High Frequency Poverty Monitoring in the Face of Data Deprivation

Peter Lanjouw (Vrije Universiteit Amsterdam)
Nobuo Yoshida (World Bank)

LIS Summer Lecture
July 6, 2021

Introduction

- Development community has committed itself to the ambitious goal of “ending” poverty
  - 2015 – Sustainable Development Goals (SDGs) and World Bank twin goals
- Ability to monitor progress in poverty reduction is critical
  - High frequency data are needed
  - Urgency has increased following COVID-19 pandemic
- But data needed for this purpose are scarce
  - Particularly in the most poverty-stricken regions and countries
  - Major information gaps: “data deprivation”
- This lecture discusses options to overcome this constraint
  - Focus is on the developing country context

Overcoming data deprivation

- Two options:

1. Conduct more frequent household surveys
    - Revisit current practice in the way surveys are fielded
    - Compromises?
2. Implement approaches involving survey-to-survey (S2S) imputation
    - Focus on two approaches that have recently seen acceptance and adoption in WB's Global Poverty Monitoring effort:
        - SWIFT-2
        - Rapid Consumption Survey (RCS)
    - What are relative advantages?
    - What are underlying assumptions?
    - What is the experience to date?

Revisiting data collection

- Living Standards Measurement Study (LSMS) program was launched in the 1980s
  - Aimed at collecting <u>integrated</u> household survey data
    - Detailed consumption/income data combined with wide range of household characteristics
- LSMS program proposed guidelines for the collection of survey data (Grosh and Muñoz, 1996, World Bank, 2000)
  - LSMS paid close attention to concerns about non-sampling error
    - Complexity, significant training requirements
  - Recognition that LSMS surveys are costly
- RECOMMENDATION: keep sample size small (2000-5000 households)
  - Acknowledged that this implied limited scope for disaggregation
  - Justification: limit non-sampling error; contain cost

Early LSMS surveys tended to follow sample size guidelines

Table 1. LSMS Surveys in the first 15 years of the LSMS initiative

Country	Year of first survey	Sample size
Albania	1996	1,500
Algeria	1995	5,900
Armenia	1996	4,920
Azerbaijan	1995	2,016
Bolivia	1989	4,330-9,160
Brazil	1996	5,000
Bulgaria	1995	2,000
Cambodia	1997	6,010
China (Hebei and Liaoing only)	1995	800
Côte d'Ivoire	1985	1,600
Ecuador	1994	4,500
Ghana	1987/88	3,200
Guyana	1992/93	1,800
Jamaica	1988	2,000-4,400
Kazakhstan	1996	2,000
Kyrgyz Republic	1994	2,100
Mauritania	1988	1,600
Morocco	1991	3,360-4,800
Nepal	1996	3,373
Nicaragua	1993	4,454
Pakistan	1991	4,800
Panama	1997	4,945
Paraguay	1997/98	5,000
Peru	1985	1,500-3,623
Romania	1994/95	31,200
South Africa	1993	8,850
Tajikistan	1999	2,000
Tanzania-Kagera	1991	800
Tanzania-Human Resource Development Survey	1993	5,200
Tunisia	1995/96	3,800
Turkmenistan	1997	2,350
Vietnam	1992/93	4,800-6,000

Source: World Bank (2000)

Over time sample sizes have drifted upwards

Table 2. Summary statistics of the latest household surveys in sub-Saharan Africa

Country Group	sample size (mean)	Survey year		
		Oldest	Median	Newest
All	10,700	2008	2015	2019
Blend	12,079	2011	2015	2019
IBRD	9,125	2014	2016.5	2018
IDA	10,835	2008	2015	2018
FCS	8,952	2008	2014	2019

Note: A total number of countries is 46.

Source: Global Monitoring Database (2021) World Bank

Revisiting new data collection, cont.

- Return to smaller samples can be expected to:
  - Take full advantage of new technological developments (CAPI, GPS, etc.)
  - Reduce cost (typical cost $500K-$1.5m)
  - Reduce logistical and administrative complexity
  - Facilitate increased frequency of surveys
  - Improve data quality

- HOWEVER:
  - Small sample size limits analytical value and practical relevance
    - limited profiling; etc.
  - Fully integrated LSMS remains complex and costly

Alternative: Survey-to-survey (S2S) imputation

- Some accumulated experience with S2S approaches
  - Deaton and Drèze (2002), Kijima and Lanjouw (2003), Christiaensen et al (2011), Douidiche et al (2013), etc.
- Basic idea:
  1. Predict income/consumption from one survey ("training data") into another ("output data");
  2. Estimate poverty in "output" survey based on predicted income/consumption
     - Typical application: parameter estimates from a survey in time *t-1* used to predict income/consumption in survey from time *t*
     - Approach has been proposed to: overcome issues of non-comparability of income measure; doubts regarding appropriate CPI adjustments; "knitting" together disparate datasts (DHS, employment surveys, budget surveys, etc.)
  - KEY ASSUMPTION: parameter estimates are stable between surveys ("stability assumption")

Survey to survey (S2S) imputation and model stability condition

	training data		*output data*
Dataset	$\{lny_{ht}, x_{ht}\}_{h=1}^{H}$		$\{x_{ho}\}_{h=1}^{H}$
Relationship	$lny_{ht} = x_{th}\beta_t + u_{ht}$	Model stability holds $\beta_t = \beta_o, \sigma_t = \sigma_o$	$lny_{ho} = x_{ho}\beta_o + u_{ho}$
Imputation			$\widetilde{lny}_{ho}^{r} = x_{ho}\tilde{\beta}^{r} + \tilde{u}_{ho}^{r}$
Imputed data			$\{\{\widetilde{lny}_{ho}^{r}\}_{h=1}^{H}\}_{r=1}^{100}$

Past Application of S2S

- It is common to assume welfare measures are comparable.
- But growing awareness of the fragility of this assumption
  - Beegle et al (2012): field experiment in Tanzania showing sensitivity of distributional indicators to subtle changes in consumption definition
  - Matthiasen and Wold (2019): study in Malawi shows that differences in survey implementation, sample size, etc., can contribute to non-comparability
  - Gibson et al (2008) question the suitability of official inflation statistics in capturing cost-of-living changes over time
- Christiaensen et al (2011) ask how well S2S "corrects" for potential non-comparability

Christiaensen et al (2011) findings

- Consider Vietnam and China as settings where stability assumption might not be expected to hold (major structural transformation).
  - Test S2S using data that are comparable but are treated as non-comparable
  - Assess performance of food and non-food expenditures as predictors
  - Findings:
  1. S2S can be successful, but results do vary with model specifications
     - Geographic indicators, housing and consumer durables generally perform well
     - No theory to identify “right” specification
  2. Models based on sub-components of consumption/income less successful
     - Stability of Engel Curves is uncertain

Christiaensen et al (2011) findings: Russia.

- Official Russian estimates suggest poverty rose sharply between 1994-1998 and then dropped back by 2003
  - Financial crisis of 1998
- But nutritional indicators and subjective wellbeing indicators do not corroborate these findings (Stillman and Thomas, 2008, and Gibson et al 2008)
- Gibson et al (2008) suggest official CPI for urban Russia was severely overstated
- What does S2S say about poverty trends in Russia?

Russia: inflation overstated?

Poverty Headcount	Observed Levels		SAE Predicted Levels in 1998		
Included in the model	Period 1	Period 2	(3)	(4)	(5)
Non-consumption assets					
Geographic			x	x	x
Demographic			x	-	x
Education/Profession			x	-	-
Housing quality			x	x	-
Consumer durables			x	x	-
Subjective perception of quality of life			-	-	x
Region	1994	1998			
National	**11.4**	**33.8**	14.1	12.7	13.2
	0.6	*1.1*			
Rural	**13.1**	**34.8**	22.4	18.2	16.9
	1.3	*2.0*			
Urban	**10.6**	**33.3**	18.8	17.4	11.5
	0.7	*1.3*			
	1994	*2003*			
National	**11.4**	**11.1**	8.5	8.4	9.2
	0.6	*0.6*			
Rural	**13.1**	**17.4**	9.9	13.1	12.4
	1.3	*1.5*			
Urban	**10.6**	**8.1**	9.2	11.2	7.4
	0.7	*0.6*			
No. of times difference NOT statistically different			1	0	1
average absolute difference			13.8	13.6	11.5
# observed poverty ≥ predicted poverty			5	5	6
# observed poverty < predicted poverty			1	1	0

Does model stability hold?

- Plausibility of model stability assumption remains highly context specific
  - If training data are collected before a crisis, a model developed from the training data might not reflect the relationship between household expenditures and poverty proxies in the output data
    - While assumption seemed to hold in Vietnam and China, experiments with data from Afghanistan and Gaza data are less encouraging
  - If the gap between training and output dataset grows, assumption becomes less tenable
  - Specification of model seems to matter; but no clear theory to guide how to ensure assumption holds (Christiaensen et al, 2011)
- Model stability more plausible when training and output datasets are fielded simultaneously
- We consider two approaches employing S2S based on concurrent datasets
  - aim to assure that model stability holds.
  - The approaches we consider were cleared for adoption in the WB Global Poverty Monitoring effort

Option 1: SWIFT 2.0, based on SWIFT

- Original SWIFT approach
  - Identify an existing household survey to function as training dataset;
  - Specify a model to predict consumption based on a parsimonious set of covariates
  - Implement a new survey to collect only data on covariates
  - Impute consumption into new dataset based on model in training dataset
  - Estimate poverty based on imputed consumption

  - Caveats:
    - Original SWIFT is predicated on stability assumption holding
    - are data collection practices the same between training and output datasets?

Option 1, cont.: SWIFT 2.0

- Collect a full, nationally-representative, household survey
- BUT: collect consumption data only from a sub-set of households
- Employ the subset with consumption data as training dataset
  - Estimate a consumption model based on covariates available in the full sample
  - Predict consumption into the sample of households without direct consumption measures
  - Estimate poverty based on combination of imputed and directly measured consumption in the full dataset
- RATIONALE: significant cost and time savings can be achieved by collecting consumption data from only a subset of households
  - Collecting consumption data is particularly complex and time consuming
  - Extent of cost savings will be a function of size of subset

SWIFT 2.0		
	Subset	**Rest**
modules	Cons	Cons
	Non-cons	Non-cons

SWIFT 2.0: experience in Zimbabwe

- Zimbabwe’s most recent survey was in 2017
- Hyperinflation in 2019 raised alarms as to impact on poverty
- Resource constraints prevented collection of new survey
- Zimbabwe National Statistics Agency and WB agreed to implement SWIFT 2.0
  - May-June of 2019, survey of 2710 households conducted
  - Only 509 households were fielded full consumption module
  - S2S was applied from the consumption households to the balance of surveyed households
  - Poverty was estimated over full dataset:
  - Headcount: Urban 24.3 (4.4)
Rural 72.0 (2.5)
- Estimates accepted for WB Global Poverty Monitoring

Comparison between SWIFT 2.0 and the traditional approach

- SWIFT 2.0 introduces imputation errors in addition to sampling error
  - If we want to achieve the same level of precision in poverty measurement as with a standard data collection, SWIFT 2.0 will require a larger overall sample size
  - Increasing the sample size raises the survey implementation cost
- SWIFT 2.0 can reduce interview costs but not transportation costs
  - If interview costs are marginal compared to transportation costs, SWIFT 2.0 does not reduce the survey implementation costs

Comparison between standard data collection and SWIFT 2.0 using Fujii and van der Weide (2020)

- To outperform the traditional approach, SWIFT 2.0 needs a significant reduction in interview time
- Phone survey or local enumerator approach could make SWIFT 2.0 very attractive
- Inserting SWIFT questions to another survey makes SWIFT 2.0 very attractive

Relative transportation cost

Time saving

Size of imputation errors

SWIFT interview costs to standard data collection	Transportation cost to interview cost	Household level imputation error to sampling error	Cluster level error to sampling error	Proportion of SWIFT 2.0's cost to standard data collection
0.06	4	0.4	0.6	0.776
0.36	4	0.4	0.6	0.944
0.06	2	0.4	0.6	0.735
0.06	0.1	0.4	0.6	0.597

Option 2: Rapid Consumption Survey

- Introduced by Mistiaen and Pape (2018), Pape and Wolfgang (2019)
- Collect a full, nationally-representative household survey
- Split the sample into multiple subsamples
- Each subsample collects non-consumption data, a subset of core consumption questions, and one subset of additional consumption questions
- The subsamples differ only in the subset of additional consumption questions
- Consumption expenditures from dropped partitions are imputed based on models developed in other subsamples
- Total consumption is then calculated for all subsamples
- Poverty estimates are generated on the basis of total per capita consumption
- EXAMPLE: Suppose an RCS with three subsamples

Rapid Consumption Survey: Example

RCS			
Modules	Group 1	Group 2	Group 3
Cons	P3	P3	P3
	P2	P2	P2
	P1	P1	P1
	Core	Core	Core
Non-cons	Core	Core	Core

Observations on RCS

- Time savings are greater if there are many subsamples (each collecting only a small amount of consumption information)
- Note, however, as subsamples increase in number, imputation models are based on smaller sample sizes, leading to less successful prediction models.
- Practical experience accrues from Somalia (2017) and South Sudan (2016/17) – See Mistiaen and Paper (2018) and Pape and Wolfgang (2019)
- Note also, imputation models for the subsamples need to anticipate being able to predict 0 consumption for some households
  - Takamatsu et al. (2021) addressed the issue of negative numbers and re-estimated poverty numbers for Somalia and South Sudan, which were adopted for WB Global Poverty Monitoring purposes.

Comparing SWIFT 2.0 and RCS

RCS			
Modules	Group 1	Group 2	Group 3
Cons	P3	P3	P3
	P2	P2	P2
	P1	P1	P1
	Core	Core	Core
Non-cons	Core	Core	Core

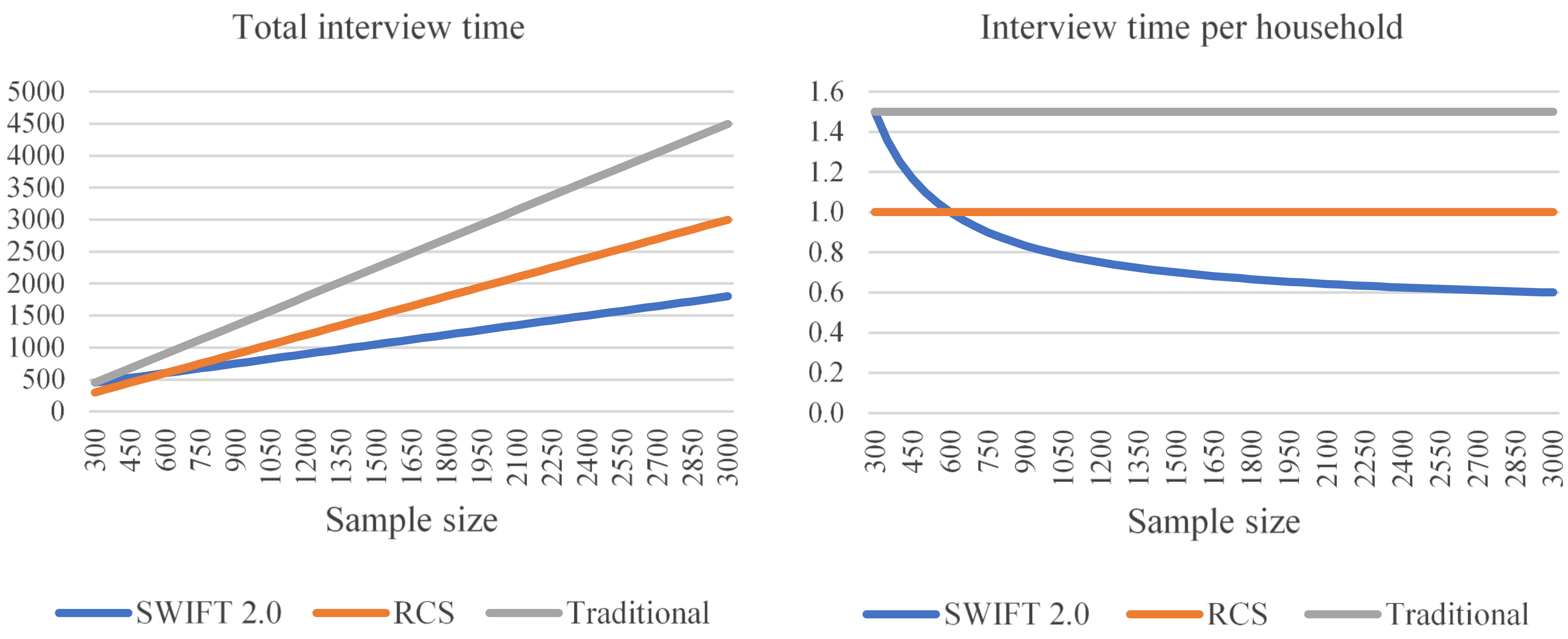
SWIFT 2.0		
Modules	Group 1	Group 2
Cons	P3	P3
	P2	P2
	P1	P1
	Core	Core
Non-cons	Core	Core

Comparing Traditional approach, SWIFT 2.0 and RCS in interview time

- Suppose collection of a standard household survey involves 90 minutes per household interviewed
  - 30 minutes non-consumption data
  - 60 minutes consumption data
- Suppose SWIFT 2.0 is assumed to take 90 minutes for consumption sub-sample; 30 minutes for rest of households
- Suppose RCS takes 60 minutes for all households

Comparison of interview time

Figure 3. Comparisons of interview time (hours) between RCS and SWIFT 2.0



Source: Authors' calculations based on assumptions described in the main text.

Concluding remarks (1)

- High cost of data collection constrains frequent poverty monitoring
- Three approaches were discussed in this paper
  - Standard data collection with a small samples (2500 to 4000)
  - SWIFT 2.0
  - RCS
- Each methodology has pros and cons
  - SWIFT 2.0 will tend to be more cost-effective than RCS as the sample size becomes bigger
  - Standard survey collection is more cost-effective than SWIFT 2.0 if transportation costs are high and time-saving of SWIFT 2.0 is limited.
  - Phone surveys and local enumerator approach can, in principle, favor SWIFT 2.0

Concluding remarks (2)

- Model stability is a crucial assumption underlying S2S
  - Model stability is not just needed for S2S
    - Many poverty projection methodologies face model stability issues
  - No reason to suppose it applies in general; difficult to test in practical applications
- SWIFT 2.0 and RCS are plausibly less exposed to this concern
  - But issues of logistics, questionnaire design, field implementation, etc. remain
    - SWIFT 2.0 implementation in Zimbabwe involved two distinct teams of interviewers and thus overcame many of these concerns
- Note: in practice assumption of comparability is commonly imposed
  - Growing awareness of both fragility of that assumption as well as of the gravity of consequences when it fails.