using easily observable criteria. We looked at factors related to the five key areas described in *Table 1*, including industry of employment, financial resources, household attributes, community amenities, and geography. We predicted that jobs in more lucrative sectors such as energy and healthcare have positive associations with consumption and that jobs in less lucrative industries, such as agriculture, likely have negative associations. We also proposed that having an increasing number of children or female head of household have a negative correlation to consumption, while financial safety nets, electricity connections, urban neighborhoods, and community amenities are likely positively correlated with consumption. That is, aside from having a political group in one's community, which we predicted to have a negative association given communities are often more politically engaged when welfare issues are salient. Furthermore, we hypothesized that different localities would have strong correlations to welfare since it is common for economic disparities to follow geographical lines. After noticing a pattern in states' average consumption, we predicted that households in the North were more likely to be disadvantaged compared to those in the South, perhaps due to economic underdevelopment in the region. Finally, we reasoned that household size may have a dynamic relationship with consumption since whether a growing number of residents represents a financial burden or increasing source of income may depend on conditions such as the local job or housing market.

Model 1.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 \underline{X_5} + \beta_6 \underline{X_6} + \beta_7 \underline{X_7} + \beta_8 X_8 + \beta_9 X_9 + \beta_n X_n \ldots$$
$$(X_1 \times X_2) \quad (X_1 \times X_3) \quad (X_1 \times X_4)$$

Table 2.

Consumption Index (Y)		Coefficient	β	Std. Err.(e)	t	p-value	[95% Conf.	Interval]
Const.		374.45	β_0	59.56395	6.29	0.000	257.6741	491.226
North	X_1	246.7226	β_1	49.46625	4.99	0.000	149.7434	343.7019
Urban	X_2	114.019	β_2	29.59316	3.85	0.000	56.00122	172.0368
Number of children	X_3	-53.50561	β_3	11.53956	-4.64	0.000	-76.12908	-30.88215
Household size	X_4	54.1459	β_4	8.867168	6.11	0.000	36.7617	71.5301
North × urban	X_5	-180.1368	B_5	42.18626	-4.27	0.000	-262.8436	-97.43011
North × number of children	X_6	51.57419	β_6	16.55533	3.12	0.002	19.11723	84.03114
North × household size	X_7	-51.21153	β_7	12.91859	-3.96	0.000	-76.5386	-25.88446
Connected to electricity	X_8	6.848573	β_8	22.03921	0.31	**0.756**	-36.35959	50.05674
Head of HH female	X_9	-21.04859	β_9	26.40819	-0.80	**0.425**	-72.8222	30.72503
Electrician	X_{10}	229.9679	β_{10}	87.98217	2.61	0.009	57.47766	402.4581
Healthcare worker	X_{11}	192.3738	β_{11}	50.71094	3.79	0.000	92.95432	291.7933
Farmer/Agriculture Sector	X_{12}	-100.1478	β_{12}	20.19159	-4.96	0.000	-139.7337	-60.56189
Insurance	X_{13}	283.944	β_{13}	57.45557	4.94	0.000	171.3016	396.5864
Bank	X_{14}	50.48225	β_{14}	21.72049	2.32	0.020	7.898936	93.06556
Political group	X_{15}	-137.8344	β_{15}	27.80966	-4.96	0.000	-192.3556	-83.31314
Youth group	X_{16}	56.43481	β_{16}	25.38644	2.22	0.026	6.664349	106.2053
†*FCT Abuja (N1)*	X_{17}	1010.42	β_{17}	101.2722	9.98	0.000	811.874	1208.965
Plateau (N1)	X_{18}	-173.1299	β_{18}	59.02946	-2.93	0.003	-288.8579	-57.40183
Niger (N1)	X_{19}	-224.8449	β_{19}	51.83049	-4.34	0.000	-326.4593	-123.2306
Adamawa (N2)	X_{20}	-269.5905	β_{20}	57.59555	-4.68	0.000	-382.5073	-156.6736
Bauchi (N2)	X_{21}	-239.2924	β_{21}	50.40645	-4.75	0.000	-338.1149	-140.4699
Gombe(N2)	X_{22}	-197.5429	β_{22}	69.41293	-2.85	0.004	-333.6279	-61.45793
Yobe (N2)	X_{23}	-360.856	β_{23}	55.42534	-6.51	0.000	-469.5181	-252.1939
Jigwa (N3)	X_{24}	-488.4607	β_{24}	58.1254	-8.40	0.000	-602.4163	-374.5051
Kano (N3)	X_{25}	-382.6886	β_{25}	47.56532	-8.05	0.000	-475.941	-289.4361
Katsina (N3)	X_{26}	-167.9625	β_{26}	48.46843	-3.47	0.001	-262.9856	-72.93954
Kebbi (N3)	X_{27}	-429.4995	β_{27}	61.59821	-6.97	0.000	-550.2636	-308.7354
Zamfara (N3)	X_{28}	-355.2672	β_{28}	65.26909	-5.44	0.000	-483.2281	-227.3063
Abia (S4)	X_{29}	-376.2958	β_{29}	65.62106	-5.73	0.000	-504.9467	-247.6448
Ebonyi (S4)	X_{30}	-107.8071	β_{30}	58.20915	-1.85	**0.064**	-221.9269	6.312755
Imo (S4)	X_{31}	166.4607	β_{31}	47.18039	3.53	0.000	73.96292	258.9585
Bayelsa (S5)	X_{32}	391.9669	β_{32}	78.14015	5.02	0.000	238.7721	545.1617
Rivers (S5)	X_{33}	264.2494	β_{33}	44.58225	5.93	0.000	176.8453	351.6535
Lagos (S6)	X_{34}	582.4924	β_{34}	55.0763	10.58	0.000	474.5146	690.4703
Oyo (S6)	X_{35}	219.968	β_{35}	45.00432	4.89	0.000	131.7365	308.1996

†*N* = north; *1* = zone number

Results

After discounting missing values and excluding outlying households with a reported AMCE above 5,000 Naira (0.32% of households), Model 1 estimates observable variables' relationship to household consumption based on the reported AMCE of 4,355 households. As expected, geospatial determinants had the largest associations with monthly consumption expenditures (Y). All northern states in *Table 2* had significant adverse relationships with consumption ($X_{17} - X_{28}$: $p < 0.005$) except for FCT Abuja ($p = 0.000$), whose significant positive relationship denotes the largest correlation to welfare in the regression ($\beta = 1010.4$.). Aside from Abia, which presented a significant negative association with consumption ($p = 0.000$), and Ebonyi, who's adverse correlation was just shy of significance at $\alpha = 0.05$ ($p = 0.064$), most southern states of notable importance were favorably related to economic welfare ($X_{31} - X_{35}$: $p < 0.000$).

More interestingly, considering interactions with the northern region (X_1) moderated the relationship between household consumption and three independent variables: urban sector (X_2), number of children (X_3), and household size (X_4). When we constrained our regression by region to verify the direction and significance of these interactions [i.e., ($X_1 \times X_2$), ($X_1 \times X_3$), and ($X_1 \times X_4$)], our results confirmed that urban dwellings were significantly correlated to increased consumption in the south ($p = 0.001$; $\beta = 116.8$), and decreased consumption in the north ($p = 0.040$; $\beta = -70.6$). Moreover, while households with more children were adversely correlated to consumption in the south ($p = 0.000$ if $X_1 = 0$), we fail to reject the null hypothesis that a household's number of children is unrelated to welfare in the North (H_0: $\beta_3 = 0$; $p = 0.777$ if $X_1 = 1$). We observed a mirrored effect for household size (H_0: $\beta_4 = 0$; $p = 0.609$ if $X_1 = 1$), which adversely related to consumption for southerners only ($p = 0.000$ if $X_1 = 0$).

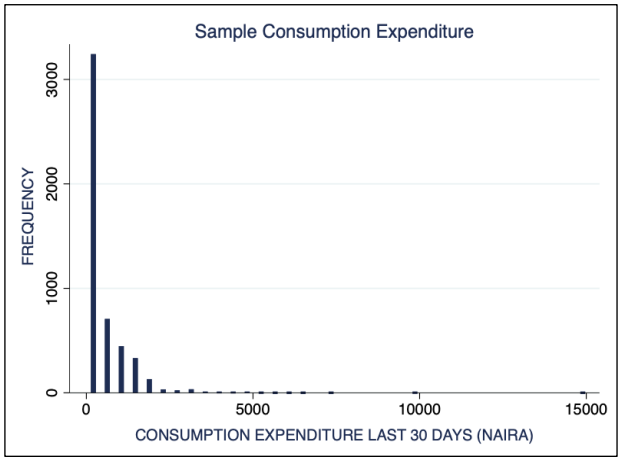
Regarding characteristics unaffected by region, our findings refute the assumption that the head of household's sex* ($p = 0.425$) and whether a property is connected to electricity ($p = 0.756$) are noteworthy indicators of welfare. In contrast, evidence supports our prediction that having a community-based political group ($p = 0.000$) and working in agriculture ($p = 0.000$) are conversely related to welfare. Finally, the model corroborates our expectations that employment as an electrician ($p = 0.009$) or healthcare worker ($p = 0.000$), and having insurance ($p = 0.000$), a bank account ($p = 0.020$), or a community youth group ($p = 0.026$) exhibit a significant, positive association with consumption. Thus, we conclude that the variables in our regression model account for 19.72% of the variation in total household consumption expenditures (R^2-adjusted = 0.1972).

Comparing Predicted and Reported Consumption

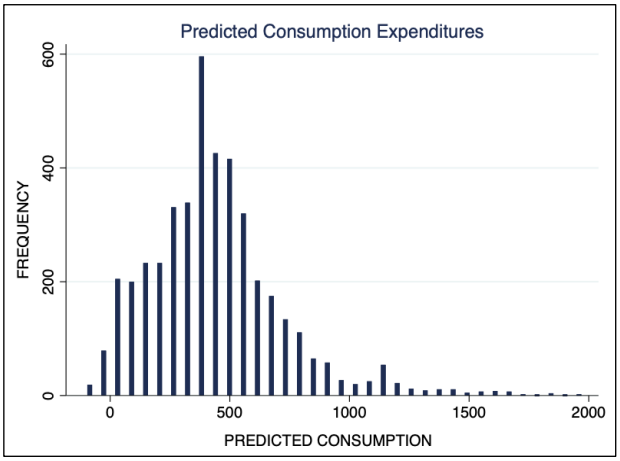
Given that the distribution of reported consumption expenditures is heavily right skewed with a markedly long tail (Graph 2), we discount outliers with consumption expenditures above 5,000 Naira to mitigate their distortive effects on the sample mean ($\bar{y}_i = 475.8$ Naira; SD = 747.7; skew = 4.47). We use the resulting sample distribution absent of missing values as the base for

*Variable is significant after discounting nonpoor households but not when accounting for the total population.

Graph 2.



Graph 3.



determining the efficacy of our model ($\bar{y}_j$ = 435.6 Naira; SD = 636.6; skew = 2.26). As illustrated in *Graph 3*, the distribution of predicted consumption expenditures maintains a slight right-skew with an elongated right-tail but is otherwise fairly balanced around the mean ($\hat{x}$ = 435.6 Naira; SD = 287.3 Naira; skew = 1.21). After conducting a two-sample t-test comparing the sample reported mean to the mean predicted by the model, we reject the null hypothesis that there is a significant difference between the two (p = 0.000). Still, the predicted sample median (425.0 Naira) is considerably higher than the sample reported median (140.0 Naira), underscoring the model's limited capacity for distinguishing between varying levels of poverty severity. Despite this shortcoming, our framework correctly identifies 75.6% of poor households. Moreover, of the households it predicts are poor, 21.3% report consumption levels above the poverty line.

Takeaways

The NEPD's primary objective is to maximize program benefits among the poor, not to preclude financially secure households from receiving benefits. As such, although the model is almost twice as likely to misidentify nonpoor households compared to poor households, the NEPD should prioritize mitigating the risk of failing to detect households living in poverty rather than minimizing this discrepancy. We therefore concede that the model's 75.6% efficacy rate is passable but should certainly be improved using mixed-methods statistical models in future research.

Data Resource

Nigeria National Bureau of Statistics. (2018–2019). General Household Survey, Panel (GHS-Panel). *The World Bank*.