Estimating poverty rates using labor force surveys in the OECS countries:

the case of st. lucia

Tanida Arayavechkit and Yuandong Qi

# Introduction

* Poverty estimates are scant in the Caribbean
* The frequency of household survey data collection in most countries is every ten years
* The labor force survey, on the other hand, is collected at a more frequent basis. For example, St Lucia and Grenada collect the LFS on a quarterly basis. St Vincent collect the LFS every two years.
* This paper builds on the existing cross-survey imputation literature to provide more frequent poverty estimates using the data already available. While the methodology is applied to St Lucia, the paper aims to lay the groundwork for other OECS countries.
* Finally, the paper proposes the variables that are not currently collected in the LFS but could contribute to more accurate estimation of poverty rates in the OECS countries.

# Methodology

Poverty rate is one of the most important welfare indicators that policymakers and researchers are interested in. Usually it is obtained by surveying a sample from the population on their incomes and/or consumption, then comparing these numbers to the official poverty line to identify if a household is in poverty or not. Unfortunately, a lot of countries do not collect income or expenditure information on a yearly basis, though it is very desirable to have poverty rate each year.

To overcome the lack of frequent budgetary survey data in St. Lucia, we implement a survey-to-survey imputation technique to project the consumption and consequently the poverty rate from the labor force survey for years when the consumption data is absent.

Specifically, we first develop a consumption model from the 2016 Household Budgetary Survey in St. Lucia. Second, we use the resulting model to impute a distribution of consumption for households in the Labor Force Survey data from 2008 to 2016 (2010 data not available and 2013 data is corrupted). Thirdly, to account for the consumption behavior that the model does not capture, we impute the estimated residuals of the consumption model from the 2016 HBS data into the LFS data and add them to the predicted consumption to obtain the final consumption for each household in the LFS data. As a result, we now can compare the final consumption to the official poverty line to gain the poverty status for each household and consequently the poverty rate for St. Lucia in each year.

The consumption model is a simple OLS model:

where y represents the annual per capita consumption (adult equivalent) of each household; X denotes all the socioeconomic and demographic variables of each household that available in both the HBS and the LFS data, see data section for more details.

There are two ways to impute residuals from the consumption model into the LFS model. One is to randomly allocate residuals to the LFS without considering the location of the households in each survey. Another way is to randomly allocate residuals within each parish. For both methods, we bootstrap the error terms 300 times to obtain the standard errors. We impute the poverty rates with both ways of assigning error terms to the LFS data.

Note that to have the method described above to deliver reliable estimates of poverty rates, we need to assume the model doesn’t change over time, e.g.: the contribution of gender on consumption in 2016 is the same as in 2008. In addition, the data without consumption values must have the same set of variables in the data with consumption so that the prediction of consumption is unbiased. Last but not least, the covariates used in the model should be defined and measured in a consistent way across surveys, otherwise, the estimated poverty rates would be incorrect.

We can validate our estimates of poverty rates by comparing the poverty rate imputed from the LFS data with the one directly calculated from the HBS data to see the closeness between the two numbers. We are able to achieve this goal for the year of 2016 as we have both LFS and HBS for that year. It would be very desirable to have the assumption of time-invariant model be checked with the help of both LFS and HBS data from the beginning (2008) and the end (2016) of the period of interest. If the poverty rate obtained from imputing consumption into LFS of beginning year using the model derived from the HBS data of the end year is sufficiently close to the poverty rate obtained from the opposite process, we would have extra confidence on our method. But due to data limitation, we are only able to conduct the first type of validation.

# Data

There are two main data sets that we use to impute poverty rates. The 2016 Household Budgetary Survey (HBS) data set which contains consumption is used to construct the consumption model for later imputation. The 2008-2016 Labor force Survey data provides the same set of variables as in the HBS data except consumption.

The 2016 HBS data covers 1440 households, the 2008, 2009, 2011, 2012, 2014, 2015 and 2016 LFS data covers 3283, 2816, 1438, 2816, 3147, 2508 and 3183 households. Note we only use observations without any of the variables used in the model being missing. Though missing values in some explanatory variables may bias the consumption model, we end up with only about 10% observations being lost so we should be fine in this regard.

Table 1 provides the summary statistics for the common set of variables present in both the 2016 Household Budgetary Survey and 2008-2016 Labor force Survey. The key variable is the consumption which is measured as the per capita consumption of each household per year deflated by 2011 CPI (adult equivalent).

The explanatory variables include the following:

Sociodemographic variables: gender, age group, education level (no school, primary school, secondary school, high school, and tertiary education), marital status (married or not), household size and its square, and the number of members at each educational level (including the head);

Labor market variables: employment sector, employment status (unemployed, employed and out of labor force), occupation type (self-employed, employer, public employee, and private employee including unpaid workers), hours of work per month, and the number of members at each occupation type (including the head);

Household asset variables: if the household own home, wash machine, refrigerator, phone, number of bedrooms, TV, computer and car;

Location variable: dummies for all ten parishes of St. Lucia.

One of the key assumptions for the success of the imputation is that the variables across surveys are consistent in terms of definition and sampling design. Fortunately, the HBS and LFS are largely comparable. (Do we need to provide some numeric evidence for this?)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1: Summary Statistics** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | HBS 2016 |  | LFS 2016 |  | LFS 2015 |  | LFS 2014 |  | LFS2012 |  | LFS2011 |  | LFS2009 |  | LFS2008 |  |
|  | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D. |
| log per capita expenditure  (adult equivalent in 2011 dollars) | 9.301 | 0.777 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| sex | 0.585 | 0.493 | 0.598 | 0.490 | 0.588 | 0.492 | 0.592 | 0.492 | 0.604 | 0.489 | 0.599 | 0.490 | 0.561 | 0.496 | 0.551 | 0.498 |
| age: 15-24 | 0.017 | 0.131 | 0.020 | 0.138 | 0.020 | 0.141 | 0.025 | 0.157 | 0.034 | 0.181 | 0.027 | 0.161 | 0.027 | 0.161 | 0.029 | 0.167 |
| age: 25-40 | 0.224 | 0.417 | 0.182 | 0.386 | 0.190 | 0.392 | 0.210 | 0.408 | 0.243 | 0.429 | 0.270 | 0.444 | 0.227 | 0.419 | 0.245 | 0.430 |
| age: 41-64 | 0.530 | 0.499 | 0.566 | 0.496 | 0.552 | 0.497 | 0.532 | 0.499 | 0.524 | 0.500 | 0.509 | 0.500 | 0.514 | 0.500 | 0.516 | 0.500 |
| age: 65+ | 0.228 | 0.420 | 0.233 | 0.423 | 0.238 | 0.426 | 0.232 | 0.422 | 0.199 | 0.399 | 0.194 | 0.396 | 0.232 | 0.422 | 0.210 | 0.407 |
| primary school | 0.378 | 0.485 | 0.050 | 0.217 | 0.057 | 0.231 | 0.054 | 0.226 | 0.051 | 0.221 | 0.053 | 0.223 | 0.569 | 0.495 | 0.568 | 0.495 |
| secondary school | 0.388 | 0.488 | 0.217 | 0.412 | 0.206 | 0.405 | 0.226 | 0.418 | 0.222 | 0.416 | 0.245 | 0.430 | 0.171 | 0.376 | 0.178 | 0.382 |
| tertiary education | 0.151 | 0.359 | 0.081 | 0.273 | 0.089 | 0.285 | 0.078 | 0.268 | 0.074 | 0.262 | 0.074 | 0.262 | 0.099 | 0.298 | 0.101 | 0.302 |
| employer | 0.172 | 0.377 | 0.174 | 0.379 | 0.161 | 0.367 | 0.162 | 0.368 | 0.145 | 0.352 | 0.153 | 0.360 | 0.171 | 0.377 | 0.177 | 0.382 |
| public employee | 0.049 | 0.215 | 0.060 | 0.237 | 0.057 | 0.231 | 0.056 | 0.230 | 0.051 | 0.220 | 0.056 | 0.231 | 0.017 | 0.130 | 0.038 | 0.191 |
| private employed | 0.110 | 0.314 | 0.091 | 0.288 | 0.114 | 0.318 | 0.137 | 0.343 | 0.115 | 0.319 | 0.115 | 0.319 | 0.103 | 0.304 | 0.112 | 0.315 |
| unemployed | 0.078 | 0.268 | 0.090 | 0.286 | 0.106 | 0.308 | 0.090 | 0.287 | 0.102 | 0.303 | 0.097 | 0.296 | 0.079 | 0.269 | 0.065 | 0.247 |
| out of labor force | 0.278 | 0.448 | 0.278 | 0.448 | 0.303 | 0.460 | 0.299 | 0.458 | 0.276 | 0.447 | 0.290 | 0.454 | 0.324 | 0.468 | 0.310 | 0.463 |
| Hours of work per month | 2.278 | 1.788 | 2.313 | 1.780 | 2.158 | 1.816 | 2.235 | 1.806 | 1.520 | 1.823 | 1.191 | 1.741 | 2.188 | 1.826 | 2.225 | 1.815 |
| parish\_2 | 0.210 | 0.408 | 0.334 | 0.472 | 0.382 | 0.486 | 0.359 | 0.480 | 0.334 | 0.472 | 0.108 | 0.311 | 0.344 | 0.475 | 0.348 | 0.476 |
| parish\_3 | 0.074 | 0.261 | 0.053 | 0.224 | 0.053 | 0.224 | 0.044 | 0.206 | 0.047 | 0.212 | 0.245 | 0.430 | 0.047 | 0.212 | 0.063 | 0.243 |
| parish\_4 | 0.076 | 0.265 | 0.058 | 0.233 | 0.050 | 0.217 | 0.065 | 0.247 | 0.056 | 0.230 | 0.061 | 0.239 | 0.050 | 0.218 | 0.061 | 0.239 |
| parish\_5 | 0.056 | 0.230 | 0.040 | 0.195 | 0.036 | 0.187 | 0.044 | 0.205 | 0.043 | 0.204 | 0.049 | 0.216 | 0.028 | 0.166 | 0.030 | 0.171 |
| parish\_6 | 0.063 | 0.242 | 0.042 | 0.200 | 0.036 | 0.187 | 0.043 | 0.204 | 0.041 | 0.199 | 0.046 | 0.209 | 0.042 | 0.201 | 0.033 | 0.179 |
| parish\_7 | 0.101 | 0.301 | 0.107 | 0.309 | 0.096 | 0.294 | 0.061 | 0.240 | 0.105 | 0.306 | 0.100 | 0.300 | 0.091 | 0.287 | 0.083 | 0.276 |
| parish\_8 | 0.101 | 0.301 | 0.121 | 0.326 | 0.115 | 0.319 | 0.111 | 0.315 | 0.115 | 0.319 | 0.124 | 0.329 | 0.104 | 0.305 | 0.087 | 0.281 |
| parish\_9 | 0.083 | 0.276 | 0.082 | 0.274 | 0.071 | 0.257 | 0.094 | 0.292 | 0.073 | 0.260 | 0.076 | 0.266 | 0.072 | 0.259 | 0.063 | 0.243 |
| parish\_10 | 0.126 | 0.332 | 0.147 | 0.354 | 0.138 | 0.345 | 0.163 | 0.370 | 0.147 | 0.355 | 0.120 | 0.325 | 0.137 | 0.344 | 0.127 | 0.333 |
| N. of unemployed | 0.301 | 0.608 | 0.315 | 0.628 | 0.390 | 0.687 | 0.335 | 0.662 | 0.331 | 0.659 | 0.280 | 0.573 | 0.182 | 0.422 | 0.178 | 0.414 |
| N. of employers | 0.069 | 0.272 | 0.079 | 0.295 | 0.075 | 0.284 | 0.078 | 0.297 | 0.070 | 0.283 | 0.078 | 0.299 | 0.020 | 0.139 | 0.048 | 0.222 |
| N. of public employees | 0.242 | 0.519 | 0.209 | 0.511 | 0.237 | 0.516 | 0.266 | 0.563 | 0.225 | 0.506 | 0.216 | 0.498 | 0.181 | 0.408 | 0.202 | 0.439 |
| N. of private employee | 0.745 | 0.855 | 0.675 | 0.822 | 0.642 | 0.834 | 0.588 | 0.799 | 0.679 | 0.836 | 0.628 | 0.795 | 0.561 | 0.696 | 0.599 | 0.713 |
| N. of secondary education | 1.347 | 1.397 | 0.840 | 1.057 | 0.868 | 1.083 | 0.831 | 1.031 | 0.816 | 1.017 | 0.811 | 0.959 | 0.597 | 0.769 | 0.576 | 0.734 |
| N. of tertiary education | 0.421 | 0.765 | 0.173 | 0.475 | 0.213 | 0.569 | 0.184 | 0.523 | 0.184 | 0.539 | 0.180 | 0.496 | 0.256 | 0.539 | 0.255 | 0.496 |
| Household size | 3.078 | 2.001 | 2.581 | 1.727 | 2.763 | 1.777 | 2.765 | 1.742 | 2.918 | 1.918 | 2.832 | 1.714 | 2.980 | 1.363 | 3.038 | 1.127 |
| Household size squared | 13.472 | 20.450 | 9.643 | 13.954 | 10.794 | 14.476 | 10.681 | 13.511 | 12.192 | 17.254 | 10.953 | 13.197 | 10.736 | 9.956 | 10.498 | 7.441 |
| wash machine | 0.548 | 0.498 | 0.592 | 0.491 | 0.617 | 0.486 | 0.591 | 0.492 | 0.607 | 0.489 | 0.587 | 0.493 | 0.531 | 0.499 | 0.518 | 0.500 |
| fridge | 0.808 | 0.394 | 0.900 | 0.300 | 0.894 | 0.308 | 0.891 | 0.312 | 0.889 | 0.315 | 0.892 | 0.311 | 0.883 | 0.322 | 0.849 | 0.358 |
| telephone | 0.910 | 0.286 | 0.838 | 0.368 | 0.875 | 0.331 | 0.866 | 0.341 | 0.850 | 0.357 | 0.843 | 0.364 | 0.815 | 0.388 | 0.792 | 0.406 |
| tv | 0.813 | 0.390 | 0.926 | 0.262 | 0.926 | 0.263 | 0.918 | 0.275 | 0.925 | 0.263 | 0.927 | 0.260 | 0.920 | 0.271 | 0.895 | 0.307 |
| N. of bedroom | 2.380 | 1.036 | 2.367 | 0.960 | 2.433 | 0.998 | 2.403 | 1.027 | 2.462 | 1.067 | 2.428 | 1.043 | 2.496 | 1.014 | 2.386 | 1.137 |
| car | 0.226 | 0.419 | 0.267 | 0.443 | 0.286 | 0.452 | 0.296 | 0.457 | 0.291 | 0.454 | 0.281 | 0.450 | 0.265 | 0.441 | 0.255 | 0.436 |
| computer | 0.405 | 0.491 | 0.382 | 0.486 | 0.384 | 0.486 | 0.381 | 0.486 | 0.366 | 0.482 | 0.334 | 0.472 | 0.280 | 0.449 | 0.264 | 0.441 |
| own home | 0.734 | 0.442 | 0.768 | 0.422 | 0.782 | 0.413 | 0.766 | 0.423 | 0.759 | 0.428 | 0.762 | 0.426 | 0.767 | 0.423 | 0.750 | 0.433 |

# Empirical model and results

To empirically impute poverty rates for years that consumption is not available, we build a consumption model using the set of explanatory variables common to both surveys.

Table 2 shows the consumption model. All the explanatory variables have the expected signs and significance. To name a few: having a male household head increases consumption; age does not have a significant effect on consumption (the base is the age group between 0 and 14); education is positively associated with consumption as expected; the more household members being employed, the more consumption a household will have regardless of employment type (private or public employee); household size has negative impact on per capita consumption; owning assets is positively and significantly associated with consumption. The R-square is 0.63 which indicates that our model can explain 63% of the variation of the consumption variable.

|  |  |
| --- | --- |
| **Table 2: Consumption Model From 2016 HBS** | |
| Dependent Variable: log per capita consumption (adult equivalent) | |
| sex | 0.06\* |
|  | (0.03) |
| age: 15-24 | 0.07 |
|  | (0.09) |
| age: 25-40 | 0.06 |
|  | (0.09) |
| age: 41-64 | 0.18\* |
|  | (0.10) |
| primary school | 0.15\*\* |
|  | (0.06) |
| secondary school | 0.28\*\*\* |
|  | (0.07) |
| tertiary education | 0.41\*\*\* |
|  | (0.09) |
| employed | 0.23\*\* |
|  | (0.10) |
| out of labor force | 0.22\*\* |
|  | (0.10) |
| Hours of work per month | 0.06\*\*\* |
|  | (0.02) |
| N. of unemployed members | -0.05 |
|  | (0.04) |
| N. of employers | 0.19\*\*\* |
|  | (0.06) |
| N. of private employee members | 0.10\*\*\* |
|  | (0.04) |
| N. of public employee members | 0.10\*\*\* |
|  | (0.02) |
| N. of member with secondary education | 0.01 |
|  | (0.01) |
| N. of member with tertiary education | 0.08\*\*\* |
|  | (0.03) |
| Household size | -0.36\*\*\* |
|  | (0.02) |
| household size squared | 0.02\*\*\* |
|  | (0.00) |
| own home | 0.02 |
|  | (0.04) |
| wash machine | 0.15\*\*\* |
|  | (0.04) |
| fridge | 0.16\*\*\* |
|  | (0.05) |
| telephone | 0.45\*\*\* |
|  | (0.06) |
| tv | 0.17\*\*\* |
|  | (0.04) |
| N. of bedrooms | 0.09\*\*\* |
|  | (0.02) |
| car | 0.32\*\*\* |
|  | (0.04) |
| computer | 0.09\*\* |
|  | (0.04) |
| sector dummy | Yes |
| occupation dummy | Yes |
| parish dummy | Yes |
| Constant | 8.29\*\*\* |
|  | (0.17) |
| Observations | 1,440 |
| R-squared | 0.63 |
| Standard errors in parentheses |  |
| \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 |  |

To account for the variation of real consumption beyond the estimated consumption model, we draw the residuals nonparametrically from the HBS data estimation and add them to the predicted consumption in the LFS data, then we impute poverty rates by comparing the final consumption (residual + predicted consumption) with the official poverty line. As mentioned before, we first randomly assign residuals to households, later we draw residuals from each parish in the HBS and assign them to households in the same parish in the LFS. The results are similar using these two residual assignment methods. Here we only present the results with clustered residual assignment. To gain robust standard errors for the imputed poverty rates, we bootstrap 300 times the process of random drawing and assigning residuals from the previous HBS consumption model to the LFS data.

Table 3 shows that the real poverty rate and the imputed poverty rates. From 2008 to 2012, poverty rate kept increasing steadily from 31.5% and reached a very high point of 36.6% in 2012. Since then it started to decrease. In 2016 it became 28.6%. Since we have both HBS data and LFS data for 2016, we are able to compare the poverty rate directly calculated with HBS data to the poverty rate imputed from the consumption model and LFS data to assess the accuracy of the method. It turns out that the imputed poverty rate is about 1.7 percentage point higher than the poverty rate from the direct calculation and our method slightly over estimate the poverty rate n St. Lucia. With better data, such as more explanatory variables in both data sets, we can further improve our poverty rate imputation.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 3: Imputed Poverty Rates for St. Lucia** | | | |
|  | Direct calculation | mean | se |
| HBS 2016 | 0.269(0.018) | 0.286 | 0.013 |
| LFS 2016 | NA | 0.279 | 0.008 |
| LFS 2015 | NA | 0.307 | 0.008 |
| LFS 2014 | NA | 0.314 | 0.009 |
| LFS 2012 | NA | 0.366 | 0.010 |
| LFS 2011 | NA | 0.340 | 0.011 |
| LFS 2009 | NA | 0.325 | 0.008 |
| LFS 2008 | NA | 0.315 | 0.006 |

# Recommendations

# Conclusion

In this paper, we implement a survey-to-survey imputation method to impute poverty rate for St. Lucia from 2008 and 2016. Our contribution is twofold. First, we successfully fill the gap of poverty rates in St. Lucia for the time period that consumption data is absent. Second, our estimation lays the foundation for measuring poverty rates and other welfare indicators in other OECS countries where we face similar issue of data availability.