

How Survey-to-Survey Imputation Can Fail

D. Newhouse
S. Shivakumaran
S. Takamatsu
N. Yoshida



WORLD BANK GROUP

South Asia Region

Economic Policy and Poverty Unit

&

Poverty Reduction and Economic Management Network

Poverty Reduction and Equity Unit

July 2014

Abstract

This paper proposes diagnostics to assess the accuracy of survey-to-survey imputation methods and applies them to examine why imputing from the Household Income and Expenditure Survey into the Labor Force Survey fails to accurately project poverty trends in Sri Lanka between 2006 and 2009. Survey-to-survey imputation methods rely on two key assumptions: (i) that the questions in the two surveys are asked in a consistent way and (ii) that common variables of the two surveys explain a large share of the intertemporal change in household expenditure and poverty. In addition, differences in sampling design can lead validation tests to underestimate the accuracy of survey-to-survey predictions. In Sri Lanka, the causes of failure

differ across sectors. In the urban sector, the primary culprit is differences between the two surveys in the design of the questionnaire. In the rural and estate sectors, the set of common variables in the prediction model does not adequately capture changes in poverty. The paper concludes that in Sri Lanka, survey-to-survey imputation between the Household Income and Expenditure Survey and the Labor Force Survey cannot produce accurate poverty estimates unless the Labor Force Survey adds additional questions on assets and is redesigned to use a questionnaire that is compatible with the Household Income and Expenditure Survey. Alternatively, a new welfare-tracking survey that satisfies these conditions could be established.

This paper is a joint product of the Economic Policy and Poverty Unit, South Asia Region; and Poverty Reduction and Equity Unit, Poverty Reduction and Economic Management Network. 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 dnewhouse@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

How Survey-to-Survey Imputation Can Fail

D. Newhouse, S. Shivakumaran, S. Takamatsu, and N. Yoshida

JEL codes: O10, D30, C53

Keywords: Poverty, Prediction models, Sri Lanka

Introduction

This paper examines whether survey-to-survey imputation (SSI) methods produce reliable poverty estimates in Sri Lanka.¹ SSI methods aim to predict changes in the distribution of household per capita expenditure, and therefore poverty rates, using information contained in other surveys that are conducted more frequently. In Sri Lanka, a Labor Force Survey (LFS) is fielded quarterly, while the Household Income and Expenditure Survey (HIES) is carried out once every three years. Therefore, survey-to-survey imputation has the potential to generate much more frequent estimates using existing HIES and LFS data, at negligible additional cost. Doudich et al. (2013) applied this approach to Moroccan data, and found that projected poverty rates based on the LFS data closely matched official poverty rates estimated from actual consumption data.

Two key assumptions must be satisfied for SSI to produce reliable poverty estimates. First, the two surveys must ask the same questions. If the common variables in the two surveys measure different versions of a concept, then the coefficients estimated using the expenditure survey will be inconsistent with the poverty proxies taken from the LFS data. Second, the prediction model must be fixed over time. For this latter assumption to hold in a strict sense, changes in the variables present in both surveys must explain all of the average change in household expenditures. The weaker the model's ability to explain changes in the household consumption distribution, the greater the bias in predicted changes in poverty that will result from assuming that the model's coefficients are fixed. In Sri Lanka, the common set of variables in both surveys excludes key factors that track changes in consumption in rural areas, such as accurate measures of local wages and profits, housing characteristics, and assets. Therefore, the model's predictions severely underestimate the extent of poverty reduction in the rural and estate sectors. For SSI to successfully estimate changes in poverty, the common set of variables used in the model must not only be strong and stable predictors of consumption in a cross-section, but also need to explain a large share of the intertemporal variation in expenditure.

We use the 2006 and 2009 rounds of the HIES and LFS surveys to test whether SSI can explain changes in observed headcount poverty rates, following Doudich et al. (2013).² First we estimate prediction formulas using the 2006 HIES data, and apply those formulas to data from the 2009 LFS to estimate household expenditure and poverty in 2009. Then, we check how close the projected poverty rates are to actual poverty rates computed from the household expenditure data collected in the 2009 HIES. We also repeat this exercise in the reverse direction, by estimating formulas using the 2009 HIES data, and using them to impute household expenditures into the 2006 LFS data. The projected poverty rates of 2006 are then compared with those estimated directly from household expenditures in the 2006 HIES.

In Sri Lanka, we find major discrepancies between the predicted estimates of poverty and actual poverty rates. In particular, predicted poverty substantially underestimates the decline in poverty in the rural and estate sectors, which contain about 85 percent of the country's population. In the rural sector, when imputing backwards to obtain estimates of poverty in 2006, the difference between the projected and actual poverty rates is 5 percentage points, and the model predicted less than a quarter of the actual change in poverty. This finding stands in stark contrast to the Moroccan analysis, which showed that the models predicted changes in poverty well. In Morocco, unlike in Sri Lanka, the poverty rates projected by

¹ Two examples of survey-to-survey imputation methods applied to projecting poverty in developing countries are Stifel and Christiaensen (2007) and Christiaensen et al. (2012)

² HIES surveys were carried out in 2006/7 and 2009/10, not in 2006 and 2009. As a result, these surveys are called the 2006/7 and 2009/10 HIES surveys. However, to match the survey periods between the LFS surveys and the HIES surveys, we use only the common period of both rounds of HIES and LFS surveys, July to December in 2006 and 2009. As a result, when we are running the SSI between HIES and LFS, we call HIES surveys the 2006 and 2009 HIES surveys. However, later when we use all 12 months data of HIES surveys, we call the surveys the 2006/7 and 2009/10 HIES surveys.

SSI were only about 1 percentage point different from those from actual household expenditure data, and the difference is statistically insignificant at the national and regional levels.

We investigate why predicted poverty in Sri Lanka fails to track estimated poverty rates. In urban areas, the main problem is incompatibility between the HIES and the LFS in terms of the surveys' questionnaire. Differences in the employment question between the LFS and HIES leads to biased predictions in the LFS. The difference in the two surveys' sampling strategy, on the other hand, does not cause bias, but it does increase the discrepancy between predicted and directly estimated poverty rates. To see the effect of the differences in questionnaire and sample design, we impute poverty based on welfare indicators and sampling weights taken from the HIES instead of the LFS, using only variables that are common to both surveys. When the predictor variables and the sampling weights both come from the HIES, estimated poverty rates correspond closely to those directly estimated from household expenditures. This confirms that the main problem with the use of SSI in urban areas is differences between the HIES and LFS in survey design.

In rural areas and the estate sector, however, differences in the values of the common variables and sampling weights between the LFS and HIES are negligible. Instead, the errors in the predictions arise mainly because the common variables fail to capture intertemporal variation in consumption. The set of common variables used in the model mainly consists of pre-determined variables such as education and location that, while highly correlated with household consumption in the cross-section, vary little over time. Furthermore, the labor market outcomes included in the model, such as the households' source of income and the head's sector of work, do not capture welfare changes in the rural and estate sector as well as they do in urban areas. Therefore, these common variables explain little of the observed changes in consumption, and predicted consumption greatly understates the increase in actual consumption.

In rural areas, which account for roughly 80 percent of the population, including additional variables largely fixes the problem. In particular, when imputing within the HIES, adding ownership of consumer durables and housing conditions to the model generates large improvements in performance. This is because as households become better off, they purchase assets and more expensive housing materials, meaning that these variables better reflect changes in consumption. In the estate sector as well, adding more variables improves the accuracy of SSI, but the projected poverty rates remain far from those estimated from actual household expenditures.

To generate accurate and frequent poverty estimates in Sri Lanka based on alternative survey data, either the design of the LFS needs to be revised, or a new welfare proxy monitoring survey should be created. Although it is typically cost-effective to use existing data to estimate poverty, ensuring that the LFS uses sampling and questionnaires that are comparable to the HIES may be a challenging endeavor, and would introduce a structural break in the measurement of labor market trends. It may be possible in the Sri Lankan context to create a new welfare proxy monitoring survey, which would collect only those variables needed to project household expenditure and poverty accurately. Recent technological advances, such as rapid improvements in computer assisted personal interviews using tablet devices, have lowered the cost of conducting such a survey.

This paper consists of seven sections. Section II reviews the literature, Section III describes the data, and Section IV explains the imputation method developed by Elbers, Lanjouw and Lanjouw (2003). Section V describes the results of the HIES to LFS imputation and Section VI conducts analyses to investigate the sources of biases in SSI. Section VII concludes.

I. Literature review

Collecting detailed consumption or income data is very costly and time-consuming, and SSI can be a cost-effective solution to improve the frequency, timeliness and comparability of poverty data in developing countries. As discussed above, the idea of SSI is simple: suppose consumption or income data are not collected in a particular year, but non-consumption data from other representative surveys are available. Then consumption or income data can be imputed into the non-consumption data sets, using an imputation model calibrated from a cross-section collected in an earlier time period. Since this method relies on data that are already available, there is no additional data collection cost.³

Projections are needed in this case because household consumption is not observed in the labor force survey, and projections by necessity must be developed from existing data.⁴ The key challenges for SSI are that the two types of surveys must be designed in a similar way, and the model parameters must not change over time. The latter assumption satisfies the requirement to assume a value for the unobserved coefficients in the later period, but it is a strong assumption that is impossible to test in the absence of consumption data. Rather than abandoning the use of projections altogether, however, it is more constructive to conduct several tests using past rounds of consumption data, to check how reliable projection models are in practice.

SSI is now a well-established approach for estimating poverty that is growing in popularity. The first case in which SSI played a key role for poverty estimation can be traced to India, and the work of Deaton and Dreze (2002) and Kijima and Lanjouw (2003). The India National Sample Survey Organization (NSSO) made a slight change in the recall period in part of the food consumption module in the 1999-2000 NSSO survey, and it was widely argued that the resulting data likely overestimated household consumption expenditure and therefore underestimated poverty. To project total household expenditures as if the recall periods for food modules were defined consistently over time, Deaton and Dreze (2002) applied a SSI approach. They first created a model to project from the unchanged part of the consumption module to total household expenditures using a previous round of the NSSO survey. This model was then used to project total household expenditures in 1999-2000 from expenditures in the unchanged part of the food module. Kijima and Lanjouw (2003) applied the same approach, but included only non-consumption data in the model. Whether using a portion of consumption or non-consumption characteristics leads to more accurate predictions remains open to debate. Subsequent analysis using consumption data from the thin-round household surveys, which are collected annually but are not used to estimate official poverty rates, suggests that the approach adopted by Kijima and Lanjouw (2003) produced more plausible results. However, no further test to examine the reliability of either approach has been conducted.

While the Indian analyses imputed household consumption within the same survey series, they raised the prospect that these methods could be used to impute consumption across multiple surveys. Stifel and Christiaensen (2007) is an early attempt to do this, by applying consumption models based on the 1997 Kenya Welfare Monitoring Survey (WMS) to three consecutive rounds of Demographic Health Survey (DHS) between 1993 and 2003. The 1997 WMS collects both consumption and non-consumption data while the DHS contains only non-consumption data. The paper also provides theoretical guidance regarding the choice of variables to be included when imputing consumption, so as to maintain comparability and reliability of the poverty predictions. It recommends including variables that change over time, but excluding variables whose rates of return are likely to change markedly in the face of evolving economic conditions. This argument makes sense in theory, but it is difficult ex-ante to identify

³Ravallion (1996) is one of the early advocates of this approach. He experimented with how well household consumption expenditure can be estimated from information on housing conditions and subjective well-being, which can be collected quickly and cheaply.

⁴Ahmed et al. (2013) propose an approach that fully overcomes this limitation.

which variables would satisfy these conditions. For example, Stifel and Christiaensen (2007) included ownership of several consumer durables in their imputation models, but Harttgen, Klasen and Vollmer (2012) later criticized this decision on the grounds that improvement in asset ownership outpaced income growth, creating “asset drift”. Identifying variables that satisfy the recommendations of Stifel and Christiaensen (2007) is intrinsically an empirical question, but the performance of the Kenya model could not be tested in the Kenya data because only one round of consumption data was available.

Christiaensen et al. (2012) moved one step further by analyzing a series of the past household budget surveys available in the Russian Federation, Vietnam, Kenya, and rural China. Their empirical strategy was to first create projection models using one round of household budget survey, impute household expenditure data into other rounds of household budget survey, and then compare projected poverty rates from the imputed expenditure with those estimated directly from actual consumption data. The comparison between projected poverty rates and directly estimated poverty rates validated the reliability of SSI projections. Models that included a comprehensive set of assets performed particularly well. These results are encouraging, and suggest that imputation models tend to remain stable over time.

Doudich et al. (2013) is the first attempt to validate the reliability of SSI using different types of surveys. They created consumption models in one round of the Moroccan Household Expenditure Survey (HES), imputed consumption data into more frequent Labor Force Surveys (LFS) and estimated poverty rates using the imputed data. To examine the accuracy of the projected poverty statistics, they compared them with poverty rates directly estimated from HIES consumption data collected in the same year as the LFS. The results are very encouraging. Projected poverty rates are very close to directly estimated poverty rates, irrespective of which round of HES is used to create consumption models. But, this raises the question of whether the results generalize to other countries and time periods. The objective of this paper is to use the same approach to test whether SSI between the HIES and LFS works in Sri Lanka.

II. Data

Household Income and Expenditure Survey (HIES)

We test the performance of SSI using the 2006 and 2009 rounds of the HIES and the LFS in Sri Lanka. Two rounds of the HIES were collected during this period, namely the 2006/7 and 2009/10 HIES surveys. HIES data are collected over a one-year period to capture seasonal changes in income and household expenditures. The 2006/7 HIES contains 18,544 households, while the 2009/10 HIES contains 19,958 households. The questionnaire includes questions on income, household expenditures, household and individual characteristics, housing conditions, and ownership of key consumer durables. Changes in the design of the questionnaire were minimal between two rounds. Due to the civil war, the 2006/7 HIES excludes the Northern and part of the Eastern Provinces, while the 2009/10 HIES covers the entire island (excluding only 3 districts from Northern Province) for the first time since 1983.

Labor Force Survey (LFS)

The LFS is collected each quarter. The questionnaire includes household and individual characteristics and a wide variety of labor statistics. Roughly 18,000 households are interviewed each year. Key statistics are designed to be nationally representative every quarter and representative at the district level each year. In both years, LFS excluded Northern Province, and included Eastern Province only in 2009.

Comparability of the LFS and HIES

To conduct SSI, the variables used in the consumption models need to be available in both the LFS and HIES data, which significantly restricts the set of available variables. For example, Christiaensen et al. (2012) found that ownership of consumer durables and housing conditions are correlated with household expenditures. But in Sri Lanka, only the HIES includes this information. As a result, for the HIES to LFS

imputation, we cannot include ownership of consumer durables and housing conditions in consumption models.

The time period of the LFS and HIES also should match. The LFS collects data from January to December, while the HIES data is fielded from July to June. To ensure consistency between the two surveys, we restricted the data to the common six month periods in 2006 and 2009 (i.e., July to December of each year) so that poverty rates projected from the SSI are comparable to those estimated from actual consumption data in the HIES.

Finally, we ensure that the geographic coverage is consistent across the surveys and across time. Table 1 shows missing provinces and districts in each survey, compared with HIES 2009/10 which covered the entire island. As mentioned above, Northern Province was excluded in every survey except the 2009/10 HIES. In addition, the 2006 LFS did not include Eastern Province. Furthermore, three districts in the Estate sector are not covered in the LFS. Given these differences in coverage, the subsequent analysis is restricted to those areas covered by both HIES and LFS in both 2006 and 2009, so that the two surveys cover identical time periods and districts.

**Table 1: Geographic areas not covered by
HIES and LFS in 2006 and 2009**

	2006	2009
HIES	N	None
LFS	N, E, NC*, D*	N, NC*, D*

Notes: N=Northern Province, E=Eastern Province, NC=North Central Province; D= Three districts (Gampaha, Hambantota and Puttalam). * Only missing in the estates sector.

Despite all these efforts to make these survey data comparable, incompatibilities between the surveys remain. In particular, both the reference period for employment statistics and the sampling strategies of the HIES and LFS differ, which make the two data sources not be directly comparable. This is a concern, given that differences in questionnaire design can have large and systematic effects on efforts to elicit labor force statistics in developing countries (Beegle, et. al., 2010). We discuss these issues below in Section V in greater detail.

Other variables

Variables that are not collected in either the HIES or LFS can also be included in the model, as long as they are aggregated at a regional level. For example, Stifel and Christiaensen (2007) include regionally aggregated rainfall data in their consumption models. In a similar vein, we include one additional regionally aggregated variable in the prediction model, namely industry level wage data published by the Central Bank of Sri Lanka. This wage data is based on information on monthly wage rate provided annually by the Ministry of Labor, which collects nominal wages for different trades fixed by Wage Boards and the Salaries and Wages in Public Administration Circulars in Sri Lanka. The Central Bank annually publishes separate real and nominal wage rate indices for each major industry and sector such as Workers in Agriculture, Workers in Industry and Commerce, Workers in Services, Central Government Employees and Government School Teachers. This wage data, even if they are aggregated to the industry level, are strong predictors of household expenditures.

Poverty trends

According to the official poverty statistics, the prevalence of poverty declined significantly between 2006 and 2009 in all three sectors (see the first main column in Table 2). The official statistics include all households in the HIES, but as mentioned above, we evaluate the performance of SSI using only geographic areas and time-periods common to the HIES and LFS data. The third to sixth columns in

Table 2 show poverty trends when we restrict the samples to the common areas covered in both surveys during the two years. As discussed before, the common survey period includes only the first six months for both rounds. These samples are the ones that are used for the HIES to LFS imputation analysis. Meanwhile, the fifth and sixth columns show poverty rates when we include all 12 month data but focus on common geographic areas of both rounds of HIES. These are the samples are used for the HIES to HIES imputation analysis, which will be discussed in detail later.

The changes in the survey period and geographic coverage have a modest effect on estimated poverty rates. For both rounds and all areas, poverty rates are in general higher when focusing on the common survey duration and geographic coverage of LFS and HIES, which are reported under columns named “Common districts: 6 months”.⁵ It is clear that in general, poverty declined in all areas and all combinations of samples and duration.

Table 2: Poverty Trends in Sri Lanka between 2006/07 and 2009/10

Poverty headcount (%)	Full sample		Common districts: 6 months*		Common districts: 12 months**	
	2006/7	2009/10	2006	2009	2006/7	2009/10
National	13.3	8.9	15.3	8.6	13.5	8.2
Urban	6.0	5.2	7.4	6.1	6.2	4.4
Rural	13.7	9.4	15.4	8.9	13.8	8.6
Estate	27.7	11.4	33.6	10.4	27.7	11.4

Source: Authors’ estimations using HIES 2009/10 data, using revised poverty lines based on CCPI (2002=100) are used for computations.

Note: *6 months and common districts between two years and HIES and LFS; ** 12 months and common districts between two years and HIES and LFS.

III. Imputation model

The imputation process follows the methodology developed in ELL (2003). The methodology has two stages: in the first stage, a model of log per capita consumption expenditure of household h is estimated in one round of HIES data. The model is defined as

$$y_{ch} = \alpha + X'_{ch}\beta + Z'_c\gamma + u_{ch} \quad (1)$$

where y_{ch} is log per capita expenditure, α is the intercept, X'_{ch} is the vector of explanatory variables for household h and location c , β is the vector of regression coefficients, Z'_c is the vector of location specific variables for location c , γ is the vector of coefficients for these variables, and u_{ch} is a stochastic error term. This error term is composed of two independent components: $u_{ch} = \eta_c + \varepsilon_{ch}$ where η_c is a cluster-specific effect and ε_{ch} is a household-specific effect. As is standard, both η_c and ε_{ch} are assumed to be orthogonal to X and Z and independent of each other and have a mean of zero. This error structure allows for both a location effect—common to all households in the same area—and heteroskedasticity in the household-specific errors. Location-specific variables Z can be defined at any level, such as province,

⁵This could be expected due to temporal monthly price changes not adjusted. Average CPI for first half of the survey 2006/07 is 145.0 compared to 156.0 in the second half. The difference between two half periods is high in 2006/07 compared to 2009/10 as the comparable CPI for 2009/10 are 209.9 and 215.7

district, or village—and can be drawn from any data source that includes all locations in the sample. All parameters regarding the regression coefficients (β, γ) and distributions of the error terms are estimated by feasible generalized least squares (FGLS).

In the second stage, the consumption model is used to impute household expenditure using a data set that contains the explanatory and location specific variables but lacks information on consumption. Poverty estimates and their standard errors can be calculated based on the imputed expenditure data.

The estimated regression coefficients $\hat{\beta}, \hat{\gamma}$ and the error terms η_c and ε_{ch} are stochastic, and ELL proposes a way to account for this when calculating poverty estimates and their standard errors. An imputed value of expenditure for each household is a sum of predicted log expenditure $(X'_{ch}\hat{\beta} + Z'_c\hat{\gamma})$ and error terms $(\tilde{\eta}_c \text{ and } \tilde{\varepsilon}_{ch})$ that are drawn randomly from the empirical distributions of the error terms, η_c and ε_{ch} . This random draw of error terms is repeated 100 times and for each round, household expenditures and corresponding poverty rates are estimated. The mean across the 100 simulations of a poverty indicator provides a point estimate of poverty, and the standard deviation is treated as a standard error of the point estimate of poverty rate. Because of the clear differences in consumption patterns between urban, rural, and estate sectors, we estimate separate consumption models for each sector.

IV. Survey-to-survey imputations between LFS and HIES

This section examines whether the SSI imputation analysis between the LFS and the HIES produces reliable poverty estimates. We test the performance of the SSI by comparing poverty rates imputed by SSI with those estimated from the actual consumption data available in the HIES. More specifically, we create two consumption models, one based on the 2006 HIES and the other on the 2009 HIES. These consumption models are then used to impute household expenditure into the LFS of the other survey year. Thus, the imputation exercise is conducted both backward and forward.

Forward projections

The forward projection means that we first estimate projection models using the 2006 HIES data and then impute household expenditures into households in the 2009 LFS data. To allow for regional differences, we create consumption models for urban areas, rural areas, and estates, separately. The third column of table 3 shows the results. Compared with the poverty rates estimated directly from the 2009 consumption data, imputing into the 2009 LFS overestimates the poverty rates for rural areas and estates and underestimates the poverty rate for urban area. As a whole, the forward imputation overestimates the poverty rate of 2009 at the national level, which is 3 percentage points higher than the direct estimation.⁶

⁶ Various specifications were tested, and province or district dummy variables as the location-specific variables are not included in the final model in order to decrease the number of time-invariant variables such as these variables and increase time-variant variables that would be able to explain the change across time. The estimation results are robust with or without these location dummy variables since the estimates and standard errors do not change as shown in Table A4 in Appendix. In addition, the final specifications do not include a cluster-specific error since the estimated cluster-specific variation was very small with the final model specification. This was measured by the ratio of the estimated cluster-level variance to the mean squared error, and the ratios are less than 0.03 in urban and rural sectors. In the estate sector, this ratio was about 0.08, but we could include the error in the estimate sector due to a problem in the software package (Povmap 2) as of now.

Table 3: Forward Imputation (HIES “2006” to LFS 2009 Imputation)

	Direct estimation		Imputation - 2009	
	2006	2009	Variables only from LFS	With wage data
National	15.3 (0.7)	8.6 (0.5)	12.3 (0.5)	11.1 (0.5)
Urban	7.4 (1.3)	6.1 (1.1)	4.4 (1.4)	4.1 (1.4)
Rural	15.4 (0.8)	8.9 (0.6)	12.3 (0.8)	10.8 (0.8)
Estate	33.6 (3.0)	10.4 (1.7)	31.5 (3.3)	31.7 (3.1)

Source: Authors’ calculations using HIES 2006/7 and 2009/10 data.

Notes: Tables A2-1 to A2-3 show the consumption models with means of these variables included in the final models for urban areas, rural areas and estates. The figures in parentheses are cluster-robust standard errors for the direct estimations and the standard errors that include the modeling and sampling errors for imputation.

The extent to which the imputed estimates overstate poverty varies by sector. In rural areas, the projected reduction in poverty was 3.1 percentage points, which accounts for less than half of the true reduction in poverty of 6.5 percentage points. In the estate sector, the differences are much larger. Projected poverty fell only 2 percentage points, while actual poverty fell 20 percentage points. On the other hand, the projection overestimates the reduction of poverty in urban areas. The projected reduction in poverty is 3 percentage points, while actual poverty fell only 1.3 percentage points.

Adding wage data

The mediocre performance of the imputation from the 2006 HIES to the 2009 LFS is likely due to a limited number of variables that are common in HIES and LFS data. In particular, the LFS does not include powerful poverty predictor such as ownership of consumer durables and housing conditions. Stifel and Christiaensen (2007), which also faced the same problem, proposed a solution – adding area-specific variables from other sources. In particular, they use rainfall data at the district level to improve the prediction of consumption models, in part because levels of rainfall vary over time. Of course, including area-specific variables can only explain variation across areas and not within them. Choosing the level of disaggregation therefore requires balancing the additional precision gained from using smaller areas, against the potential loss of precision in the data itself. For example, average wages at the village level may not be measured precisely enough to predict well in the model, while average wages at the national level will not be sufficiently disaggregated to be useful.

In Sri Lanka, area-specific wage data are not published. We therefore use national-level wage data disaggregated by sector. As explained above, this comes from a database of wage rates created by Department of Labor that captures monthly wage rates for four sectors.⁷ The data set shows substantial improvements in wage rates for some sectors (see Table 4). These data, however, are only available at the national level and are not separately collected for the urban, rural and estate sectors.

⁷ Because the wage variable is measured at the sector level, the estimates should account for clustering at the level of the sector in addition the level of the PSU. Because neither ELL nor simple probit models are well set-up to cluster at multiple levels, the standard errors for the models that include wage data are likely to be underestimated.

Table 4: Real Minimum Wage rate Indices of Workers in Wage Boards and Government Employees

December 1978 = 100	Agriculture		Industry and commerce		Services		All Central Government Employees	
Month	2006/07	2009/10	2006/07	2009/10	2006/07	2009/10	2006/07	2009/10
7	80.7	82.2	55.6	71.8	39.9	54.1	163.7	156.5
8	81	82.4	55.8	72.1	40.1	54.2	164.4	156.9
9	80.2	82.2	55.2	71.9	39.7	54.1	162.7	156.5
10	78.7	81.9	54.2	71.6	39	53.9	159.8	155.9
11	75.4	81.3	51.9	71.1	37.3	53.5	153	160.9
12	73.7	80.3	54.1	70.2	36.5	52.9	149.5	159
1	72.7	112	54	69.3	36	52.1	172.4	156.9
2	73.7	111.7	54.7	69.1	36.5	52	174.7	156.6
3	74.9	112.8	55.6	69.8	37.1	52.5	177.7	158.1
4	75	114	55.7	70.5	37.1	53.1	177.9	159.7
5	73.5	112.2	78.5	69.4	55.3	52.3	172.7	157.3
6	89.3	111.3	76.1	68.9	53.5	51.8	171	156

Source: Tables 86 and 87 in 2008 and tables 7.12 and 7.13 in 2011 publications of Economic and Social Statistics of Sri Lanka, Central Bank of Sri Lanka.

In order to incorporate the wage data in our consumption model, we construct a wage rate index that assigns 100 to the wage rate for the 2009/10 public sector. The obtained wage rate index is shown in Table 4. To control for heterogeneity between employed and self-employed workers and urban and rural areas and estates, we followed the following steps:

1. We identified seven industry clusters that covered most workers (4 clusters for those employed in public, agriculture, industry and commerce, and service sectors, and 3 clusters for those self-employed in agriculture, industry and commerce, and service sectors) in both the HIES and LFS data using individual questions on economic activities and main industries.
2. Using HIES 2009/10 data, the 2009/10 wage rate index for the seven industry clusters in three areas (urban and rural areas and estates) were derived by using the median wage income for individuals who mainly engaged in each of seven clusters in each area, as shown on the third column in Table 5. This step is necessary since the minimum wage indices in Table 4 do not indicate the difference in wage levels across sectors but only reflect changes within a sector across time.⁸
3. The 2006/7 wage rate index for the seven industry clusters in three areas was calculated using the obtained wage rate index in 2009/10 and the across-time changes in Table 4.
4. After merging the obtained wage rate index with the individuals in HIES or LFS data, the household-level wage index is calculated by assigning the sector-specific value of the index to each household member based on which sector he or she worked in and summing them up for each household. Finally, the per capita wage index is calculated by dividing it by a household size.

⁸HIES data are used for this purpose since HIES data include both employed and self-employed income, but LFS data include employed income only. Also, we did not calculate the wage levels using 1978 wage level data since the available 1978 wage data are old to be unreliable.

Table 5: Wage levels in Industry clusters in three areas

Industry clusters	Median income from HIES09/10	Wage rate index in 2009/10	Cross-time Change in wage (09/06) in Table 4	Wage rate index in 2006/7
Urban				
Public	25000	100.0	1.06	105.8
Agriculture, employed	10651	42.6	0.80	34.0
Industry, employed	13872	55.5	0.83	46.0
Service, employed	14500	58.0	0.77	44.5
Agriculture, self-employed	9400	37.6	0.80	30.0
Industry, self-employed	17100	68.4	0.83	56.7
Service, self-employed	22200	88.8	0.77	68.1
Rural				
Public	22166	88.7	1.06	93.8
Agriculture, employed	9679	38.7	0.80	30.9
Industry, employed	12500	50.0	0.83	41.5
Service, employed	12180	48.7	0.77	37.4
Agriculture, self-employed	11767	47.1	0.80	37.6
Industry, self-employed	13500	54.0	0.83	44.8
Service, self-employed	16710	66.8	0.77	51.3
Estate				
Public	8886	35.5	1.06	37.6
Agriculture, employed	7500	30.0	0.80	23.9
Industry, employed	8580	34.3	0.83	28.5
Service, employed	8000	32.0	0.77	24.5
Agriculture, self-employed	5390	21.6	0.80	17.2
Industry, self-employed	10000	40.0	0.83	33.2
Service, self-employed	10500	42.0	0.77	32.2

Source: Authors' calculations using HIES 2006/7 and 2009/10 data.

The results of including the wage index are mixed (see the fourth column of Table 3). The poverty projection for rural areas shows significant improvement. It declines from 12.3 percent to 10.8 percent, which is much closer to the direct estimation of 8.9 percent. But it still only explains about 60 percent of the actual fall in poverty. For the estate sector, the poverty projection shows a negligible reduction toward the actual poverty rate (31.5 percent from 31.7 percent) and the difference from the direct estimation (10.4 percent) remains very large. For urban areas, the poverty estimation declines from 4.6 percent to 4.1 percent, further away from the direct estimation of 6.1 percent. As a whole, the imputed poverty rate at the national level declines from 12.3 percent to 11.1 percent, but it is still 2.5 percentage points higher than the direct estimation of 8.6 percent.

Given the magnitude of the differences between the projected and direct estimates and the large size of the two household surveys, it is likely that these differences are statistically significant. To verify this, we conduct two types of tests. First, we examine whether the 95 percent confidence interval of the direct poverty estimates contains the corresponding projections. Second, we test whether a difference between a projected estimate and a direct estimate is statistically significant. In this second case, it is important to be careful in interpreting the results of these tests. In particular, the test could fail to reject the null

hypothesis of equality because the projection is noisy. In other words, if a projection is very imprecise and its standard error is large, the test is less likely to find the difference is statistically significant.

For rural areas and estates, even after including the wage data, the projected poverty estimates for rural areas and estates are outside the 95 confidence interval of the direct estimates and the differences are statistically significant at the one percent level. The projected poverty estimate for urban areas performed slightly better. The difference between the projected and the direct estimates is not statistically significant at the 95 percent level, but the projected estimate lies outside the 90 percent confidence interval of the direct estimate. At the national level, the projected poverty estimate is above the upper bound of the 95 percent confidence interval of the direct estimate, and the difference is statistically significant at the one percent level.

Backward projections

We now reverse the direction of the survey-to-survey imputation. We estimate the consumption model in the 2009 HIES, and impute household expenditure into the 2006 LFS. To assess the accuracy of the prediction, we compare the imputed poverty rates with those directly estimated from the 2006 HIES.

When the wage variable is excluded, the results are dismal. In urban areas, the imputation predicts that poverty in 2006 was 4.1 percent when the direct estimate was 7.4 (Table 6). Therefore, the imputation suggests that poverty increased from 4.1 to 6 percent between 2006 and 2009 when in fact it declined. The rural model predicts that poverty in 2006 was 8.8 percent, implying virtually no change between 2006 and 2009, when in fact rural poverty fell 6.5 percentage points. Similarly, the imputation suggests little movement in the estate sector, despite a huge reduction in poverty of 23 percentage points. At the national level, the imputed poverty estimate of 2006 is 8.3 percent. This is 7 percentage points lower than the actual national poverty rate of 15.3 percent, and slightly lower than the 2009 direct estimate of 8.6 percent. This method therefore predicts a small increase in poverty between 2006 and 2009, when in fact poverty fell substantially.

The estimates after adding the wage variable are slightly closer to the direct estimates, but remain far off the mark. For example, the imputed poverty rate for urban areas increases from 4.1 percent to 4.6 percent, but the direct estimate is 7.4 percent. For rural areas, the imputed poverty rate rises from 8.8 percent to 10.3 percent, as compared to a direct estimate of 15.4 percent. At the national level, the imputed poverty rate increases slightly to 9.8 percent but remains far lower than the direct estimate of 15.3 percent.

Like the forward projection, for rural areas and estates, the projected poverty estimates of 2006 are outside the 95 percent confidence intervals of the corresponding direct estimates. Also, the differences between the projected and the direct estimates are statistically significant at the one percent level. For urban areas, the projected poverty estimate lies outside the 95 percent confidence interval of the direct estimate. Furthermore, the difference is also statistically significant at the 10 percent level. At the national level, the imputed poverty rate is below the 95 percent confidence interval of the direct estimate and the difference is statistically significant at the one percent level.

Table 6: Backward imputation (2009 HIES to 2006 LFS Imputation)

	Direct estimation		Imputation - 2006	
	2009	2006	Variables only from LFS	With wage data
National	8.6 (0.5)	15.3 (0.7)	8.3 (0.8)	9.8 (0.8)
Urban	6.1 (1.1)	7.4 (1.3)	4.1 (1.5)	4.6 (1.6)
Rural	8.9 (0.6)	15.4 (0.8)	8.8 (0.9)	10.3 (1.0)
Estate	10.4 (1.7)	33.6 (3.0)	10.8 (3.5)	14.0 (3.6)

Source: Authors' estimations using HIES 2006/7 and 2009/10 data.

Notes: Tables A3-1 to A3-3 show the consumption models with means of these variables included in the final models for urban areas, rural areas and estates. The figures in parentheses are cluster-robust standard errors for the direct estimations and the standard errors that include the modeling and sampling errors for imputation.

V. Two sources of failures in SSI

There are potentially two underlying causes for the failure of SSI to accurately predict changes in poverty. First, the HIES and LFS surveys may not be compatible, because the questionnaires differ in important ways. Finally, the HIES to LFS imputation is limited to the set of variables that are available in both the HIES and LFS. If these variables do not closely track changes in consumption, the resulting models will not accurately project changes in poverty rates. In addition, a different sampling scheme can also affect how well the imputed results from the LFS match the HIES. This section investigates these three possibilities.

To see the effects of the two potential sources of error, we impute from the HIES into itself. In other words, we create a consumption model using one round of the HIES and use it to estimate consumption based on proxy indicators in the other round of the HIES. To make the imputed poverty rates comparable to those in the HIES to LFS imputations discussed above, we use the same survey period, geographic coverage, and consumption models.⁹ The only difference is the source of the welfare proxies used to predict consumption, which are now taken from the HIES instead of the LFS. If using proxies from the HIES instead of the LFS substantially improves the accuracy of the estimates, then it is likely that the HIES and LFS surveys are using incompatible samples or questions.

Table 7 shows that for rural areas and estates, the estimated poverty rates change little when using the HIES instead of the LFS, implying that the effect of different survey designs is minimal in these areas. For urban areas, however, the poverty rates projected by the HIES to HIES imputation become much closer to the corresponding direct estimates. For example, for the backward imputation, the HIES to HIES imputation projects a poverty rate of 7.2 percent, which is just 0.2 percentage point different from the direct estimate of 2006 and inside its 95 percent confidence interval. For the forward projection, the HIES to HIES imputation projects 6.6 percent, which is just 0.5 percentage point higher than the direct estimate of 2009 and inside the 95 percent confidence interval.

⁹ Furthermore, all models do not include wage data.

Table 7: Comparisons of poverty rates (%) between the HIES to HIES and the HIES to LFS imputations

	Backward imputation		Direct estimation of 2006	Forward imputation		Direct estimation of 2009
	HIES to HIES	HIES to LFS		HIES to HIES	HIES to LFS	
Urban	7.2 (1.6)	4.1 (1.5)	7.4 (1.3)	6.6 (1.5)	4.4 (1.4)	6.1 (1.1)
Rural	9.5 (0.9)	8.8 (0.9)	15.4 (0.8)	12.7 (0.9)	12.3 (0.8)	8.9 (0.6)
Estate	11 (3.4)	10.8 (3.5)	33.6 (0.7)	30 (2.7)	31.5 (3.3)	10.4 (1.7)

Source: Authors' estimations using the corresponding rounds of HIES and LFS

Notes: For the backward imputation, consumption models are estimated using HIES 2009/10 data; for the forward imputation, consumption models are estimated from HIES 2006/07 data. For both imputations, we did not include district level wage data. Also, for comparisons, the HIES to HIES imputations use exactly the same models as the HIES to LFS imputations. The figures in parentheses are cluster-robust standard errors for the direct estimations and the standard errors that include the modeling and sampling errors for imputation

The LFS and the HIES generally use a similar sample design, but there are several differences that can partly explain these differences in predicted poverty rates. Both surveys are stratified on urban, rural, and estate sectors within each district, and each selected around a total of 2500 census blocks and ten household per block, based on the 2001 population census. However, the surveys use different sampling strategies. In the HIES, the number of PSUs in each sector are selected using a Neymann allocation to minimize the variance of estimated mean per capita consumption, while the LFS uses a similar approach that minimizes the variance of the estimated unemployment rate. As a result, the HIES includes nearly two and a half times more primary sampling units than the LFS in the urban and estate sectors.

Table 8: Number of household respondents by sector and survey

	HIES: July-December 2009		LFS: July-December 2009	
	Primary Sampling Units	Households	Primary Sampling Units	Households
National	1001	8,927	907	8,464
Urban	261	2,240	108	966
Rural	637	5,768	747	7,005
Estate	103	919	52	493

Source: Authors' calculation based on 2009/10 HIES and 2009 LFS data

This difference in sampling allocation strategy does not lead to biased estimates in either survey, because both samples are weighted to be representative of the 2001 census. However, the unique sampling strategy of the LFS hampers our ability to test the accuracy of SSI. This is because the households

included in each survey are less comparable, compared to a situation in which each survey uses the same sampling strategy. This in turn makes SSI based on the LFS appear to be less accurate when benchmarking it against actual expenditures from the HIES, reducing the apparent accuracy of the SSI results. In other words, the different sampling design likely increases the discrepancy between the HIES and the LFS. This design difference could be partially corrected by adjusting the weights on the LFS to make the sample more comparable to the HIES, but whether this would substantially improve the accuracy of the validation is an open question left for future work.¹⁰

Given the differences in the sampling strategies used to collect the HIES and LFS, it is important to check whether the surveys are comparable, both in terms of the geographic distribution of households and the types of household interviewed. To get a better handle on this, we compare the population shares and estimated poverty rates for the seven provinces within each sector, to check if the LFS underrepresents poor provinces in urban areas (see Table 9). In rural areas, both the poverty rates at the province level and the population shares across provinces differ little. In the estate sector, the province level poverty estimates are also very similar across surveys, except for Western Province. However, Western province makes up a small share of the estate sector, and as a result, the national average poverty rate for estates differ little between the HIES to HIES and the HIES to LFS imputations.

Table 9. Comparisons in population shares (%) and poverty rates (%) between the HIES to HIES and the HIES to LFS imputations – HIES “2009”

Province	Urban				Rural				Estate			
	Population share		Poverty rate		Population share		Poverty rate		Population share		Poverty rate	
	HIES	LFS	HIES	LFS	HIES	LFS	HIES	LFS	HIES	LFS	HIES	LFS
All			6.6	4.4			12.7	12.3			30.0	31.5
Western	71.5	78.2	7.2	4.4	28.0	29.4	9.5	9.7	5.8	3.6	27.6	15.9
Central	8.8	8.2	5.2	3.2	13.4	11.4	12.3	13.0	51.5	54.1	30.0	32.3
Southern	8.6	8.2	4.3	4.8	16.7	15.6	15.6	13.4	3.8	1.0	27.0	33.1
North Western	3.8	2.3	8.1	6.5	14.9	16.0	13.9	12.6	1.1	0.2	25.5	27.7
North Central	2.1	1.2	2.9	3.3	8.1	7.8	13.2	12.5	0	0		
Uva	2.6	0.7	4.1	1.2	7.7	7.8	14	13.8	15.9	23.4	30.1	32.2
Sabragamuwa	2.8	1.3	3.5	7	11.2	12.1	13.8	14.9	22	17.6	31.1	31.2

Source: Authors' estimations using the corresponding rounds of HIES and LFS

Note: Poverty rates are predicted using the same periods, geographic coverage and consumption models of the HIES to HIES imputation for both HIES to HIES and HIES to LFS imputations.

There are more discrepancies between the two surveys in urban areas, where we see the largest differences in mean household characteristics. For example, the poverty rates for Western provinces differ by 3 percentage points. In addition, Western province's share of the population is 71 percent in the HIES and 78 percent in the LFS. Given that the LFS underestimates poverty within the urban areas of Western

¹⁰ One potential strategy for balancing the two samples on observables would be estimate the probability that a particular household is in the LFS as opposed to the HIES, given its observable characteristics in 2006 or 2009. The estimated probabilities could then be used to reweight the LFS sample in order to balance the two samples on observables.

province and has a larger share of the sample in that province, it is not surprising that imputation using the LFS understates poverty in urban areas nationally.

What is causing these differences in imputed poverty rates within strata? In fact, the HIES and LFS have important differences in the wording of the questionnaire. In particular, for questions regarding employment the reference period of the LFS is one week, while the HIES asks about the respondent's current activity. In addition, the LFS classifies people as employed if they typically engage in economic activity, even if they did not in the past week. How questions are worded and the reference period used can have major effects on responses to labor-related questions.

In this case the different reference periods and questions affect trends in employment in addition to levels (Figure 1). Employment rates are consistently higher in the LFS than the HIES for the total population, because the LFS definition includes people that engaged in a few hours of work. For heads, who tend to work longer hours, these differences in wording do not appear to lead to a systematic bias. The differences in employment rates due to the different wording of the questionnaires is only moderately large, however, and never exceeds four percentage points.

For the purposes of survey-to-survey imputation, however, what is critical is the discrepancy in the two surveys in employment trends. Between 2006 and 2009, the employment rate for the HIES declined substantially for both heads and adults, by about 2 percentage points. On the other hand, the LFS employment rate changed little, declining slightly for heads and increasing slightly overall. Because of differences in the wording of the questionnaire, in conjunction with differences in sampling strategy, employment changes in the LFS do not reflect those in the HIES.

One way to see how using the HIES rather than the LFS affects the results is to examine how mean predicted consumption is affected. Table 10 shows that using the LFS instead of the HIES raises predicted mean household expenditure per capita by 11 percent. For rural areas and estates, however, the results are identical whether the HIES or the LFS is used to predict consumption. This is consistent with the above results.

Two variables explain a large portion of this discrepancy in imputed per capita consumption in urban areas: Whether a person is employed, and whether the household head has an advanced level of education (see Table 11).¹¹ The differences in these two variables across the surveys explains 5.5 percent of the difference in imputed household expenditure, which is about half of the 11 percentage point difference in imputed household expenditure.¹² In rural areas, the employment ratio is also higher in the LFS, but the impact on the imputation is negligible since the magnitude of the coefficient is half of that in urban areas. The differences in screening questions and the reference period used by each survey likely account for the observed discrepancy in employment rates. With respect to education, in contrast, the two surveys use identical questions. Therefore, the inconsistent reported rates of education between the two surveys likely stems from the different sampling strategy employed by each survey. Further analysis will be necessary to fully understand the source of the differences in estimated means between HIES and LFS.

¹¹ In this paper, we grouped completed education levels into four categories: (1) Passed Grade 5 or less, (2) Passed Grade 6 to 10, (3) Passed Grade 12 or General Certificate of Education (GCE) Ordinary Level (hereafter OL), (4) Passed GCE Advanced Level or above (hereafter AL).

¹² These numbers are not strictly comparable, because the 11 percent difference reported in table 9 is based on the ratio of averages while the percentage difference reported in table 10 are based on differences in logs, but they are indicative of the important role of these two variables.

Figure 1: Employment rate by survey and year

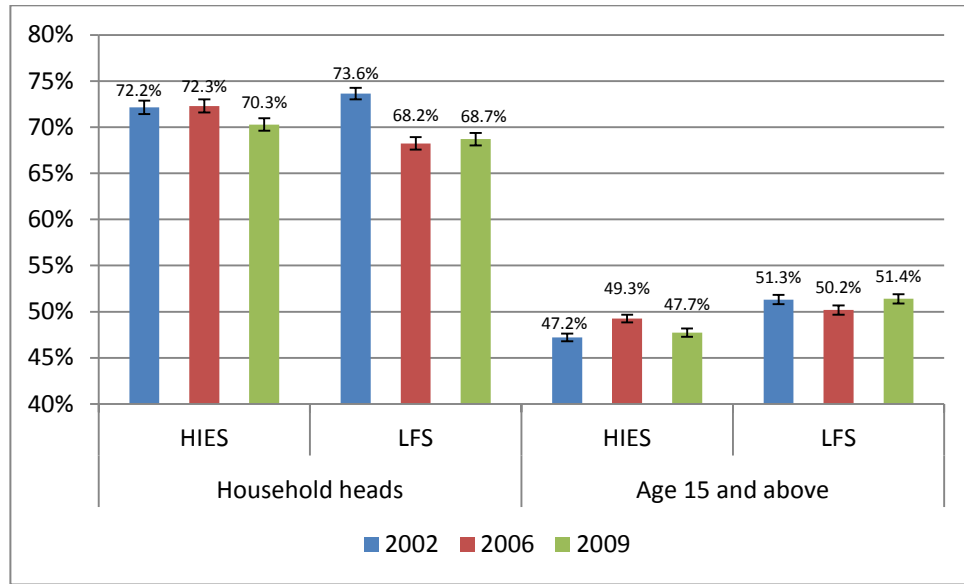


Table 10: 2009 imputed per capita consumption expenditures (Rs.) evaluated at means of the right hand side variables based on HIES and LFS

	HIES	LFS	Ratio of prediction (LFS/HIES)
Urban	4251	4701	1.11
Rural	3015	3017	1.00
Estate	1928	1926	1.00

Source: Authors' estimations using the corresponding rounds of HIES and LFS

Table 11: Two variables explain about half of the 11 percent discrepancy in imputed in urban areas for imputed household expenditures per capita

Right hand side variables (X)		Coeff.	Mean of X		X*Coeff		Difference in prediction (approx. pp)
			HIES	LFS	HIES	LFS	
Urban	Employed ratio	0.99	0.350	0.380	0.348	0.377	2.9
	Head's education passed AL	0.82	0.190	0.228	0.156	0.187	2.5
Rural	Employed ratio	0.45	0.379	0.411	0.172	0.186	0.7
	Head's education passed AL	0.54	0.102	0.100	0.055	0.054	0.1
Estate	Head's education passed AL	0.48	0.020	0.027	0.010	0.013	0.1

Source: Authors' estimations using the corresponding rounds of HIES and LFS. Note: The samples of HIES and LFS data are restricted to the common survey period and geographic coverage. In the estates, "employed ratio" was not included in the estimation model.

HIES to HIES imputation using additional variables

The analysis so far has shown that differences in survey design bias imputations based on the LFS in urban areas, but shows almost no evidence that it affects the imputations for rural areas and estates. Then, what is causing the large difference between the HIES to LFS imputation and the direct estimation in rural areas? This sub-section investigates another potential cause of failure in survey-to-survey imputation – bias due to the inability of common variables to adequately track changes in household expenditure. We assess this by adding variables only available in the HIES to the model. This analysis builds on the HIES to HIES imputation in Table 7, which was limited to variables common to the LFS and the HIES.

Expanding the set of variables in the model is likely to reduce the bias. Many of the common variables, such as household demographics and education status, change little over the course of the three years. For example, the rural average household size, which is one of the strongest predictor of poverty, changed only from 4.35 to 4.25 in rural areas between the two rounds of HIES surveys. Even though employment variables can change rapidly in many contexts, they remained nearly constant in the data sets used in this analysis. As mentioned above, variables that are strong predictors of consumption in the cross-section will not be useful indicators of changes in poverty if they do not reflect changes in consumption over time.

This potential limitation can be addressed by adding additional variables to the HIES model that are not available in the LFS, which improves the accuracy of consumption models. The main variables excluded from the HIES to LFS imputation are ownership of consumer durables and housing conditions. These variables are known to be powerful predictors of poverty and household expenditures (see for example Stifel and Christiaensen, 2007, and Doudich et al. 2013). Collecting information on ownership of consumer durables and housing conditions is easy, quick, and cheap.

Tables 12, 13, and 14 show the results of HIES 2006/7 to HIES 2009/10 imputations with these additional variables. The final models still include measures of household demographics and employment status, which change little over time. They are included in the final models because they explain a large share of the cross-sectional variation in consumption, which remain important to distinguish the poor from the non-poor in the 2009/10 data. Ownership of key consumer durables and housing conditions became significantly more widespread over time. The improvements are particularly striking in rural areas and estates, which are the two sectors in which the previous models underestimated the decline in poverty.

The asset for which ownership expanded most rapidly is cell phone ownership. But, this rapid expansion may reflect reductions in cell phone prices and service fees. As a result, owning cell phones is likely to have very different financial consequences in 2009 than in 2006. Therefore, we follow previous studies that exclude cell phone ownership from consumption models (see Revilla et al.(2010) for Zambia and Allwine et al. (2013) for Lesotho).

Table 12. HIES to HIES imputation from 2006/07 to 2009/10 for urban areas

Description	Coefficient	Std. Err.	Mean of X in	
			2006/7	2009/10
Intercept	8.3644	0.0533		
Bicycle	0.1162	0.0158	0.311	0.271
Boats	0.5614	0.1153	0.012	0.004
Bus lorry	0.4368	0.1016	0.015	0.014
Motor car van	0.4174	0.028	0.122	0.120
Computers	0.152	0.0225	0.176	0.263
Cookers	0.173	0.0176	0.745	0.772
Lighting-electricity or solar electricity	0.2067	0.0337	0.953	0.965
Electric Fans	0.1107	0.02	0.785	0.804
Flat house	0.0986	0.0308	0.057	0.076
Fridge	0.1156	0.019	0.604	0.656
Head's major industry: Agriculture	0.1232	0.0675	0.012	0.014
Household size	-0.2769	0.0126	4.4	4.3
Household size squared	0.0136	0.001	22.4	21.7
Head's education completed primary	0.1144	0.018	0.408	0.445
Head's education passed OL	0.1309	0.022	0.221	0.186
Head's education passed AL	0.2722	0.0273	0.176	0.199
Telephone	0.1329	0.0182	0.503	0.554
Floor>750 sqft	0.0744	0.0162	0.409	0.458
Line room/Row house	-0.0743	0.0221	0.104	0.072
Motor Bicycle	0.1012	0.0207	0.150	0.179
Only paid income	-0.2023	0.0299	0.575	0.573
Only self-income	-0.2085	0.0315	0.192	0.180
Paid and self-income	-0.1613	0.0415	0.122	0.116
Per capita wage index	0.0062	0.0006	20.9	23.6
Three wheeler	0.1024	0.0308	0.057	0.066
Has own or shared toilet	-0.07	0.036	0.965	0.935
TV	0.0558	0.0235	0.881	0.901
VCD	0.0879	0.0156	0.448	0.517
Washing machine	0.1618	0.0209	0.283	0.317
Water pumps	0.1784	0.0618	0.017	0.002
Source: Authors' calculations using HIES 2006/06 and 2009/10 data				

Table 13. HIES to HIES imputation from 2006/07 and 2009/10 for rural areas

Description	Coefficient	Std. Err.	Mean of X in	
			2006/7	2009/10
Intercept	8.3354	0.0388		
Head's age	-0.0024	0.0003	51.2	51.5
Bicycle	0.0768	0.0078	0.430	0.388
Boats	0.2085	0.0732	0.003	0.003
Bus lorry	0.3962	0.0452	0.018	0.021
Motor car van	0.3411	0.0245	0.053	0.056
Computers	0.1437	0.0204	0.059	0.120
Cookers	0.1631	0.0114	0.320	0.351
Dependency ratio (age -14, 64+ to HH size)	-0.0364	0.0173	0.318	0.323
Electric Fans	0.0924	0.0101	0.422	0.488
Flat house	0.0933	0.0376	0.014	0.015
Mud for floor	-0.0825	0.0126	0.121	0.093
Fridge	0.08	0.0115	0.338	0.403
Head's major industry: Agriculture	-0.03	0.0099	0.236	0.226
Household size	-0.267	0.0098	4.0	3.9
Household size squared	0.0145	0.0009	18.5	17.8
Head's education completed prim	0.0227	0.0094	0.442	0.480
Head's education passed OL	0.0492	0.015	0.150	0.145
Head's education passed AL	0.0956	0.0201	0.093	0.100
Lighting-kerosene	-0.0476	0.0124	0.183	0.120
Telephone	0.1653	0.0108	0.304	0.495
Floor>750 sqft	0.1007	0.0086	0.386	0.408
Motor Bicycle	0.1475	0.0102	0.220	0.281
Maximumhousehold education completed primary	0.0881	0.0179	0.437	0.430
Maximumhousehold education passed OL	0.162	0.0203	0.245	0.245
Maximumhousehold education passed AL	0.2001	0.0219	0.257	0.272
Only paid income	-0.1908	0.0178	0.473	0.455
Only self-income	-0.1588	0.0179	0.281	0.294
Source of drinking water within premises	0.0187	0.0085	0.691	0.760
Paid and self-income	-0.209	0.0225	0.147	0.131
Per capita wage index	0.0079	0.0004	17.3	19.8
Roof: concrete	0.1784	0.0314	0.017	0.026
Sewing Machine	0.0396	0.0089	0.447	0.451
Pesticider	0.0816	0.0201	0.043	0.048
Three wheeler	0.2041	0.0212	0.048	0.074
Has own or shared toilet	0.0561	0.0242	0.976	0.983
Tractor 2 wheel	0.1571	0.024	0.031	0.038
Tractor 4 wheel	0.1487	0.0709	0.007	0.009
TV	0.067	0.011	0.777	0.818
Unemployed ratio to household size	-0.0784	0.0315	0.048	0.052
VCD	0.0691	0.0099	0.259	0.346
Wall: pressed soil blocks	-0.09	0.0162	0.051	0.041
Washing machine	0.0907	0.018	0.088	0.121

Source: Authors' calculations using HIES 2006/07 and 2009/10

Table 14. HIES to HIES imputation from 2006/07 to 2009/10 for the estate sector

Description	Coefficient	Std. Err.	Mean of X in	
			2006/7	2009/10
Intercept	8.1312	0.0496		
Motor car van	0.3413	0.188	0.011	0.011
Computers	0.3564	0.0856	0.014	0.022
Cookers	0.1071	0.0293	0.106	0.115
Dependency ratio (age -14, 64+ to HH size)	-0.1315	0.039	0.328	0.324
Head's major industry: Inactive	-0.0654	0.0198	0.271	0.248
Household size	-0.2078	0.0127	4.2	4.2
Household size squared	0.0088	0.001	21.4	20.6
Head's education passed AL	0.3811	0.0825	0.018	0.023
Telephone	0.1465	0.0294	0.097	0.389
Floor>750 sqft	0.1207	0.0388	0.060	0.059
Line room/Row house	-0.0612	0.0179	0.693	0.640
Maximum household education completed primary	0.0826	0.0207	0.602	0.620
Maximumhousehold education passed OL	0.1427	0.0344	0.075	0.090
Maximumhousehold education passed AL	0.18	0.0519	0.051	0.062
Per capita wage index	0.0028	0.0015	11.8	14.3
Radio	0.0378	0.0186	0.693	0.694
Three wheeler	0.2134	0.0843	0.011	0.023
TV	0.0969	0.0192	0.617	0.705
VCD	0.0694	0.0217	0.230	0.399
Washing machine	0.2095	0.0826	0.013	0.016

Source: Authors' calculations using HIES 2006/07 and 2009/10

Table 15 summarizes the results of the HIES to HIES imputations with additional variables. For each sector, the left-most column shows the estimates using only variables common to the LFS and the HIES. The next column adds wage data, and the final column adds the consumer durable and housing data available in the HIES. The direct estimates of poverty differs slightly from those reported previously (Table 7) because it reflects the full HIES sample.¹³

For urban areas, adding the additional variables on consumer durables has little effect. This is because, as discussed above, imputing from the HIES to the HIES based on variables common to the LFS works well in urban areas, so the benefit of including additional information is small. This reconfirms that the main source of biases in the imputations for urban areas is the difference in survey designs between the HIES and the LFS. However, there is noticeable improvement when wage data are added to the model.

For rural areas and estates, adding ownership of consumer durables and other variables to the model greatly improves the accuracy of the predictions. This is particularly noticeable for the backward imputations. For example, for rural areas, backward imputation with the common variables only estimates a poverty rate of 11.1 percent for 2006/7. This is about midway between the direct estimates of 2006/7 and 2009/10, which suggests that common variables of LFS and HIES are not fully tracking the increase in consumption between 2006/7 and 2009/10. Adding wage data improves the imputed poverty rate slightly to 12.6 percent. However, adding further ownership of consumer durables and housing conditions, which are available only in HIES, improves the imputed poverty rates significantly. The imputed poverty rate of 2006/7 is now 13.4 percent, which is just 0.4 percentage point lower than the direct estimate and inside its 95 percent confidence interval. For estates, although including durable goods

¹³ Table 14 is based on the full HIES sample from July to June, whereas table 7 is based on a subsample from July to December covering a subset of districts.

and housing conditions improves the poverty estimate, it remains far below the lower bound of the direct estimate.

Table 15. The HIES to HIES imputations

	Estimation of poverty rates (%) for 2009/10				Estimation of poverty rates (%) for 2006/07			
	Forward imputation			Direct estimation	Backward imputation			Direct estimation
	Variables only from LFS	With wage data	With wage & durables		Variables only from LFS	With wage data	With wage & durables	
National	11.6 (0.51)	10.2 (0.53)	9.0 (0.47)	8.2 (0.39)	10.3 (0.55)	11.7 (0.56)	12.5 (0.55)	13.5 (0.46)
Urban	5.0 (0.99)	4.7 (0.99)	4.1 (0.99)	4.4 (0.66)	4.6 (0.95)	5.1 (0.92)	5.2 (0.87)	6.2 (0.75)
Rural	11.7 (0.67)	10.0 (0.61)	9.0 (0.61)	8.6 (0.45)	11.1 (0.64)	12.6 (0.66)	13.4 (0.64)	13.8 (0.53)
Estate	26.4 (2.71)	26.8 (2.23)	21.6 (2.23)	11.4 (1.63)	13.2 (2.54)	15.4 (1.44)	19.0 (2.57)	27.7 (2.24)

Source: Authors' estimations using the corresponding rounds of HIES.

Notes: Tables A1-1 to A1-3 show the consumption models with means of these variables included in the final models for urban areas, rural areas and estates. The figures in parentheses are cluster-robust standard errors for the direct estimations and the standard errors that include the modeling and sampling errors for imputation.

At the national level, the backward imputation estimates a poverty rate of 12.5 percent for 2006/7 once both wage data and ownership of durables are included. This is a 2.2 percentage point increase compared to the prediction based on variables that are common to the LFS. It still underestimates the poverty rate compared with the direct estimate of 13.5 percent. The imputed poverty rate is now within the 95 percent confidence interval of the direct estimation, and the difference between the imputed and direct estimates is no longer significant at the five percent level. The result of forward imputation is also qualitatively the same – expanding the variable set by adding ownership of consumer durables improves the precision of imputation substantially.

In sum, the HIES to LFS imputation fails to accurately project poverty estimates in the context of Sri Lanka between 2006/7 and 2009/10. The above analyses clarified likely causes of the biases in the imputations for urban areas and rural areas, but still failed to find causes of the biases in the imputations for estates. The main reason for the drop for estate sector in direct estimation may be due to wage increase in the plantation sector employees in late 2009, which was not captured in either the wage index or the common variables used in the model.

For urban areas, the LFS fails to predict well because of differences in the design of the LFS and HIES. In particular, the reference period for employment differs between the HIES and the LFS, and the two surveys also follow slightly different sampling schemes. But, the subsequent analyses using the HIES to HIES imputation suggest that if the LFS questionnaire and sampling scheme were made consistent with the HIES, the survey-to-survey imputation would work well.

For rural areas, the biases are caused by the limited availability of variables that change over time. As a result, the variables in the model explained little of the actual change in consumption, which were instead captured in changes in the model's intercept or other coefficients. When other variables that better reflect

short-term changes in welfare, such as ownership of consumer durables and housing conditions were added to the model, the projected poverty changes were far more accurate.

For estates, the survey design effects do not appear to cause any bias and adding consumer durables and housing conditions makes the imputations somewhat accurate. However, the fall in poverty in this sector was very rapid and the variables included in the model could not adequately capture it.

Although this analysis clarifies that using additional variables substantially improves the accuracy of the imputation in rural areas, these additional variables are not currently available in LFS. Therefore, we conclude that it is not possible to accurately predict poverty using the existing LFS. If the LFS were used to create a frequent welfare tracking system in Sri Lanka, either the LFS should expand its questionnaire while modifying its questionnaire and sampling strategy to become comparable to the HIES, or a new survey containing the full set of welfare proxies in the HIES should be established.

Imputing poverty directly using a Probit model

The imputation approach developed by Elbers et al. (2003) has become one of the most popular approaches in projecting poverty statistics. It has been utilized widely by researchers and practitioners and continues to evolve in response to technical issues raised by the community of users. However, this is by no means the only methodology that can project poverty statistics. Here, we use a simpler and more direct method by projecting poverty using a probit model, as suggested by Tarozzi and Deaton (2007), and Tarozzi (2011).

Table 16 shows the results of backward and forward imputations, respectively. In general, the results are similar to those generated using ELL, but slightly less accurate. For example, when imputing poverty in 2006, adding wages to the probit model only changes estimated poverty from 4.0 to 4.2 percent, while in ELL estimated poverty rises from 4.6 to 5.1. When imputing 2009 poverty in rural areas, the probit estimator projects a poverty rate of 9.8 percent for 2009/10, which is still 1.2 percentage points higher than the direct estimation. This compares unfavorably with the imputation based on the ELL method, projects a poverty rate of 9.0 percent, which is only 0.4 percentage point larger than the direct estimation. The forward imputation also shows qualitatively similar results. For estates, even after ownership of consumer durables and housing conditions are added into the consumption models, both forward and backward projections fail to produce accurate estimates of poverty rates. As in rural areas, the probit based imputations produce less accurate projections of poverty rates than the ELL based imputations.

Table 16. The HIES to HIES imputations using Probit as imputation formula

	Estimation of poverty rates (%) for 2009/10				Estimation of poverty rates (%) for 2006/07			
	Forward Imputation (probit)			Direct estimation	Backward imputation (probit)			Direct estimation
	Variables only from LFS	With wage data	With wage & durables		Variables only from LFS	With wage data	With wage & durables	
National	12.3	11.0	9.7	8.2	8.8	9.6	11.1	13.5
Urban	6.1	4.7	3.8	4.4	4.0	4.2	5.3	6.2
Rural	12.5	11.1	9.8	8.6	9.3	10.1	11.7	13.8
Estate	25.1	24.9	22.0	11.4	14.0	16.7	17.1	27.7

Source: Authors' estimations using HIES 2006/07 and 2009/10 data

VI. Concluding remarks

Survey-to-survey imputation, when applied to existing data, can be a cost-effective and powerful tool to produce frequent and timely estimates of trends in poverty. This paper proposes several tests to examine whether this technique can be used to increase the frequency of poverty estimates in Sri Lanka, where an expenditure survey is conducted every three years while a labor force survey is conducted every quarter. Doudich et al. (2013) explored this approach for Morocco, where an LFS is conducted every quarter but Household Expenditure Surveys (HES) are conducted every seven years. That study showed that in Morocco, imputations based on LFS data can produce statistically reliable poverty rates in each quarter. In contrast, this paper presents strong evidence that unless the sampling and questionnaire of the LFS are changed or a new household survey is created, survey-to-survey imputation in Sri Lanka will not work.

The analysis demonstrates two potential pitfalls to successful survey-to-survey imputation. First, the wording of the questionnaires of the two surveys and households' responses to the questions can differ between the two surveys. Second, imputing from one survey to another restricts the prediction model to the set of variables available in both surveys, which may not accurately capture changes in household expenditure per capita. In addition, if the two surveys use different sampling strategies, validation tests can make the imputation appear to be less accurate than it actually is.

The analysis shows that in urban areas, SSI produces biased estimates of poverty rates due to the differences in wording of the employment question in the two surveys. In addition, differences in the sampling design, while not causing bias, contributed to the observed discrepancy between the two surveys. In the rural and estate sectors, however, the estimated poverty rates were biased because the limited set of proxies common to both surveys do not accurately reflect changes in household expenditure. As a result, both forward and backward projections severely underestimate the decline in poverty observed during this period. In urban Sri Lanka, labor market outcomes such as employment rates, sectors, and wages track changes in living standard fairly well. On the other hand, in part because self-employment income and agricultural income is difficult to collect, the labor market outcomes collected in the labor force survey did not adequately capture the substantial improvements in welfare enjoyed by rural and estate households.

Unless the LFS expands its questionnaire to include ownership of consumer durables and housing conditions, and makes its sampling design and questionnaire consistent with the HIES, the HIES to LFS imputation cannot produce reliable poverty estimates in the rural and estate sectors. In the estate sector in particular, the only possible solution would be to add additional variables not currently collected in the HIES or the LFS, which could, for example, measure short-term fluctuations in the wages of estate workers.

In Sri Lanka, creating a new household survey is a feasible option to generate frequent and reliable poverty estimates in a cost effective way. While it would also be possible to alter the questionnaire of the labor force survey to ensure consistency with the HIES, this would require a substantial revision that would also create a structural break in Sri Lanka's labor statistics. To implement a new household survey, a short questionnaire can be fielded that includes the variables necessary to impute household expenditures accurately. If such a survey utilizes mobile technology and Computer Assisted Personal Interview (CAPI) techniques, survey implementation costs can be minimized while improving the precision of the survey-to-survey imputations.

An important lesson from this analysis is that there is no guarantee that survey-to-survey imputation works, and we know very little a priori about what variables are useful to explain both cross-sectional and inter-temporal variation. Theory can offer only broad guidance about which variables track changes in poverty or living standards closely, and it is therefore impossible to predict ex-ante whether and under what conditions survey-to-survey imputation performs well. We now know that imputing poverty based on labor market outcomes and asset indicators worked well for a particular period in Morocco, but that

using labor market outcomes alone failed during a particular period in Sri Lanka. This emphasizes the importance of verifying that a particular survey-to-survey imputation model performs well, using as many past periods of consumption data as possible, before using it to generate new poverty projections.

We propose four practical steps to assess whether imputing from one survey to another across different time periods is feasible. These are: Closely examining the surveys to ensure that the questions are compatible, examining whether the sampling strategy of the two surveys is comparable, verifying that the means of the proxy poverty measures are balanced between the two surveys, and ensuring that common variables explain a large portion of changes in household expenditure per capita. For survey-to-survey imputation to succeed, both surveys must contain a set of common variables based on substantively identical questions that track changes in household expenditure per capita closely. Therefore, it is important to better understand which variables are strongly correlated with changes in welfare and poverty in a variety of contexts. For example, in urban Sri Lanka between 2006 and 2009, a small set of labor outcomes captures changes in household welfare quite well, whereas in rural and estate sectors, information on assets are necessary to adequately capture changes in poverty.

Looking forward, additional empirical analysis can examine which types of variables best track changes in poverty in a variety of developing country contexts, to better inform future survey design. Future work can also further explore the extent to which ex-post adjustments to survey weights can make imputations across surveys more accurate. Finally, efforts could be made to develop more sophisticated diagnostic techniques to better identify which set of existing common variables can best track changes in poverty.

References

- Ahmed, F., C. Dorji, S. Takamatsu, and N. Yoshida. 2014. "Conducting a Hybrid Survey to Improve the Reliability and Frequency of Poverty Statistics in Bangladesh." Mimeo, World Bank, Washington, DC.
- Allwine, M., H. Uematsu, S. Takamatsu, and N. Yoshida. 2013. "A Technical Note on Lesotho Poverty Measurement," Mimeo
- Beegle, K., J. De Weerd, J. Friedman, and J. Gibson. 2010. "Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania." Policy Research Working Paper Series 5501, World Bank, Washington, DC.
- Christiaensen, L., P. Lanjouw, J. Luoto, and D. Stifel. 2012. "Small Area Estimation-Based Prediction Methods to Track Poverty: Validation and Applications." *Journal of Economic Inequality* 10 (2): 267–97.
- Deaton, A. and J. P. Dreze. 2002. "Poverty and Inequality in India: A Reexamination" *Economic and Political Weekly*, September 7, 2002.
- Doudich, M., A. Ezzrari, R. Van der Weide, and P. Verme. 2013. "Estimating quarterly poverty rates using labor force surveys : a primer," *Policy Research Working Paper Series* 6466, The World Bank.
- Elbers, C., J. O. Lanjouw, and P. Lanjouw. 2003. "Micro-Level Estimation of Poverty and Inequality." *Econometrica* 71 (1): 355–64.
- Harttgen, K., S. Klasen, and S. Vollmer. 2012. "An African Growth Miracle? Or: What do Asset Indices Tell Us about Trends in Economic Performance?" *Poverty, Equity, and Growth Discussion Paper* 109, Courant Research Centre PEG.
- Kijima, Y. and P. Lanjouw. 2003. "Poverty in India during the 1990s - a regional perspective," *Policy Research Working Paper Series* 3141, The World Bank.
- Ravallion, M. 1996. "How Well Can Method Substitute for Data? Five Experiments in Poverty Analysis." *World Bank Research Observer* 11(2): 199–21.
- Revilla, J., R. Katayama, N. Yoshida, and L. Fox. 2010. "Analysis of Poverty Trends using Poverty Mapping (Small Area Estimation-Based) Methods," available in *Zambia Poverty Assessment: Stagnant Poverty and Inequality in a Natural Resource-Based Economy*, 2012, World Bank, WashingtonDC, USA.
- Stifel, D., and L. Christiaensen. 2007. "Tracking Poverty Over Time in the Absence of Comparable Consumption Data." *World Bank Economic Review* 21 (2): 317–41.
- Tarozzi, A. and A. Deaton. 2009. "Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas". *Review of Economics and Statistics*, 91(4), 773-792.
- Tarozzi, A. 2011. "Can Census Data Alone Signal Heterogeneity in the Estimation of Poverty Maps?" *Journal of Development Economics* 95(2), 170-185.

Appendix

Table A1-1: HIES-HIES Imputation from 2009/10 to 2006/07 - Urban

Description	Coefficient	Std. Err.	Mean of X in	
			2009/10	2006/7
Intercept	8.3915	0.0422		
Boats	0.4718	0.1037	0.004	0.012
Bus lorry	0.2791	0.0575	0.014	0.015
Motor car van	0.4379	0.0254	0.120	0.122
Computers	0.1343	0.0179	0.263	0.176
Cookers	0.2355	0.0156	0.772	0.745
Dependency ratio (age -14, 64+ to HH size)	0.0822	0.0281	0.314	0.306
Employed ratio to household size	0.2565	0.0314	0.342	0.350
Electric Fans	0.1087	0.0173	0.804	0.785
Fridge	0.1208	0.0166	0.656	0.604
household size	-0.3224	0.0122	4.3	4.4
household size squared	0.0172	0.0011	21.7	22.4
Head's education completed prim	0.0427	0.0163	0.445	0.408
Head's education passed OL	0.1101	0.0219	0.186	0.221
Head's education passed AL	0.2102	0.0241	0.199	0.176
Telephone	0.1291	0.0146	0.554	0.503
Floor>750 sqft	0.0927	0.0142	0.458	0.409
Motor Bicycle	0.1408	0.0167	0.179	0.150
Only self-income	-0.0445	0.0157	0.180	0.192
Own Other water	0.0814	0.0204	0.913	0.885
Radio	0.0598	0.0157	0.802	0.849
roof: concrete	0.1125	0.021	0.110	0.083
Three wheeler	0.1626	0.0253	0.066	0.057
VCD	0.0634	0.0139	0.517	0.448
Wall: pressed soil blocks	-0.1703	0.062	0.009	0.013
Washing machine	0.1386	0.0185	0.317	0.283

Source: Authors' calculations using HIES 2006/07 and 2009/10

TableA1-2: HIES-HIES Imputation from 2009/10 to 2006/07 - Rural

Description	Coefficient	Std. Err.	Mean of X in	
			2009/10	2006/7
Intercept	8.3055	0.0396		
Head's age	-0.0026	0.0003	51.497	51.164
Bicycle	0.0366	0.0079	0.388	0.430
Bus lorry	0.325	0.0385	0.021	0.018
Motor car van	0.2973	0.0206	0.056	0.053
Computers	0.1213	0.0139	0.120	0.059

Cookers	0.1386	0.0109	0.351	0.320
Dependency ratio (age -14, 64+ to HH size)	-0.0526	0.0177	0.323	0.318
Employed ratio to household size	-0.1521	0.0321	0.379	0.391
Electric Fans	0.0782	0.0097	0.488	0.422
Fishing nets	0.1948	0.0526	0.005	0.004
Flat house	0.0843	0.0334	0.015	0.014
Mud for floor	-0.101	0.0132	0.093	0.121
Fridge	0.065	0.0107	0.403	0.338
head major industry: Inactive	0.0464	0.0109	0.288	0.274
household size	-0.2787	0.0086	3.9	4.0
household size squared	0.0145	0.0009	17.8	18.5
Head's education completed prim	0.0501	0.0094	0.480	0.442
Head's education passed OL	0.0905	0.0145	0.145	0.150
Head's education passed AL	0.1504	0.0194	0.100	0.093
Telephone	0.0908	0.0085	0.495	0.304
Floor>750 sqft	0.0926	0.0086	0.408	0.386
Motor Bicycle	0.1298	0.0091	0.281	0.220
Maximum household education passed OL	0.0525	0.0107	0.245	0.245
Maximum household education passed AL	0.0986	0.0125	0.272	0.257
Only paid income	-0.0235	0.008	0.455	0.473
Own Other water	0.0514	0.0091	0.760	0.691
Per capita wage index	0.006	0.0006	19.809	17.312
Sewing Machine	0.0365	0.0088	0.451	0.447
Pesticider	0.0678	0.0186	0.048	0.043
Three wheeler	0.1037	0.0163	0.074	0.048
Has own or shared toilet	0.1394	0.0276	0.983	0.976
TV	0.079	0.0109	0.818	0.777
Unemployed ratio to household size	-0.1302	0.0302	0.052	0.048
VCD	0.0888	0.0091	0.346	0.259
Wall: pressed soil blocks	-0.1099	0.0182	0.041	0.051
Washing machine	0.1096	0.0148	0.121	0.088

Source: Authors' calculations using HIES 2006/07 and 2009/10

TableA1-3: HIES-HIES Imputation from 2009/10 to 2006/07 - Estate

Description	Coefficient	Std. Err.	Mean of X in	
			2009/10	2006/7
Intercept	8.3247	0.0661		
Head's age	-0.0032	0.0006	49.571	49.514
Bus lorry	0.7342	0.1767	0.004	0.005
Motor car van	0.4918	0.1786	0.011	0.011
Computers	0.242	0.0772	0.022	0.014
Cookers	0.1162	0.0311	0.115	0.106
Dependency ratio (age -14, 64+ to HH size)	-0.1482	0.0417	0.324	0.328
Employed ratio to household size	-0.4561	0.1279	0.449	0.449
Electric Fans	0.112	0.0314	0.114	0.064
Mud for floor	-0.0803	0.0221	0.160	0.219
head major industry: Industry	-0.0694	0.032	0.073	0.064
household size	-0.2063	0.0182	4.182	4.207
household size squared	0.011	0.0017	20.598	21.409
Head's education passed AL	0.3328	0.1072	0.023	0.018
Telephone	0.1503	0.0177	0.389	0.097
Floor>750 sqft	0.1359	0.0488	0.059	0.060
Line room/Row house	-0.047	0.0172	0.640	0.693
Motor Bicycle	0.1694	0.0466	0.040	0.027
Maximum household education completed primary	0.083	0.0205	0.620	0.602
Maximum household education passed OL	0.1332	0.034	0.090	0.075
Maximum household education passed AL	0.2522	0.0546	0.062	0.051
Only self-income	-0.0682	0.0404	0.055	0.050
Per capita wage index	0.0224	0.0039	14.341	11.790
TV	0.0885	0.0194	0.705	0.617
Unemployed ratio to household size	-0.1427	0.0669	0.054	0.061
Source: Authors' calculations using HIES 2006/07 and 2009/10				

Table A2-1: HIES-LFS Imputation from 2006 to 2009 - Urban

Description	Coefficient	Std. Err.	Mean of X in	
			HIES 2006	LFS 2009
Intercept	8.5211	0.0755		
Head's ethnicity=SL Tamil	-0.1785	0.0375	0.119	0.122
Female head	-0.1193	0.0323	0.266	0.288
Head's major industry: Inactive	0.2094	0.0328	0.351	0.371
Head's major industry: Agriculture	0.3511	0.1234	0.014	0.017
Household size	-0.1731	0.0198	4.360	4.066
Household size squared	0.0071	0.0016	22.901	19.837
Head's education completed prim	0.2568	0.0335	0.411	0.421
Head's education passed OL	0.5054	0.0452	0.240	0.208
Head's education passed AL	0.8625	0.0559	0.162	0.228
Maximum household education completed primary	-0.1404	0.0344	0.317	0.313
Maximum household education passed AL	0.1107	0.0399	0.352	0.438
Only paid income	-0.2354	0.0263	0.567	0.513
Per capita wage index	0.0078	0.0009	20.501	25.951

Source: Authors' calculations using HIES 2006 and LFS 2009

Table A2-2: HIES-LFS Imputation from 2006/07 to 2009/10 - Rural

Description	Coefficient	Std. Err.	Mean of X in	
			HIES 2006	LFS 2009
Intercept	8.2095	0.0451		
Dependency ratio (age -14, 64+ to HH size)	-0.0627	0.0309	0.321	0.319
Female head	-0.0722	0.0171	0.222	0.241
Head's major industry: Agriculture	-0.1536	0.0176	0.242	0.238
Head's major industry: Industry	-0.0536	0.0186	0.181	0.166
Household size	-0.1935	0.0164	4.040	3.857
Household size squared	0.0099	0.0015	18.797	17.324
Head's education completed prim	0.1531	0.0161	0.438	0.462
Head's education passed OL	0.3336	0.0251	0.151	0.142
Head's education passed AL	0.4851	0.0364	0.100	0.100
Maximum household education completed primary	0.1732	0.0313	0.428	0.441
Maximum household education passed OL	0.3671	0.0352	0.249	0.237
Maximum household education passed AL	0.5551	0.0377	0.264	0.257
Only paid income	-0.3477	0.0327	0.468	0.425
Only self-employed income	-0.1443	0.0327	0.281	0.295
paid and self-employed income	-0.3087	0.0412	0.155	0.151
Per capita wage index	0.0126	0.0008	17.306	20.586
Unemployed ratio to household size	-0.2005	0.0557	0.048	0.038

Source: Authors' calculations using HIES 2006 and LFS 2009

Table A2-3: HIES-LFS Imputation from 2006/07 to 2009/10 - Estate

Description	Coefficient	Std. Err.	Mean of X in	
			HIES 2006	LFS 2009
Intercept	8.0418	0.061		
Dependency ratio (age -14, 64+ to HH size)	-0.2563	0.0555	0.319	0.340
Head's major industry: Agriculture	0.0799	0.0249	0.522	0.487
Household size	-0.1645	0.0175	4.280	4.078
Household size squared	0.0073	0.0014	22.552	20.188
Head's education completed prim	0.0901	0.0281	0.259	0.267
Head's education passed OL	0.1653	0.09	0.024	0.030
Head's education passed AL	0.5357	0.2198	0.008	0.027
Maximum household education passed AL	0.2942	0.0812	0.039	0.067
Only self-employed income	0.2753	0.059	0.059	0.058
Per capita wage index	0.0019	0.0022	11.951	14.684

Source: Authors' calculations using HIES 2006 and LFS 2009

Table A3-1: HIES-LFS Imputation from 2009 to 2006 - Urban

Description	Coefficient	Std. Err.	Mean of X in	
			HIES 2009	LFS 2006
	8.5841	0.0723		
Female head	-0.064	0.0252	0.289	0.249
Head's major industry: Agriculture	0.1372	0.1027	0.014	0.008
Head's major industry: Industry	-0.1095	0.0321	0.140	0.180
Household size	-0.2283	0.0171	4.281	4.340
Household size squared	0.0097	0.0013	22.251	22.097
Head's education completed prim	0.0966	0.0321	0.439	0.381
Head's education passed OL	0.3449	0.044	0.189	0.221
Head's education passed AL	0.5144	0.0507	0.190	0.225
Maximum household education completed primary	0.1942	0.0656	0.334	0.319
Maximum household education passed OL	0.3716	0.0693	0.266	0.270
Maximum household education passed AL	0.5226	0.0726	0.367	0.395
Only paid income	-0.1982	0.027	0.577	0.474
paid and self-employed income	-0.1195	0.049	0.117	0.184
Per capita wage index	0.0053	0.0008	23.940	21.240

Source: Authors' calculations using HIES 2009 and LFS 2006

Table A3-2: HIES-LFS Imputation from 2009 to 2006 - Rural

			Mean of X in	
			HIES 2009	LFS 2006

Description	Coefficient	Std. Err.	HIES	LFS
			2009	2006
	8.3007	0.0477		
Dependency ratio (age -14, 64+ to HH size)	-0.1322	0.031	0.321	0.312
Employed ratio to household size	-0.3526	0.0604	0.379	0.425
Head's ethnicity=Sinhalase	0.0626	0.0227	0.917	0.936
Head's major industry: Agriculture	-0.0815	0.0192	0.226	0.242
Head's major industry: Industry	-0.0384	0.0184	0.163	0.184
Household size	-0.1694	0.0161	3.918	3.915
Household size squared	0.008	0.0015	17.803	17.816
Head's education completed prim	0.1634	0.0161	0.479	0.450
Head's education passed OL	0.3437	0.0251	0.153	0.138
Head's education passed AL	0.474	0.032	0.102	0.095
Maximum household education completed primary	0.0713	0.0332	0.425	0.443
Maximum household education passed OL	0.2206	0.0371	0.245	0.243
Maximum household education passed AL	0.369	0.0387	0.277	0.243
Only paid income	-0.2724	0.0317	0.468	0.408
Only self-employed income	-0.1162	0.0326	0.281	0.313
paid and self-employed income	-0.2121	0.0416	0.130	0.170
Per capita wage index	0.0133	0.001	19.891	18.304
Unemployed ratio to household size	-0.2795	0.0506	0.056	0.041

Source: Authors' calculations using HIES 2009 and LFS 2006

Table A3-3: HIES-LFS Imputation from 2009 to 2006 - Estate

Description	Coefficient	Std. Err.	Mean of X in	
			HIES 2009	LFS 2006
Intercept	8.4754	0.0953		
Age	-0.0036	0.0011	49.916	48.331
Dependency ratio (age -14, 64+ to HH size)	-0.2115	0.0534	0.328	0.307
Employed ratio to household size	-0.7105	0.1609	0.450	0.487
Head's ethnicity=Indian Tamil	0.047	0.0244	0.526	0.654
Female head	-0.1371	0.0293	0.233	0.276
Head's major industry: Inactive	0.1093	0.0386	0.234	0.234
Household size	-0.2273	0.0227	4.148	4.136
Household size squared	0.0133	0.0022	20.207	20.651
Head's education passed OL	0.2071	0.0889	0.017	0.015
Head's education passed AL	1.0965	0.2522	0.020	0.012
Maximum household education completed primary	0.1514	0.029	0.594	0.612
Maximum household education passed OL	0.1349	0.0494	0.102	0.084
Maximum household education passed AL	0.5783	0.0723	0.058	0.054
Per capita wage index	0.028	0.0045	14.272	12.580

Source: Authors' calculations using HIES 2009 and LFS 2006

Table A4: Estimates and Standard Errors With or Without Location Dummy Variables: HIES-LFS Imputation from 2006 to 2009

Direct estimation			Imputation - 2009 (ELL)					
			District dummies		Province dummies		No location dummies	
	2006	2009	Variables only from LFS	With wage data	Variables only from LFS	With wage data	Variables only from LFS	With wage data
National	15.3 (0.7)	8.6 (0.5)	12.4 (0.5)	11.1 (0.5)	12.4 (0.4)	10.9 (0.4)	12.3 (0.5)	11.1 (0.5)
Urban	7.4 (1.3)	6.1 (1.1)	4.4 (1.5)	4.1 (1.1)	4.5 (1.5)	3.2 (1.3)	4.4 (1.4)	4.1 (1.4)
Rural	15.4 (0.8)	8.9 (0.6)	12.4 (0.8)	10.9 (0.8)	12.3 (0.8)	10.7 (0.8)	12.3 (0.8)	10.8 (0.8)
Estate	33.6 (3.0)	10.4 (1.7)	29.3 (2.9)	30.0 (3.2)	33.2 (2.9)	31.8 (3.2)	31.5 (3.3)	31.7 (3.1)

Source: Authors' estimations using HIES 2006/7 and 2009/10 data.

Notes: Tables A2-1 to A2-3 list the consumption models except for the used location dummy variables. The figures in parentheses are cluster-robust standard errors for the direct estimations and the standard errors that include the modeling and sampling errors for imputation. The results for "No location dummies" in the last columns are the same as those in Table 3.