WPS7024

Policy Research Working Paper 7024

Estimating Poverty in the Absence of Consumption Data

The Case of Liberia

Andrew Dabalen
Errol Graham
Kristen Himelein
Rose Mungai

WORLD BANK GROUP
Poverty Global Practice Group
&
Macroeconomics and Fiscal Management Global Practice Group
September 2014

Abstract

In much of the developing world, the demand for high frequency quality household data for poverty monitoring and program design far outstrips the capacity of the statistics bureau to provide such data. In these environments, all available data sources must be leveraged. Most surveys, however, do not collect the detailed consumption data necessary to construct aggregates and poverty lines to measure poverty directly. This paper benefits from a shared listing exercise for two large-scale national household surveys conducted in Liberia in 2007 to explore alternative methodologies to estimate poverty indirectly. The first is an asset-based model that is commonly used in Demographic and Health Surveys. The second is a survey-to-survey imputation that makes use of small area estimation techniques. In addition to a standard base model, separate models are estimated for urban and rural areas and an expanded model that includes climatic variables. Special attention is paid to the inclusion of cell phones, with implications for other assets whose cost and availability may be changing rapidly. The results demonstrate substantial limitations with asset-based indexes, but also leave questions as to the accuracy and stability of imputation models.

This paper is a joint product of the Poverty Global Practice Group and the Macroeconomics and Fiscal Management Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The authors may be contacted at khimelein@worldbank.org.

Estimating Poverty in the Absence of Consumption Data: The Case of Liberia

Andrew Dabalen, Errol Graham, Kristen Himelein, and Rose Mungai[1]

World Bank

1818 H Street NW, Washington DC, 20433, USA

Keywords: poverty measurement, asset indices, cross-survey imputations, Liberia

JEL classification: D6, I30

Sector: POV

[1] Andrew Dabalen, Kristen Himelein, and Rose Mungai are members of the World Bank's Poverty Global Practice, and Errol Graham is in the Macroeconomics and Fiscal Management Global Practice. Kristen Himelein is the corresponding author and can be reached at khimelein@worldbank.org. This paper was written as a methodological background paper to the "Liberia Poverty Note Tracking the Dimensions of Poverty Report" report (World Bank, 2012). The authors benefited from information from Shea Rutstein at ICF International on the calculation of the DHS indices.

1. Introduction

Liberia remains one of the poorest countries in Africa. It ranks 174 of 187 countries on the United Nation Development Programme's Human Development Index, and in 2007 an estimated 64 percent of the population, or more than one and a half million Liberians, lived below the national poverty line, with 48 percent of the population living in extreme poverty[2], as defined by the national poverty line in 2007 (30,224 Liberian dollars per year per adult equivalent). There have, however, been dramatic improvements in a number of key indicators of well-being since the end of the brutal civil conflict a decade ago. Sustained economic growth and stability have allowed the country to make notable progress on key Millennium Development Goal (MDG) targets. The proportion of the population living below the poverty line is estimated to have fallen since 2007. The country has also achieved the target for reducing under-nutrition, while surpassing the target for reducing child mortality and making notable progress in reducing maternal mortality. Given progress to date, Liberia will also likely meet the MDG targets for eradicating hunger, promoting gender equality and empowering women and establishing a global partnership for development. As many aspects of household well-being are changing rapidly, it is essential to have frequent and accurate data to assess poverty trends and anchor poverty reduction policies on evidence.

However, despite having a great need for data, Liberia also has limited financial and human capital resources with which to collect it. Funding is limited for all programs, including statistics, and conflict interrupted the education for an entire generation of Liberians, limiting the available pool of statisticians. In this context, all possible sources of data must be leveraged to their maximum usefulness for poverty monitoring. Most household surveys implemented in the country, however, do not collect the detailed consumption information necessary to construct aggregates or poverty lines. Surveys of this nature, such as Living Standards Measurement Studies or Household Income and Expenditure Surveys, are conducted maybe once or twice a decade. They are costly to implement, and because of concerns about seasonality, require 12 months to collect, followed by time to clean and compile data. While these surveys offer a wealth of complex integrated data, they are not sufficiently nimble to generate real-time monitoring information and provide feedback to improve the allocation of resources in the short term.

The last survey which included a consumption module was the Core Welfare Indicator Survey (CWIQ) conducted in 2007. Beyond the large integrated household well-being survey, Liberia Institute of Statistics & Geo-Information Services (LISGIS) collects data through a number of lighter, more specialized instruments. Also in 2007, LISGIS, in partnership with ICF International, conducted a Demographic and Health Survey (DHS). In 2010, a lighter version of the CWIQ survey, excluding the consumption section, was conducted. This paper makes use of the fact that the 2007 and 2010 CWIQ surveys had similar questionnaires outside of the consumption section, and that the 2007 DHS and CWIQ surveys shared a common listing operation (samples were drawn from the same population of households living in a cluster), to explore two alternative methodologies for estimating well-being in the absence of consumption data. The first is an asset-based index, commonly used in the DHS surveys. The second is a survey-to-survey imputation method which uses the relationship

[2] National poverty line estimates from 2007 Core Welfare Indicator Questionnaire. See Annex 2 in World Bank (2012) for details on the calculation.

between the covariates of poverty in the survey with consumption to predict poverty levels in the survey without consumption.

Section 2 of the paper provides a brief review of the relevant literature, section 3 describes the data, section 4 discusses the asset model while section 5 describes the construction and analysis of the survey-to-survey model. Section 6 concludes with some suggestions for future research.

2. Literature Review

The need for timely poverty estimates for evidence-based policies in the face of the high cost of fielding comprehensive surveys to track income and/or expenditure has led to the development of a variety of approaches for estimating poverty in the absence of consumption expenditure or income data. The two most popular methodologies currently in use are the construction of proxy measure based on observed assets and other household characteristics and survey-to-survey imputation.

The most common approach is to construct an asset index to proxy consumption based on the information available in the data set. This is often the only alternative if a comparable survey including consumption is not available. It is also preferred by some analysts since it is able to take into account dimensions of well-being beyond the simple monetary consumption or income measures. There are a number of different methodologies which can be used to construct the asset index, including factor analysis (see Sahn and Stifel, 2000) and multiple correspondence analysis (see Booyen et al, 2008), but the most common is principal component analysis, as employed by Filmer and Pritchett (2001).

Filmer and Pritchett use data from India to estimate the relationship between household wealth and children's school enrollment. As a proxy for wealth they constructed a linear index of asset ownership indicators (which included both assets and housing characteristics) using principal component analysis to derive weights. Filmer and Pritchett argue that the econometric evidence from their work suggests that the asset index, as a proxy of economic status for use in predicting enrollments, is at least as reliable as conventionally measured consumption expenditures, and sometimes more so. Filmer and Scott (2011), however, find that while asset indices are robust estimators of inequalities in characteristics related to education, health, and labor, there are substantial differences in the poverty rankings of the two measures. Howe et al (2009) conducted a review of the correspondence between wealth indices and consumption measures for 33 surveys in developing countries, and found wealth indices to be a poor proxy. They further found the most favorable results to be in middle-income over low income settings, and urban areas over rural areas. Furthermore, a technical note compiled by the World Bank in 2003 found "living standard indices based on principal components analysis often have a weak relationship with consumption, with correlation coefficients often in the region of 0.2-0.4" (World Bank 2003, p.8). In light of these findings, analysts have increasingly been seeking alternative methods of estimation.

The survey-to-survey imputation requires that at least one previous comparable survey contains household-level consumption information. The method draws upon the imputation literature (see Brick and Kalton, 1996 for a discussion of various techniques), which utilizes non-missing data in a larger data set to predict the values for missing variables, and from the poverty targeting literature

(see Grosh and Baker, 1995) which seeks proxies for poverty status from household characteristics. The survey-to-survey imputation method extends that to using common variables in two data sets, only one of which contains consumption to predict consumption values in the second data set. A common application of this method in poverty analysis is "poverty mapping," which uses consumption and poverty estimates from a household survey imputed into census data to achieve very fine levels of geographic disaggregation. (See Rao, 2003, for a discussion of the general technique of small area estimation, and Elbers et al., 2002, specifically related to poverty mapping.) For examples of survey-to-survey imputation for poverty analysis, see Stifel and Christiaensen (2007), Tarozzi (2007), Grosse et al (2009), and Douidich et al (2013).

3. Data

Between 2007 and 2010, the Government of Liberia, through the Liberia Institute of Statistics and Geo-Information Services (LISGIS), conducted two nationally representative Core Welfare Indicator Surveys (CWIQs) and a Demographic and Health Survey (DHS). Since the surveys were conducted so close in time, the sample design and listing operation were shared between the 2007 CWIQ and 2007 DHS. Both surveys used the 1984 Population Census as the sampling frame to select the enumeration areas (EAs), and then a household listing operation was conducted between March and May 2006. For EAs outside of the capital city of Monrovia, a single listing operation was done, and then the selected households were allocated between the two surveys. In Monrovia, separate EAs were selected and listing operations were independent.

The DHS survey was collected first, between December 2006 and April 2007, and the CWIQ in August and September, 2007. Both surveys used a two-stage stratified sample, with 300 EAs selected (though two were dropped from the DHS due to selection error). The CWIQ survey collected information from 12 households in each EA and the DHS from 25. Both surveys were designed to be representative at the urban/rural level and at the level of the six regions: Greater Monrovia, North Central, North Western, South Central, South Eastern A, and South Eastern B. Data is available for 6,824 households in the DHS final data set and 3,595 households in the final CWIQ data set.

The sample design for the 2010 CWIQ was similar to that of 2007, with a two-stage stratified sample with 500 EAs in which 12 households were selected. The survey had a total sample size of 5,990 households, and the bulk of the data was collected in March and April, 2010. The 2010 CWIQ was designed to be representative at the urban/rural level and at the level of Monrovia and the 15 counties. The main difference between the two surveys was that the 2007 CWIQ contained an extended consumption and income module that was not included in the 2010 survey.

4. Asset Model

Since the DHS surveys do not collect consumption information to construct a poverty aggregate, an asset-based methodology is employed to proxy poverty quintiles. The methodology uses factor analysis with the following variables: housing characteristics (water supply, toilet, type of floor, walls and roof), cooking fuel, asset ownership (generator, table, chairs, cupboard, mattress, sewing machine, computer, boat/canoe, radio, television, refrigerator, bicycle, motorbike, car, cellphone, and watch), having a bank account, and having electricity. We constructed a similar aggregate to the

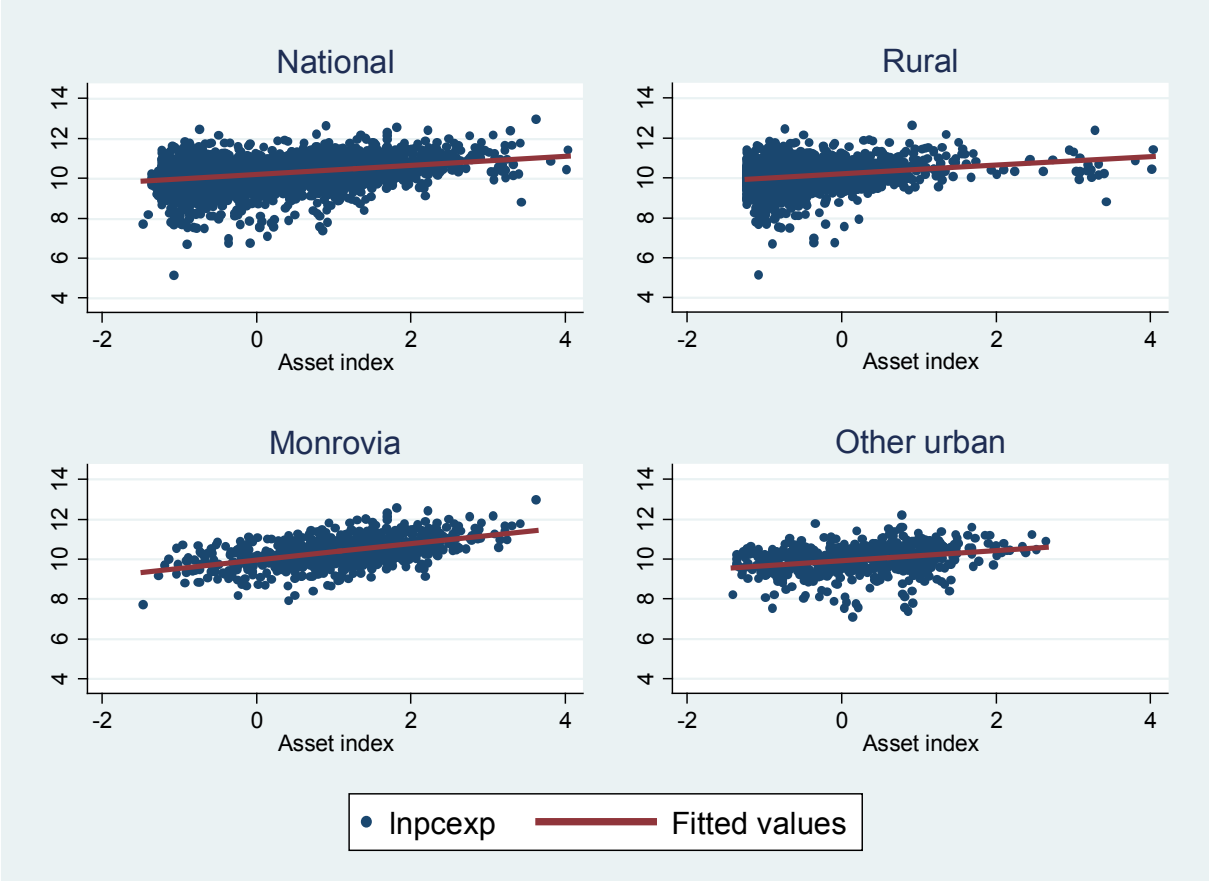
DHS using the CWIQ data.[3] Since the 2007 DHS and 2007 CWIQ were conducted so close in time period and shared a listing operation, we would expect the results from the two to be comparable, and indeed there was no statistically significant difference between the two measures at the national level. At the sub-national level, however, some differences were found. The DHS asset score showed consistently higher results in urban areas over the CWIQ scores, but no difference in rural areas. There were also weakly significant differences within some of the strata. Monrovia, South Central, and South Eastern B are higher in the DHS, but largest differences are in South Eastern A, which is higher in the CWIQ. While there is a relationship between the degree of urbanization and the differences observed – Monrovia is obviously higher in the DHS as it is 100 percent urban – this does not completely explain the discrepancies. Table 1 in the appendix shows the mean for the asset-based index by survey and strata. For example, South Eastern A, South Eastern B, and North Central have roughly equal proportions of urban population, but the CWIQ estimates are higher in South Eastern A, lower in South Eastern B, and there is no difference in North Central. There is also substantial variation in the ranking of the regions based on data sets. The DHS data classifies South Eastern A as the poorest region while the CWIQ data ranks it as the second wealthiest after Monrovia.

It is not immediately clear what is driving the differences between the two surveys as the design of the index is identical and the respondents selected from the same enumeration areas. Seasonal differences may be a factor as the DHS was collected first and over a longer period, though this is somewhat unlikely due to the static nature of most of the included variables. It could also be a data quality effect. For example, one or more of the teams or enumerators could have been more diligent in collecting data on assets, leading to higher estimations.

By using the data from the 2007 CWIQ, we are able to compare the rankings of the households using the asset index and the consumption data. We find that there is a positive and significant correlation of 0.319 between the two measures, which is consistent with previous findings (see World Bank 2003, 8). The predictive power, however, is weak as the asset index explains only 10.8 percent of the variation of per capita expenditure. The predictive power is also weaker in rural areas than in urban. In rural areas the model explains only 4.4 percent of the variation compared to 24.5 percent in urban areas. This result though is driven largely by Monrovia. In Monrovia the asset index explains 26.2 percent of the variation, compared with only 8.2 percent in urban areas outside Monrovia. Figure 1 below shows a series of scatterplot graphs of the log per capita expenditure and the constructed asset index at national and sub-national levels.

[3] The construction varies only slightly between the two surveys, due mainly to the difference in statistical analysis software. The DHS analysis was conducted in SPSS, and it was necessary to replace missing values with the mean. The corresponding analysis was done in Stata for the HIES variables which excluded variables with zero variance (all households had or did not have an item). Overall, the results should be considered comparable.

Figure 1 : Scatterplot of Log Per Capita Expenditure and Asset Index



Source: Authors' calculation based on CWIQ 2007

The asset index used in the DHS generally does not assign a strict ranking nor designate a poverty line as is commonly done with consumption aggregates. Instead, quintiles are constructed. Tables 2a through 2c in the appendix show the frequency of observations for population quintiles constructed using the asset index and the consumption aggregate. The findings corresponded to the results of the scatterplots above. The Pearson chi-square relationship between the rows (consumption quintile) and columns (asset quintiles) is positive and significant. The Cramér's V statistics however are low, indicating the explanatory powers are low. The Cramér's V statistic was similarly higher also for urban areas than rural areas.

In addition, the findings from the asset indices contradict some standard findings in the literature regarding correlates of poverty. One finding that is fairly common across poverty analyses is that poor households have larger household sizes than rich households. The consumption quintiles show an average household size of 4.2 members for the richest quintile, which increased to 5.4, 5.7, 6.2, and finally 6.3 members in the poorest quintile. The asset index quintiles show the reverse trend. They show an average household size of 5.7 in the richest quintile, decreasing to 5.5, 5.5, 5.4, and finally 5.1 in the poorest quintile

5. Survey-to-Survey Imputation

5.1 Model Building

5.1.1 Standard Model

Given the discussion above, it would be of interest to be able to generate better proxies for well-being in surveys that do not include consumption modules. In the specific case of Liberia, in order to monitor poverty reduction, it was necessary to estimate well-being from the 2010 CWIQ data, which collected many of the same variables as the 2007 survey but without the consumption module. For these purposes, a survey-to-survey imputation model was developed. The standard steps in building such a model are, first, to construct an exhaustive list of overlapping variables from the two surveys. These variables should focus on the correlates of poverty, such as demographic information, household head characteristics, household assets (including characteristics of the physical dwelling), landholdings, assets, and geographic location. Once this list is constructed, a stepwise regression is run on the 2007 CWIQ data which contains consumption, with log per capita household expenditure as the dependent variable. In this case, we used a backwards-selection model in which variables with p>=pr(0.05) were removed from the model.[4] Once the parsimonious model has been developed, these variables are regressed against log household consumption. The β coefficients from this model are then applied to the same variables in 2010 CWIQ to predict poverty levels in that survey.[5] Since the counterfactual distribution is built using the parameter estimates from the original survey, in this case 2007 Liberian dollars, it is not necessary to adjust the existing poverty line for inflation[6]. The same line can be used and standard poverty analysis can follow.

There are some limitations with using this methodology to impute values. The most central assumption is that the parameters used for prediction are relatively stable over time. For example, if the level of education among household heads was to increase substantially between the two surveys, but the returns to education remain stable, the model would capture the drop in poverty associated with the higher overall level of education. If, however, the returns to education were to substantially change, either increasing as the structure of economy changes or decreasing as more

[4] It should be noted that there are many valid and well-documented concerns regarding the use of stepwise regression modeling in poverty analysis, see Thompson (1995) for a discussion. The author raises a number of issues with stepwise techniques, but the one most applicable to survey-to-survey imputation is the "tendency towards non-replicable results." This stems from the overly large influence of sampling error in the selection of included variables. If a variable is selected because of a spurious correlation with the variable of interest, this variable may be selected into the model at the expense of other variables with true correlations. In subsequent rounds of data, including the second survey used in the imputation, this relationship would not be correct. While this is a concern for survey-to-survey imputation, it is somewhat mitigated by the relatively small degree of sampling error associated with large household surveys.

[5] There are a number of methods that can be used for this imputation process. The World Bank Development Economics Research Group has developed a free software (PovMap) which can execute these calculations. For the purposes of this paper, we used the multiple imputation package in Stata to perform the estimations. The results for multiple imputations are similar but not identical to those generated by PovMap. Further research into the two methods is ongoing. Our code for "mi impute" is provided in the appendix.

[6] Note that the imputation procedure is akin to a quantity index where the X (quantities varies across surveys) but parameters ("prices") are kept unchanged. In this way, the new counterfactual distribution is valued in the original "prices" and so everything is in real terms and there is no need for further price adjustment.

educated graduates enter the market, the model would become less accurate in predicting poverty. In addition, there are areas of concern related to assets, particularly with regard to cell phones (discussed in more detail below). Since the model relies on this stability, the accuracy of the predictions deteriorates as more time elapses from the base year to the imputation year. Fresh data collection including consumption is necessary periodically to recalculate the base model, and modeling therefore can never eliminate the need for new survey data.

If consumption data were available for 2010, we could have performed checks of our analysis similar to what was done in the previous section with regard to the asset index. This type of analysis was performed by Douidich et al (2013) for Morocco and showed that the imputation model was fairly accurate in predicting even large changes in poverty. Since the data was not available for these types of validation exercises, we focused on the explanatory power of the prediction model with the 2007 data before exploring the robustness of these estimates to other specifications.

Using the complete national sample, a stepwise regression was used to reduce the list of variables to the parsimonious model listed in the appendix as "base national." The base national model gives an overall R^2 of 0.339, including an adjustment for the stratified cluster design. This means that just over one-third of the variance in consumption is explained by the variables we have been able to include in our model. Using Stata's *nestreg* function, we decomposed the variance into component parts. In the base model, dwelling characteristics explain the largest component of the variance at 31.4 percent of the total, followed closely by the demographics and household assets with 27.8 percent and 21.6 percent of the total variance, respectively. The included characteristics of the household head explained most of the remainder, with 12.6 percent, and land ownership, 2.3 percent. An alternative specification including county fixed effects was also used as a robustness check. The overall R^2 increased to 0.363, and of the total variable 7.7 percent was explained by the county variables. The addition of these variables, however, reduced the degrees of freedom of the model and posed between-term collinearity issues when included in more disaggregated models. Facing the choice of excluding the county or other explanatory variables (mainly describing dwelling characteristics), we decided to remove the county variables as they may cause the poverty measures to be overly sticky when predicted over time.

These results partially illustrate the limitations of the asset index methodology. Overall, the explanatory powers of the predictors of poverty available in the survey are limited, and beyond this, the two main components of the asset index methodology, dwelling characteristics and household assets, together account for only half of the explained variation in consumption.

5.1.2 Urban / Rural Models

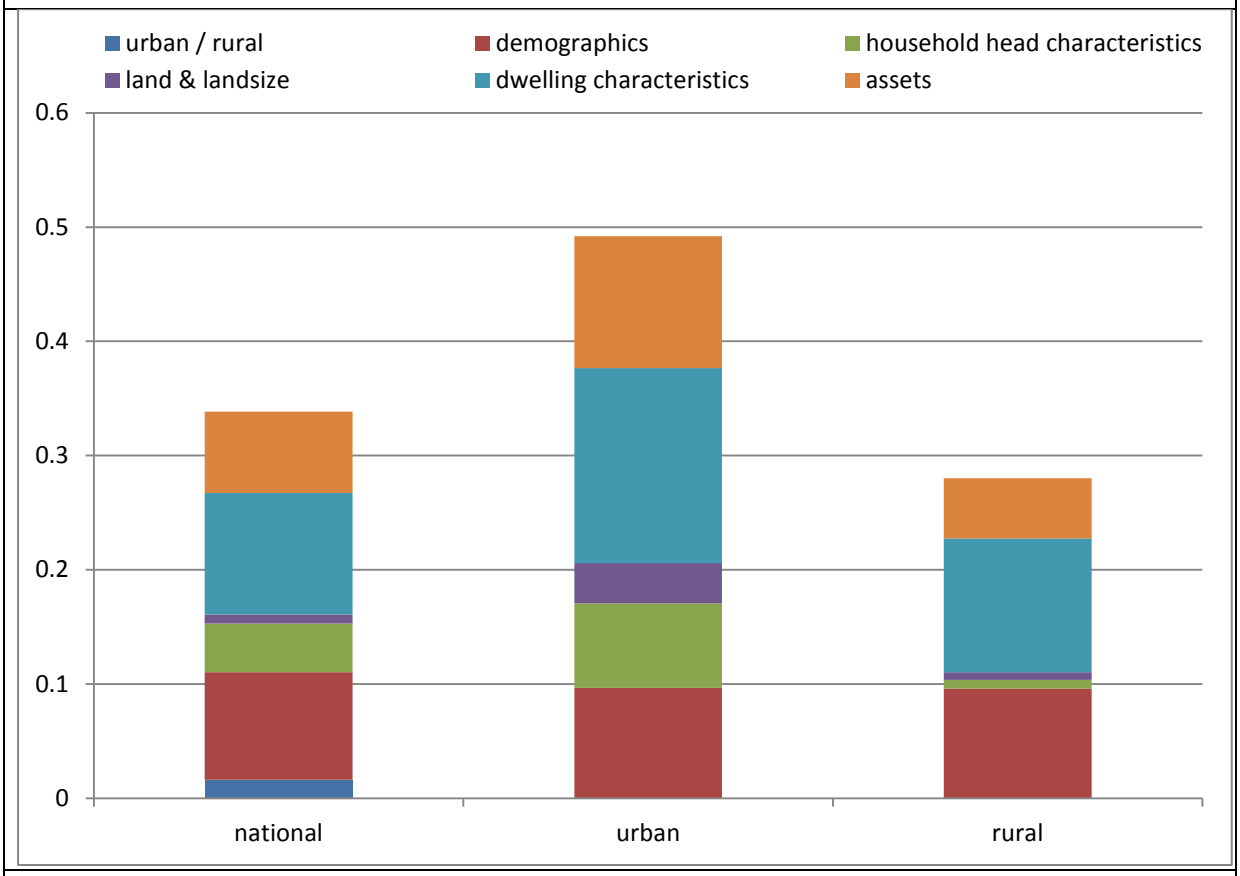
One possible explanation for the relatively poor performance was the substantial difference between urban and rural areas. To increase the flexibility, we estimated a new set of models separately for urban and rural areas. We began with the same universe of possible variables and re-ran the stepwise regressions with the individual sub-samples. Using these results, we predicted poverty within urban and rural areas, and then appended the results to estimate national poverty levels.

Separately estimating the model substantially increased the explanatory power in urban areas, but it was slightly lower in rural areas. The R^2 for urban areas was 0.492 and was 0.280 for rural areas,

compared with 0.339 in the overall model. As nearly 40 percent of the population is classified as living in urban areas, including just over one-quarter of the total population in Monrovia itself, the greater precision in urban areas benefits the overall explanatory power.

The relative importance of the variance components also varied between urban and rural areas. In urban areas, the most important component of the variance was dwelling characteristics, which explained 34.7 percent of the total, followed by the asset ownership, which explained 23.6 percent. As these two components form the majority of the DHS model, this demonstrates the higher correlation between the two methodologies in urban areas. Of the other components, demographics explained an addition 19.6 percent, the characteristics of the household head 15.1 percent in urban areas, land ownership and land size 7.1 percent. In contrast, in rural areas the two most important components were the dwelling characteristics and the household demographics, which explained 41.7 and 34.2 percent of the variance, respectively. Household asset ownership explained 18.8 percent, household head characteristics 2.7 percent, and land ownership 2.5 percent. Figure 2 shows the relative importance of the variance components nationally, and in the urban and rural sub-samples.

Figure 2: Components of Total Variance

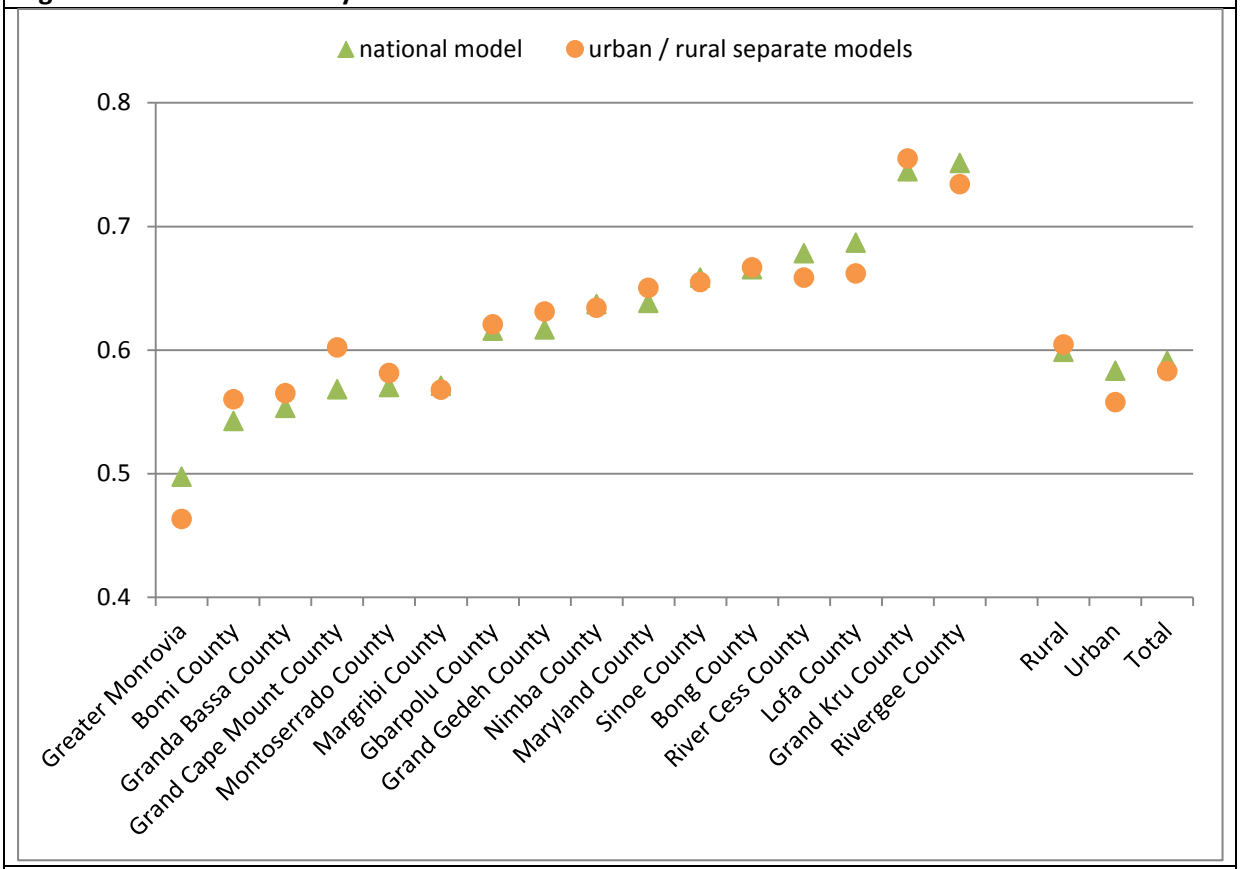


Source: Authors' calculation based on CWIQ 2007

Next, we use results of the 2007 model to estimate poverty outcomes with the 2010 data. Figure 3 shows the comparison between the estimated poverty headcount measures between the national model and the split urban/rural model. Nationally, there is a one percentage point difference between the headcount poverty using the national model and the split model, with the split model

giving the lower estimation. In seven of 16 counties, the estimations are lower with the national model, but the magnitude of the difference is again small. The largest difference was three percentage points in Gbarpolu County, representing less than five percent of the total headcount for that county, and this difference was not statistically significant.

Figure 3: Estimated Poverty Headcount Measures



Source: Authors' calculation based on CWIQ 2007 & CWIQ 2010

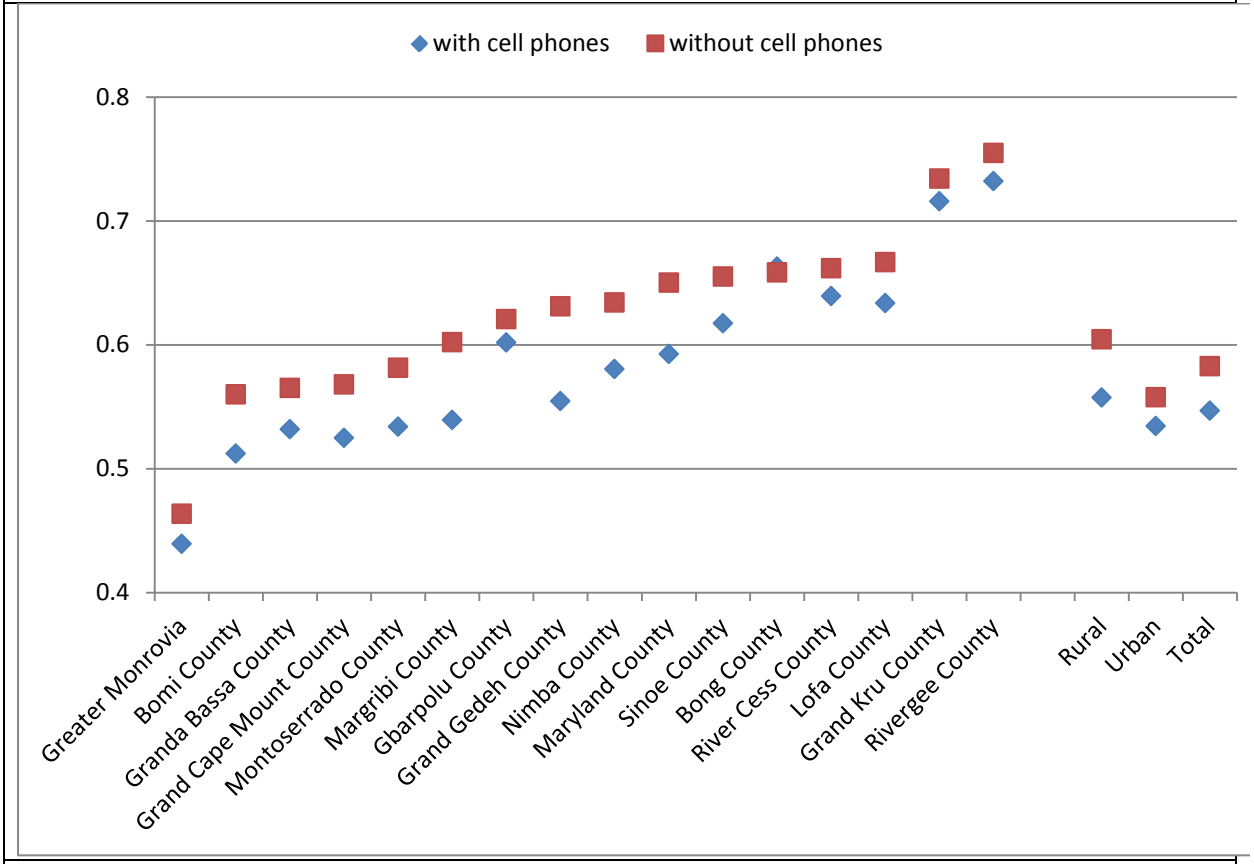
5.1.3 Cellphones

As mentioned in section 5.1.1, a central assumption of the survey-to-survey imputation methodology is the stability of relationship between the included covariates of poverty and poverty status over time. It has, however, recently emerged in the practice of survey-to-survey imputation that certain types of assets introduce substantial instability in the estimates of poverty over time, in particular cell phones and mass produced electronics. When cell phones first arrived in Africa, the handsets were expensive and the geographic coverage limited to a few major cities. Only the top echelon of urban residents would have owned cell phones and therefore ownership would have predicted high levels of household consumption. As the price of the handsets fell and more people could afford cell phones, the relationship between cell phone ownership and poverty would have changed. Ownership would no longer be exclusively limited to the wealthy, and it would be associated with a lower level of consumption (so that the estimate may be biased towards zero, meaning no relationship). By definition, survey-to-survey imputation must hold the relationship unchanged from the base year. Therefore from a base reference year, predictions including cell phones would likely overestimate predicted consumption in later years as the β coefficients may be

too high. This would lead to an underestimation of poverty[7]. Similarly, the opposite would be true for estimates going backwards from a base year. The extent of the bias introduced in the estimates would depend on the degree of change in the coefficients. Since we only have one year with consumption data, there is no further information available as to the stability of the coefficient.

The change in cell phone ownership is large between the two rounds of data collection. In 2007, 30.5 percent of households indicated that at least one member owned a cell phone. By 2010, this figure had risen to 56.6 percent. Outside of Monrovia, the change was even more dramatic. In 2007, 18.4 percent of households had a cell phone. By 2010, it was almost half at 46.6 percent. The inclusion of cellphones as an eligible variable in the model has a substantial effect on the predicted poverty outcomes. The poverty estimates for 2010 including cell phones in the model show a decrease in the headcount poverty number from 64.0 percent in 2007 to 54.7 percent in 2010 using the split urban/rural model. Excluding cell phones from the model, the decrease is approximately one-third smaller, declining to 58.3 percent. In rural areas, due to the large change in ownership, the differences were even greater. Including cell phones in the model showed a decline from 68.2 percent to 55.7 percent, while their exclusion limits that decline to 60.5 percent. Comparing the difference by county, the inclusion of cell phones gives consistently lower results, differences which are in some cases significant. For the reasons listed above cell phones were excluded from eligible variables, though further research is necessary to fully understand these dynamics.

Figure 4: Estimated Poverty Headcount Measures Including & Excluding Cell Phones



Source: Authors' calculation using separate urban / rural models based on CWIQ 2007 & CWIQ 2010

[7] This discussion is true in a bivariate context as an illustration of how changing asset prices and ownership influences the accuracy of imputation methods. How, the empirical relationship could be different in a multivariate estimation.

5.1.4 Rainfall

Since the explanatory power of the imputation model in rural areas was troublingly low, we expanded the analysis to incorporate auxiliary data on agro-climatic conditions. The RFE Rainfall and the eMODIS Normalized Difference Vegetation Index (NDVI) data are collected by the United States Geological Survey and are available through the FEWS NET Africa Data Portal. The RFE rainfall measure publishes county-level statistics for rainfall in 10-day periods since 1995. It should be noted that these are not specific to the household but rather county-level averages. GIS information for individual households is not available. From this measure, we construct three variables: total rainfall in the previous month prior to the survey (as measured in completed 10-day periods), total rainfall in the previous three months, and total rainfall in the previous 12 months. The NDVI is a measure of "greenness" estimated every five days based on satellite remote sensing observations and is available since 2001. From this variable, we construct the current NDVI at the time of the survey, the average NDVI in the three months prior to the survey, and the average NDVI in the 12 months prior to the survey.

Though there is a high correlation between the rainfall and NDVI measures, we include both because rainfall alone interacts with rural well-being in a number of different ways. First there is the obvious link between rainfall and crop production and potential profits from sales, the basis for the rural economy. Since there is a lag between rainfall and crop growth, we included the three-month variable. Also, since there is an even longer lag between rainfall and harvest, we also included the 12-month variable. But since rainfall can also deteriorate roads and infrastructure, making it harder to access markets and services, it can also have a negative effect on well-being. These negative effects can be immediate, in terms of flooding, or cumulative, in the case of deterioration of infrastructure. Since rain can therefore be associated with a positive or negative effect, we include all three measures with no priors on the expected direction of the effects.

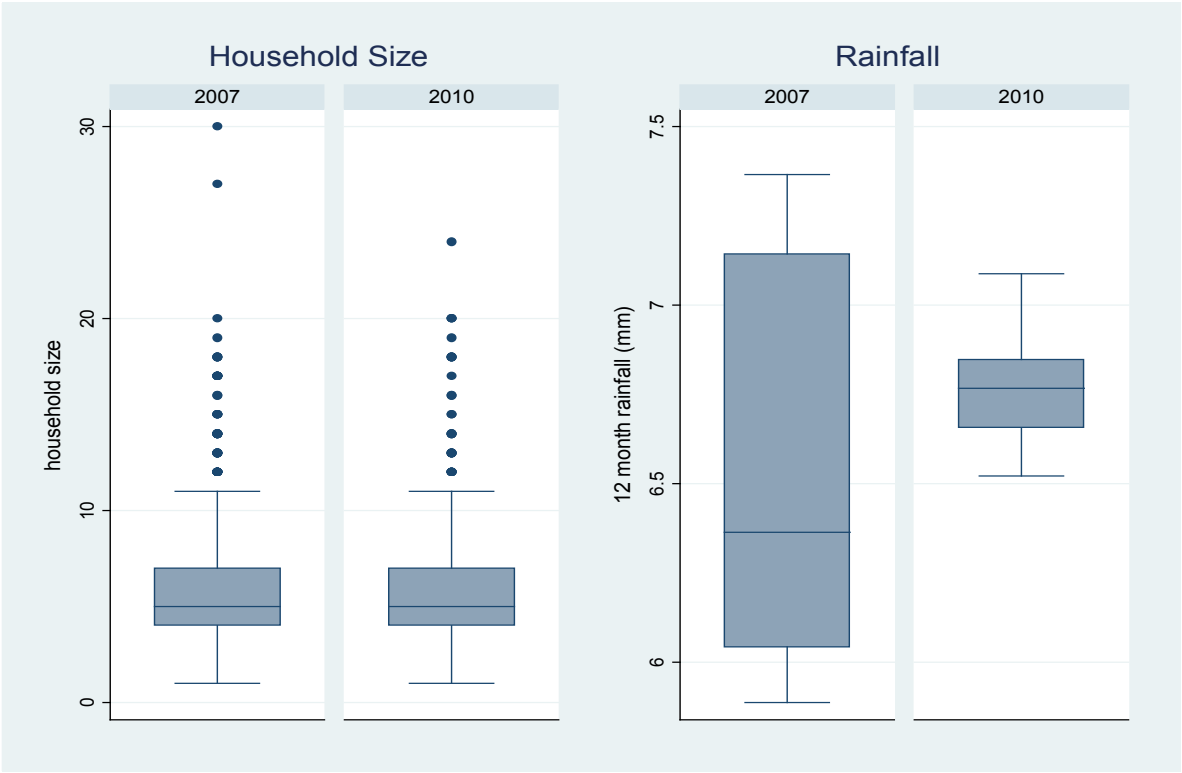
To further try to separate the infrastructure impacts of rain from the agriculture effects, we also include the three measures of NDVI. Higher NDVI measures are indicative of greener ground cover, and generally higher likelihoods of better harvests. It is also sensitive to flooding as too much water would also destroy ground cover. We also initially included the square values of rainfall and NDVI variables, though removed them from the final analysis as they did not add further information.

We re-run the models to see if adding the six rainfall and six NDVI measures change the results or increase the explanatory power of the model. The variables are incorporated into the list of eligible variables and the stepwise analysis is repeated for the expanded list. In the national model, the three-month measure of NDVI, the 12-month measure of NDVI, and the 12-month measure of rainfall are retained. The overall R^2 increased from 0.339 to 0.350, and additional variable explained 4.1 percent of the total variation. The other shares remained nearly unchanged. In the urban model, the three-month measure of NDVI, the three-month rainfall, and the 12-month rainfall variables are retained. Here the R^2 increased from 0.492 to 0.502, and the additional variables

explained 5.9 percent of the total variation. Again the relative contributions from the other shares remained nearly constant. Finally, in the rural areas, the current NDVI, the three-month NDVI, the 12-month NDVI, the three month rainfall, and the 12-month rainfall are retained in the model. The overall explanatory power increased from 0.280 to 0.302. This additional variable explains 9.3 percent of the total variation in the expanded model and the remaining shares are again almost unchanged.

When we attempted to incorporate the rainfall and NDVI data into the 2010 imputation model, the addition was rejected by the multiple imputation package in Stata. In order to include variables in the imputation process, Stata performs a series of tests to ensure that there are no substantial systematic differences between the covariates of the missing and non-missing variables. Permitting some variation is necessary for survey-to-survey imputation to work. For example, in a simple one-variable model with per capita household consumption and years of education of the household head, if variation in the mean were not allowed, poverty could never decrease. The higher mean of education in the second data set (in which the consumption variables are to be imputed) leads to higher mean consumption in the imputed values. Stata will therefore cause the imputation to fail only if the difference in the mean and distribution of the covariate are extreme between the missing and non-missing values. Figure 4 shows box plot results from two continuous variables, household size and 12-month rainfall, found to be significant predictors in the backward selection model. Though there is a statistically significant decrease in household size between the two years, the mean and distribution remain comparable and Stata does not reject the model. In contrast, both the mean and distribution of the 12-month rainfall are vastly different in the two years. Since this would introduce a high degree of instability into the imputations, Stata blocks the execution of the command.

Figure 5: Distribution of household size and 12-month rainfall 2007 and 2010



Source: Authors' calculation based on CWIQ 2007 & CWIQ 2010

5.2 In-Sample Comparisons

Since a comparison data set is not available to test the accuracy of the predictions, the best verification methodology available is an in-sample comparison. To perform this check, a random sub-sample is selected from the overall 2007 data set. Then, using the same model and coefficients as were used to do the previous imputation forward into the 2010 data, the sub-sample data are used to impute the remaining values in the full 2007 data set. We varied the size of the sub-sample from five to 50 percent, testing in five percentage point increments, with the procedure being repeated 100 times for each sub-sample. Figures 6a and 6b show the results for the national model including and excluding cell phones as eligible variables. The shaded area around the imputed estimate represents the imputation component of the total error. For both models, the imputation estimate is well within the confidence interval of the survey estimate, which has a mean of 64.0 percent and with a lower bound of 60.9 percent and an upper bound of 67.1 percent. The estimates including cell phones range from 62.5 and 64.2 percent, and for the model excluding cell phones, the range is between 63.0 and 64.4 percent. Neither model displays substantial bias, indicating that the coefficients of the other variables in the model adjust and compensate for the exclusion of cell phones. Caution should be used, however, before assuming that the in-sample test validates the imputation to different points in time. Recent research from Nguyen and Van der Weide (2014) using evidence from Peru and Malawi shows that while in-sample checks are generally unbiased even for very small sub-samples, bias develops quickly for imputation between years, even when the sub-samples are relatively large.

Figure 6a: Imputation error including cell phones

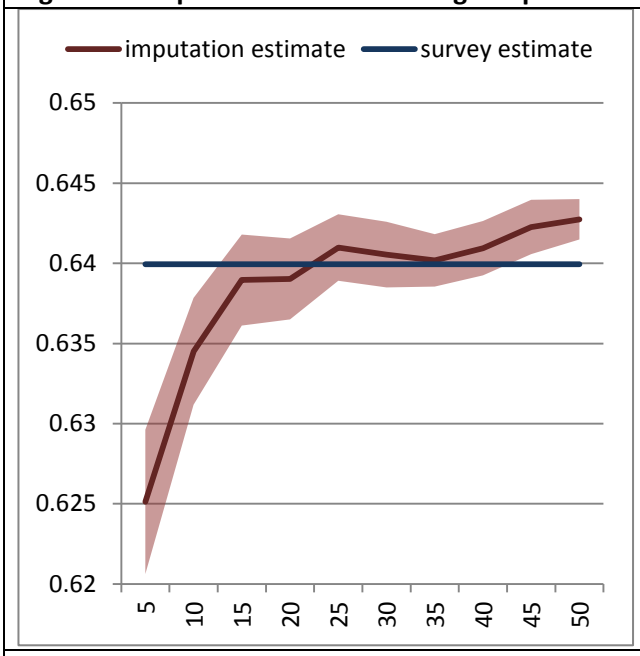
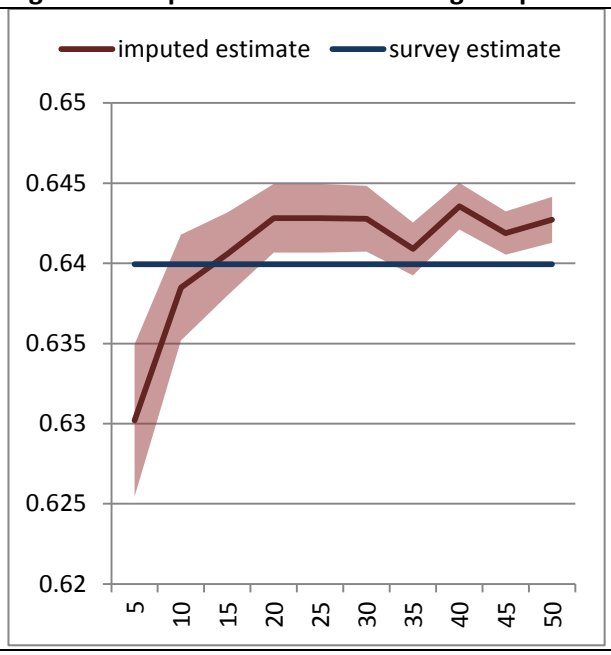


Figure 6b: Imputation error excluding cell phones

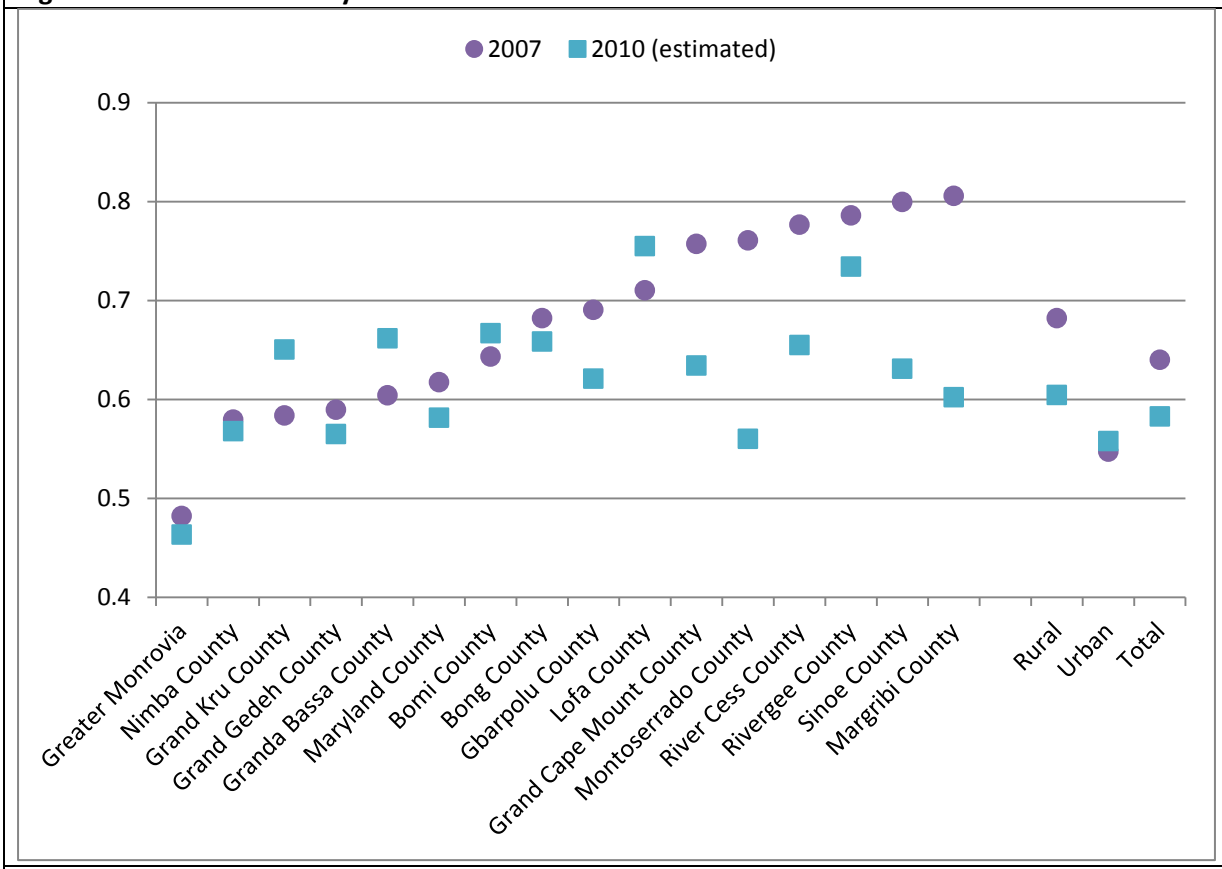


Source: Authors' calculation based on CWIQ 2007

5.3 Overall Poverty Outcomes

Overall, the imputed estimates, using separate models for urban and rural areas predict a 14.5 percent decline in national poverty, from 64.0 percent to 54.7 percent when cell phones are included in the models, but only an 8.9 percent decline, to a headcount of 58.3 percent, when cell phones are excluded. In addition, the estimates at the level of the counties are more vulnerable to the choice of method than the national model. Table 3 in the appendix shows the predicted values across the four combinations. It is difficult, however, to attribute the large predicted changes in poverty at the county level to actual changes. The lower levels of disaggregation is likely to contain a large component of sampling error in addition to the imputation error as the 2007 data set was not designed to be representative at the county level.

Figure 7: Estimated Poverty Headcount Measures



Source: Authors calculation based on CWIQ 2007 & CWIQ 2010

6 Conclusion

As the demand for high frequency quality household data for program design and poverty reduction monitoring will continue to outpace the production, alternative methodologies will continue to grow in importance. This paper examines two common methods used in filling consumption data gaps when estimating poverty: asset-index construction and survey-to-survey imputation. Asset-index construction has the main benefit of being straight forward to construct and not relying on any additional sources of data. As our analysis shows, however, the correlation with household per capita consumption, while positive and significant, is not very high. Alternatively survey-to-survey imputation has the advantage of being based in empirical relationships between consumption and

poverty correlates, but this too has its drawbacks. First, it requires at least one comparable survey that includes consumption. Second, the prediction of the outcome of interest (in this case poverty) could be very sensitive to model specification (e.g. inclusion or exclusion of cell phones). Finally, imputation relies on the assumption of the stability of the relationship between covariates and poverty over time, which can be a strong assumption in rapidly evolving economies where alternative estimation methodologies are the most useful.

In the case of Liberia, survey-to-survey imputation explained a relatively low percentage of the variation, particularly in rural areas. Efforts to improve the rural model by incorporating climate information were unsuccessful in this analysis because they did not substantially add to the explanatory power of the model and additionally varied too much between years for stable imputation. It is clear further research is needed on this issue. Potential areas of future exploration could be comparisons of the stability of the relationship between covariates and poverty over time using multiple cross-sectional waves that include consumption data, improvements to the model using more specific geodata tied to the household/plot location, such as more granular rainfall and NDVI measures, soil quality, slope elevation, and so forth, and finally the incorporation of agriculture specific information to rural-only models, such as crop diversity, input use, and commercialization.

References

Booyen, F., Van der Berg, S., Burger, R., Von Maltitz, M., and Du Rand, G. (2008) Using an Asset Index to Assess Trends in Poverty in Seven Sub-Saharan African Countries. *World Development* 36 (6), 1113-1130.

Brick, J.M. and Kalton G. (1996) Handling Missing Data in Survey Research. *Statistical Methods in Medical Research* (5), 215-238.

Douidich, M., Ezzrari, A., Van der Weide, R. and Verme, P. (2013) Estimating Quarterly Poverty Rates Using Labor Force Surveys. Policy Research Working Paper 6466. Washington DC : World Bank.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003) Micro–level Estimation of Poverty and Inequality. *Econometrica*, 71(1), 355-364.

eMODIS TERRA Normalized Difference Vegetation Index (NDVI) pentadal data and monthly means downloaded from FEWS NET http://earlywarning.usgs.gov/fews/africa/index.php

Filmer, D. and Pritchett, L. (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. (2011). Assessing asset indices. Demography, 49(1), 359-392.

Grosh, M. and Baker, J. (1995) Proxy means tests for targeting social programs: Simulations and speculation. Policy Research Working Paper 118. Washington DC : World Bank.

Grosse, M., Klasen, S. and Spatz, J. (2009) Matching Household with DHS Data to Create national Representative Time Series of Poverty: An Application to Bolivia. Courant Research Centre: Poverty, Equity and Growth – Discussion Paper 21. Courant Research Centre PEG.

Gwatkin, D. R., Rutstein, S., Johnson, K., Pande, R.P., and Wagstaff, A. (2000). Socio-economic differences in health, nutrition and poverty. Washington, DC: World Bank: HNP/Poverty Thematic Group, World Bank.

Howe, L. D., Hargreaves, J. R., Gabrysch, S., and Huttly, S. R. (2009) Is the wealth index a proxy for consumption expenditure? A systematic review. *Journal of Epidemiology and Community Health*, 63(11), 871-877.

NOAA Climate Prediction Center Famine Early Warning System African Rainfall Estimation Algorithm Version 2 (RFE 2.0). Available from http://www.cpc.ncep.noaa.gov/products/fews/rfe.shtml (accessed July 24, 2013).

Nguyen, M.C., Van der Weide, R. (2014) A split questionnaire survey design - an empirical simulation using Peru's continuous household surveys. Mimeo

Rao, J. N. K. (2003) Some new developments in small area estimation. *Journal of Iranian Statistical Society*, 2(2), 145-169.

Sahn, D.E, Stifel, D.C. (2000) Urban-rural Inequality in living standards in Africa. *Journal of African Economies* 12, 564-597.

Tarozzi, A. (2007) Calculating comparable statistics from incomparable surveys, with an application to poverty in India. *Journal of Business and Economic Statistics* 25(3), 314-336.

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement*, 55(4), 525-534.

USGS Earth Resources Observation and Science Center (2012a). eMODIS NDVI Africa (monthly means). Available at: http://earlywarning.usgs.gov/fews/africa/index.php (accessed July 24, 2013).

World Bank (2003) Measuring living standards: Household consumption and wealth indices. Quantitative techniques for household equity analysis – Technical Note 4. Washington DC: World Bank.

World Bank (2012) Liberia Poverty Note: Tracking the Dimensions of Poverty, World Bank Other Operational Studies 12320, The World Bank.

Appendix

Table 1: Mean of Asset Index by Survey & Strata

	DHS	CWIQ	difference	
National	0.0996	-0.0199	-0.1195	
Urban	1.099303	0.956043	-0.1433	*
Rural	-0.47343	-0.47091	0.0025	
Monrovia	1.3191	1.1767	-0.1424	*
North Western	-0.4491	-0.3616	0.0875	
South Central	-0.0751	-0.3572	-0.2821	*
South Eastern A	-0.6420	-0.1492	0.4928	***
South Eastern B	-0.5051	-0.6630	-0.1579	*
North Central	-0.3915	-0.4957	-0.1042	

note: *** p<0.01, ** p<0.05, * p<0.1

Table 2a: Comparison of Population-Weighted Quintiles (National)

		Quintiles of Asset Index					
		Poorest	Poorer	Middle	Rich	Richest	Total
Quintiles of Consumption	Poorest	217	168	109	118	43	655
	Poorer	174	139	126	131	61	631
	Middle	163	157	138	152	103	713
	Rich	137	111	145	159	160	712
	Richest	107	131	119	196	331	884
	Total	798	706	637	756	698	3,595

Pearson chi2(16) = 385.5744 Pr = 0.000
Cramér's V = 0.1637

Table 2b: Comparison of Population-Weighted Quintiles (Rural)

		Quintiles of Asset Index					
		Poorest	Poorer	Middle	Rich	Richest	Total
Quintiles of Consumption	Poorest	125	101	77	44	30	377
	Poorer	110	85	86	69	40	390
	Middle	108	104	99	79	57	447
	Rich	94	86	75	95	95	445
	Richest	86	100	104	97	158	545
	Total	523	476	441	384	380	2,204

Pearson chi2(16) = 140.8365 Pr = 0.000
Cramér's V = 0.1264

Table 2c: Comparison of Population-Weighted Quintiles (Urban)

Quintiles of Consumption	Quintiles of Asset Index					
	Poorest	Poorer	Middle	Rich	Richest	Total
Poorest	165	69	40	14	8	296
Poorer	80	80	56	29	10	255
Middle	80	63	46	43	25	257
Rich	44	53	60	66	46	269
Richest	24	57	55	63	115	314
Total	393	322	257	215	204	1,391

Pearson chi2(16) = 363.1373 Pr = 0.000
Cramér's V = 0.2555

Table 3: Imputed poverty levels under national and split imputation (cell phones included)

	2007 (actual)	"final" model (rural / urban no cell phones)	(national with cell phones)	(national no cell phones)	(rural/urban with cell phones)
Greater Monrovia	0.482	0.464	0.464	0.498	0.439
Bong County	0.643	0.667	0.633	0.665	0.634
Lofa County	0.604	0.662	0.638	0.687	0.640
Nimba County	0.757	0.634	0.593	0.637	0.581
Bomi County	0.761	0.560	0.492	0.543	0.512
Grand Cape Mount County	0.806	0.602	0.527	0.568	0.539
Gbarpolu County	0.691	0.621	0.631	0.616	0.602
Granda Bassa County	0.590	0.565	0.548	0.553	0.532
Margribi County	0.580	0.568	0.545	0.571	0.525
Montoserrado County	0.617	0.581	0.514	0.570	0.534
Grand Gedeh County	0.800	0.631	0.565	0.617	0.555
River Cess County	0.682	0.659	0.694	0.678	0.663
Sinoe County	0.777	0.655	0.611	0.659	0.617
Grand Kru County	0.710	0.755	0.723	0.745	0.732
Maryland County	0.584	0.650	0.591	0.638	0.593
Rivergee County	0.786	0.734	0.717	0.751	0.716
Rural	0.682	0.605	0.564	0.598	0.557
Urban	0.547	0.558	0.549	0.583	0.534
Total	0.640	0.583	0.557	0.591	0.547

Table 4 : Means table with standard errors (2007 & 2010)

		2007		2010	
continuous variables	variable name	mean	se	mean	se
household size	HHSIZE*	5.44	0.039	4.98	0.030
ratio of males in household	femratio	0.50	0.004	0.50	0.004
number dependents 0-14	dep0014	0.39	0.004	0.38	0.003
number dependents 65 and older	dep65up	0.03	0.002	0.04	0.002
number active adults	HH1564	3.02	0.026	2.71	0.020
age of household head	HHAGEY*	43.72	0.230	43.25	0.182
number of rooms in dwelling	ROOMS*	2.77	0.024	2.73	0.021
land size	LANDSIZE*	11.41	0.550	2.35	0.102
number of livestock	MLIVESK	0.70	0.045	0.64	0.038
NDVI at time of survey	ndvi_curr*	0.50	0.001	0.62	0.001
average NDVI 3 months prior to survey	ndvi_3month*	0.52	0.001	0.68	0.000
average NDVI 12 months prior to survey	ndvi_12month*	0.62	0.000	0.59	0.000
rainfall month prior to survey (mm)	deka_month *	72.88	0.615	59.85	0.337
rainfall 3 months prior to survey (mm)	deka_3month*	261.65	3.365	100.56	0.630
rainfall 12 months prior to survey (mm)	deka_12month*	660.70	5.275	837.17	1.651

** Square value (suffix is 2) and cube values (suffix is 3) are also included.*

binomial variables (percentage)					
female headed household	HHSEX	25.34	0.063	24.35	0.051
head ever attended school	HHEVERATTD	56.72	0.070	61.91	0.057
household owns dwelling	OWNHOUSE	67.03	0.066	59.43	0.057
dwelling has electric connection	ELECTCON	0.62	0.011	3.63	0.022
household owns land	LAND	54.24	0.070	38.49	0.057
household has radio	RADIO	49.49	0.071	61.95	0.057
household has cell phone	CPHONE	30.48	0.022	56.62	0.017
household has TV	TV	6.10	0.034	11.32	0.037
household has refrigerator	FRIDGE	1.23	0.016	2.66	0.019
household has stove	STOVE	0.41	0.009	1.04	0.012
household has bicycle	BCYCLE	2.61	0.023	1.41	0.014
household has motorcycle	MCYCLE	1.27	0.016	3.17	0.020
household have vehicle	CAR	1.24	0.016	2.09	0.017
household has sewing machine	SEWMACH	1.42	0.017	3.10	0.020
household has computer	COMPUTER	0.58	0.011	3.57	0.022
household has electric iron	EIRON	0.88	0.013	1.54	0.014
household has VCR/DVD	VCRDVD	5.24	0.032	11.59	0.037
household has boat	BOAT	0.66	0.011	0.85	0.011

categorical variables (percentage)		2007	2010
marital status			
never married	_IHHMARST_1	7.9	9.0
married monogamously	_IHHMARST_2	63.7	56.0
married polygamous	_IHHMARST_3	6.2	5.6
living together	_IHHMARST_4	10.4	17.1
divorced	_IHHMARST_5	3.4	4.8
widowed	_IHHMARST_6	8.4	7.6
literacy of household head			
can read or write	_IHHLITERACY_1	55.3	60.5
cannot read or write	_IHHLITERACY_2	44.7	39.0
cannot be determined	_IHHLITERACY_3	0.0	0.5
education of household head			
no education	_IHHEDLEV_0	41.9	37.7
pre-school	_IHHEDLEV_1	0.3	0.5
primary, not completed	_IHHEDLEV_2	7.5	10.4
complete primary	_IHHEDLEV_3	3.9	4.2
incomplete secondary	_IHHEDLEV_4	20.8	20.2
completed secondary	_IHHEDLEV_5	17.2	20.0
post-secondary technical	_IHHEDLEV_6	5.0	0.7
university and higher	_IHHEDLEV_7	3.6	5.9
other	_IHHEDLEV_9	0.1	0.5
roof			
concrete/cement/brick/stone	_IROOF_1	1.4	4.1
bamboo/thatch	_IROOF_3	31.1	24.7
tiles/shingles	_IROOF_4	0.3	0.1
tin/metal sheets	_IROOF_5	63.3	67.4
other	_IROOF_9	4.0	3.7
walls			
Concrete/cement/brick/stone	_IWALLS_1	24.5	32.4
Wood	_IWALLS_2	0.4	0.6
Iron/metal sheets	_IWALLS_4	3.3	4.4
Clay/mud	_IWALLS_5	70.3	61.1
Other	_IWALLS_9	1.6	1.6
floor			
cement/tiles/marble	_IFLOOR_1	39.9	50.1
wood/bamboo	_IFLOOR_2	0.6	0.5
earth/clay/mud	_IFLOOR_4	59.5	49.2
other	_IFLOOR_9	0.1	0.2
water source			
pipe (own tap)	_IWATER_1	3.8	21.3
public standpipe	_IWATER_2	31.0	19.9
borehole	_IWATER_3	22.1	25.9
wells (protected)	_IWATER_4	16.8	17.8
wells (unprotected)	_IWATER_5	6.0	7.7
surface water	_IWATER_6	16.6	4.2

rain water	_IWATER_7	0.4	1.0
vendor/truck	_IWATER_8	2.5	0.8
other	_IWATER_9	1.0	1.4
cooking fuel			
firewood	_IFUELCOOK_1	66.3	52.8
kerosene	_IFUELCOOK_2	1.0	0.6
charcoal	_IFUELCOOK_3	32.6	44.2
electricity	_IFUELCOOK_4	0.0	0.8
gas	_IFUELCOOK_5	0.0	0.1
other	_IFUELCOOK_9	0.0	1.5
lighting fuel			
electricity	_IFUELLIGH_1	2.7	3.2
kerosene	_IFUELLIGH_2	77.5	49.2
battery/candles	_IFUELLIGH_3	16.9	46.9
gas	_IFUELLIGH_4	0.3	0.5
other	_IFUELLIGH_9	2.6	0.2
toilet facility			
flush toilet	_ITOILET_1	12.4	17.7
improved pit latrine	_ITOILET_2	28.4	31.8
pit latrine	_ITOILET_3	6.1	5.1
no facility	_ITOILET_4	46.8	39.9
other	_ITOILET_9	6.3	5.6
garbage collection			
collected	_IGARBDISP _1	7.4	12.8
buried/burned	_IGARBDISP _2	37.6	49.1
discarded in empty lots, street, rivers	_IGARBDISP _3	33.4	33.6
other	_IGARBDISP _9	21.7	4.6
county			
Greater Monrovia	_IPROVINCE_1	22.3	28.0
Bong County	_IPROVINCE_2	8.9	10.1
Lofa County	_IPROVINCE_3	9.9	7.7
Nimba County	_IPROVINCE_4	16.6	12.5
Bomi County	_IPROVINCE_5	2.8	2.8
Grand Cape Mount County	_IPROVINCE_6	5.3	3.5
Gbarpolu County	_IPROVINCE_7	2.1	2.4
Granda Bassa County	_IPROVINCE_8	7.6	6.5
Margribi County	_IPROVINCE_9	5.4	5.9
Montoserrado County	_IPROVINCE_10	3.2	5.0
Grand Gedeh County	_IPROVINCE_11	3.7	3.4
River Cess County	_IPROVINCE_12	1.6	2.0
Sinoe County	_IPROVINCE_13	3.6	2.9
Grand Kru County	_IPROVINCE_14	1.4	1.6
Maryland County	_IPROVINCE_15	3.5	3.7
Rivergee County	_IPROVINCE_16	2.1	2.1

Code for Imputation using Multiple Imputation Package in Stata

Note: this code assumes that there are no missing values in the variables included in the model. The multiple imputation commands can be used to first impute the missing covariate values and then to perform the imputation of the consumption aggregate. See the Stata Multiple Imputation Reference Manual (Release 13) for further details.

```
local vars   HHSIZE HHSIZE2 HHSIZE3 dep0014 dep65up femratio HH1564 HHSEX HHAGEY
HHAGEY2 HHAGEY3 i.HHMARST i.HHLITERACY HHEVERATTD i.HHEDLEV OWNHOUSE ROOMS ROOMS2
ROOMS3 i.ROOF i.WALLS i.FLOOR i.WATER ELECTCON i.FUELCOOK i.FUELLIGH i.TOILET
i.GARBDISP LAND LANDSIZE LANDSIZE2 LANDSIZE3 RADIO TV CPHONE FRIDGE STOVE BCYCLE
MCYCLE CAR SEWMACH COMPUTER EIRON VCRDVD BOAT MLIVESK

svyset cluster [w=WTA_ADQ], strata(PROVINCE)
preserve
keep if SURVEYR==2007

* stepwise (with GIS)
qui: xi: stepwise, pr(.05): reg lnpcexp `vars' RURURB i.PROVINCE  `ndvi' `deka'
local temp : colfullnames e(b)
local b _cons
local model_nat_gis : list temp - b
di "`model_nat_gis'"
nestreg: svy, subpop(if SURVEYR==2007): reg lnpcexp RURURB (HHSIZE HHSIZE2 HHSIZE3
dep0014 femratio) (HHAGEY HHEVERATTD) (LANDSIZE LANDSIZE2 LANDSIZE3)  (_IWATER_6
_IFUELCOOKa2 _ITOILET_9 _IWATER_5 _IWATER_9 _IFUELCOOKa3 _IFUELCOOKa9 _IHHEDLEVa5
_IFUELLIGH_3 _IFUELLIGH_9 _IWALLS_2 _IWATER_8 _IFLOORa2 _IFLOORa4 _IWATER_4)
(VCRDVD CPHONE RADIO MCYCLE CAR)  (_IPROVINCE_19 _IPROVINCE_8 _IPROVINCE_16
_IPROVINCE_5 _IPROVINCE_21 _IPROVINCE_20 _IPROVINCE_9 _IPROVINCE_3 _IPROVINCE_17
_IPROVINCE_12 _IPROVINCE_15) (ndvi_3month ndvi_12month deka_12month)
restore

* national
preserve
mi set flong
mi register imputed lnpcexp
mi register regular `model_nat_gis'
bys SURVEYR: sum lnpcexp
mi impute regress lnpcexp `model_nat_nogis', add(10) rseed(16071847)
mean lnpcexp [aw=WTA_ADQ], over(SURVEYR)
gen poor=lnpcexp<lnzref
mi estimate: svy: mean poor, over(SURVEYR)
mi estimate: svy: mean poor, over(SURVEYR RURURB)
mi estimate: svy: mean poor, over(SURVEYR PROVINCE)
restore
```

Models

Base National (excluding county, excluding GIS)
RURURB HHSIZE HHSIZE2 HHSIZE3 dep0014 femratio HHAGEY HHEVERATTD _IHHEDLEVa5 LANDSIZE LANDSIZE2 LANDSIZE3 _IFUELCOOKa9 _IFUELLIGH_9 _IWALLS_2 _ITOILET_9 _IWATER_4 _IWATER_8 _IWALLS_5 _IWATER_6 _IFUELCOOKa2 _IGARBDISP_9 _IFUELCOOKa3 _IWATER_5 _IFLOORa2 _IFLOORa4 _IGARBDISP_3 MCYCLE CAR VCRDVD CPHONE RADIO

Base National (with county, excluding GIS)
RURURB HHSIZE HHSIZE2 HHSIZE3 dep0014 femratio HHAGEY HHEVERATTD _IHHEDLEVa5 LANDSIZE LANDSIZE2 LANDSIZE3 _IWATER_9 _IFUELLIGH_3 _IWALLS_2 _IWATER_6 _IFLOORa4 _IWATER_5 _IFLOORa2 _IFUELCOOKa3 _IFUELCOOKa2 _IWATER_8 _IGARBDISP_3 _IFUELCOOKa9 _ITOILET_9 _IWATER_4 _IFUELLIGH_9 CAR MCYCLE VCRDVD RADIO CPHONE _IPROVINCE_3 _IPROVINCE_5 _IPROVINCE_7 _IPROVINCE_8 _IPROVINCE_9 _IPROVINCE_11 _IPROVINCE_12 _IPROVINCE_13 _IPROVINCE_15 _IPROVINCE_16 _IPROVINCE_17 _IPROVINCE_20 _IPROVINCE_21

Rural (excluding GIS)
HHSIZE HHSIZE2 HHSIZE3 femratio HH1564 HHAGEY _IHHEDLEVa2 LANDSIZE LANDSIZE2 _ITOILET_9 _IWALLS_4 _IWATER_2 _IWATER_4 _IWATER_3 _IFUELCOOKa2 _IWATER_9 _IFLOORa4 _IWALLS_5 _IFUELCOOKa3 _IFUELLIGH_9 TV BCYCLE CPHONE MCYCLE COMPUTER RADIO

Urban (excluding GIS)
HHSIZE HHSIZE2 HHSIZE3 dep0014 femratio _IHHMARST_5 _IHHLITERAC_2 LAND LANDSIZE LANDSIZE2 LANDSIZE3 _IFUELCOOKa5 _ITOILET_9 _ITOILET_4 _IFUELCOOKa3 _IWATER_8 _IROOF_9 _IFUELCOOKa9 _IGARBDISP_3 _IWATER_4 _IFLOORa4 _IWATER_9 _IGARBDISP_9 _IWATER_5 _ITOILET_2 _IWALLS_2 CPHONE RADIO VCRDVD SEWMACH CAR BOAT

Base National (excluding county, with GIS)

RURURB HHSIZE HHSIZE2 HHSIZE3 femratio dep0014 HHAGEY HHEVERATTD _IHHEDLEVa5 LANDSIZE LANDSIZE2 LANDSIZE3 _IWALLS_2 _IGARBDISP_9 _IFLOORa4 _IWATER_5 _IFUELLIGH_9 _IFLOORa2 _IWATER_6 _IGARBDISP_3 _IWATER_8 _IFUELCOOKa9 _IFUELCOOKa2 _IFUELCOOKa3 _IWATER_4 _ITOILET_9 CAR MCYCLE VCRDVD CPHONE RADIO deka_12month deka_3month ndvi_3month

Base National (with county & GIS)
RURURB HHSIZE HHSIZE2 HHSIZE3 dep0014 femratio HHAGEY HHEVERATTD LANDSIZE LANDSIZE2 LANDSIZE3 _IWATER_6 _IFUELCOOKa2 _ITOILET_9 _IWATER_5 _IWATER_9 _IFUELCOOKa3 _IFUELCOOKa9 _IHHEDLEVa5 _IFUELLIGH_3 _IFUELLIGH_9 _IWALLS_2 _IWATER_8 _IFLOORa2 _IFLOORa4 _IWATER_4 VCRDVD CPHONE RADIO MCYCLE CAR _IPROVINCE_19 _IPROVINCE_8 _IPROVINCE_16 _IPROVINCE_5 _IPROVINCE_21 _IPROVINCE_20 _IPROVINCE_9 _IPROVINCE_3 _IPROVINCE_17 _IPROVINCE_12 _IPROVINCE_15 ndvi_3month ndvi_12month deka_12month

Rural (with GIS)
HHSIZE HHSIZE2 HHSIZE3 HH1564 femratio HHAGEY _IHHEDLEVa2 LANDSIZE LANDSIZE2 LANDSIZE3 _IFUELLIGH_3 _IWATER_2 _IWATER_9 _IFUELLIGH_2 _ITOILET_9 _IFLOORa4 _IFUELCOOKa2 _IWATER_3 _IWALLS_5 _IWALLS_4 COMPUTER CPHONE MCYCLE RADIO TV ndvi_curr deka_12month ndvi_3month deka_3month ndvi_12month

Urban (with GIS)
HHSIZE HHSIZE2 HHSIZE3 dep0014 femratio HHEVERATTD _IHHMARST_5 LAND LANDSIZE LANDSIZE2 LANDSIZE3 _IFUELCOOKa3 _IGARBDISP_3 _IROOF_9 _IWATER_9 _IWATER_8 _IWALLS_2 _IFUELCOOKa9 _IWATER_4 _IFLOORa4 _IFLOORa2 _IWATER_5 BOAT MCYCLE RADIO CAR VCRDVD CPHONE deka_3month deka_12month ndvi_3month ndvi_curr