

# Using Survey-to-survey Imputation for More Frequent and Lower-cost Poverty Estimates

Evidence from Malawi

I. Yacoubou Djima

Joint with T. Kilic, H. Moylan, A. Palacios-Lopez, W. Vundru, P. Wollburg

Better Data for Better Jobs and Lives: Innovations in Survey Measurement in the Age of AI

WB LSMS & Data Academy - NU GPRI 30 / December 8/9, 2025



# Introduction



# Background

- Timely and accurate poverty statistics are essential for guiding policy and monitoring progress.
- Full consumption surveys conducted Face to Face (F2F) are costly and are not implemented infrequently.
- A growing literature shows Survey to Survey (S2S) imputation can fill poverty data gaps at low cost (Corral Rodas et al., 2025; Dang et al., 2019, 2025; Dang and Lanjouw, 2023; Mathiassen and Wold, 2021)
  - S2S imputation uses one “base” survey (with welfare measure) to predict poverty in later “target” surveys (lack direct welfare measure)
  - Update poverty estimates using recent surveys



- Most S2S applications assume comparability across surveys.
- In practice, target surveys are often **lighter** and more recently have relied on **Phone Surveys (PS)** interviews.
- Mode can affect reporting, item non-response, welfare measure
  - Castaing et al. (2025) : experiment in rural Nigeria shows PS yield systematically yield different answers than F2F —about 17–18% at the median across 20 outcomes
  - Abate et al. (2023) : experiment in Ethiopia shows lower capita consumption in PS compare to F2F
- The implications of mode effects for S2S-imputed poverty remain under-studied.
  - One exception is (Dang and Carletto, 2022) on labor outcomes in Tanzania, which finds encouraging results for F2F–phone imputation.



# This Study: Malawi Mixed-Mode Imputation Experiment

- Context:
  - Malawi IHS6 (2024–25), ~12,000 households interviewed F2F over 4 quarters.
- Design:
  - Panel phone surveys of ~2,000 Q1 households, re-interviewed in Q2, Q3, Q4 as “target” surveys.
  - Base survey: Q1: F2F IHS6
  - F2F IHS6 Q2, Q3, Q4, samples provide concurrent F2F “ground-truth” consumption.
  - Harmonized predictors across mode
- Contributions:
  - Assess mode effects on S2S-imputed poverty in a real-time monitoring setting.
  - Experimentally vary incentives to study attrition and non-sampling error in phone surveys.



# This Study: Malawi Mixed-Mode Imputation Experiment

- Context:
  - Malawi Integrated Household Survey 2024–25 (IHS6), ~12,000 households interviewed F2F over 4 quarters.
- Design:
  - Panel phone surveys of ~2,000 Q1 households, re-interviewed in Q2, Q3, Q4 as “target” surveys.
  - Base survey: Q1: F2F IHS6
  - F2F IHS6 Q2, Q3, Q4, samples provide concurrent F2F “ground-truth” consumption.
  - Harmonized predictors across mode
- Contributions:
  - **Assess mode effects on S2S-imputed poverty in a real-time monitoring setting.**
  - Experimentally vary incentives to study attrition and non-sampling error in phone surveys.



# Study Design



# Study Design

- Phone Survey Sample is a subsample from IHS6 Q1 (eligible for a phone survey: at least one member has a cell phone)
  - Balance tests: ▶ Household characteristics ▶ Dwelling characteristics ▶ Assets
  - Phone survey runs concurrently to F2F Q2, Q3, Q4 :  
▶ Monthly Interviews Count
- Phone Survey Questionnaire design
  - Roster Update
  - Housing characteristics
    - Q2: Ask over the phone again for Walls, Roof, Floor, Number of Rooms, Wash etc.
    - Q3 & Q4: Ask for confirmation about the recorded info during Q2 for dwelling characteristics
  - Assets
    - Q2, Q3, Q4: Ask over the phone again for various assets
  - Food consumption and non Food consumption : Phone has a subset list of items
  - Food insecurity experience scale



- Assess Mode Differences
  - Basic household characteristics
  - Housing characteristics
  - Assets
  - Food and non-food consumption
- S2S Imputation Exercise
  - Base Survey: Q1 F2F
  - Target Surveys: PS Q2, Q3, Q4
  - Comparison with F2F empirical rates, Predicted rates using Q1 as a base
  - Permutation of models with 4 categories of predictors
    - HH characteristics
    - Dwelling characteristics
    - Assets
    - Food Consumption

# Data



# Summary Statistics

	IHS6 Cross-section Face to Face					IHS6 Subsample Phone Survey		
	Q1–Q4	Q1	Q2	Q3	Q4	Q2	Q3	Q4
HH Head Age	41 (14)	42 (14)	41 (14)	40 (14)	41 (14)	42 (14)	42 (14)	42 (14)
HH Head Female	0.26 (0.44)	0.27 (0.44)	0.25 (0.44)	0.24 (0.44)	0.26 (0.44)	0.27 (0.44)	0.27 (0.44)	0.27 (0.44)
Head Education	2.36 (0.71)	2.31 (0.71)	2.40 (0.71)	2.36 (0.71)	2.36 (0.71)	2.30 (0.72)	2.31 (0.71)	2.30 (0.72)
Head work	0.97 (0.16)	0.98 (0.16)	0.97 (0.16)	0.98 (0.16)	0.97 (0.16)	0.97 (0.17)	0.97 (0.17)	0.97 (0.17)
Household size	4.38 (1.87)	4.53 (1.87)	4.52 (1.87)	4.16 (1.87)	4.29 (1.87)	4.82 (1.98)	4.93 (1.99)	5.02 (1.98)
Urban	0.22 (0.41)	0.18 (0.41)	0.29 (0.41)	0.17 (0.41)	0.23 (0.41)	0.19 (0.39)	0.19 (0.39)	0.19 (0.39)
N	9088	2332	2322	2168	2266	2158	2053	1932

Notes: Weighted means and standard deviations are reported in parentheses. Unit of observation is household. Sample is restricted to phone survey eligible households for F2F.

# Differences Across Survey Modes

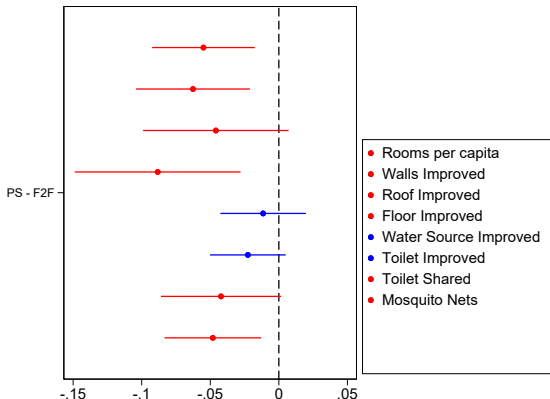


# Mode Differences - Basic Household Characteristics

	(1)	(2)	(3)	(4)	(5)
	Age	Female	Education	Work	HH Size
PS - F2F	0.895** (0.425)	0.020 (0.015)	-0.075** (0.038)	-0.004 (0.005)	0.594*** (0.067)
Mean Dep. Var.	41.316	0.260	2.342	0.973	4.601
N PS	6143	6143	6143	6143	6143
Observations	12899	12899	12899	12899	12899

*Notes: Weighted robust standard errors clustered at the EA level are reported in parentheses. Sample is restricted to phone survey eligible households for the F2F. We pooled PS samples and F2F Q2-Q4. Quarter fixed effects are included. Significance levels: \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$*

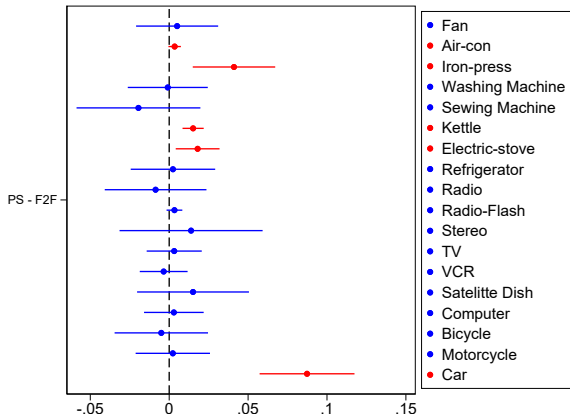
# Mode Differences - Dwelling Characteristics



**Figure:** Differences assessed via weighted regression comparing PS report in Q2 to F2F (Q2-Q4). F2F Sample is restricted to phone survey eligible households. 95% confidence intervals of the coefficients are plotted. Coefficients are red for significance level  $p < 0.1$

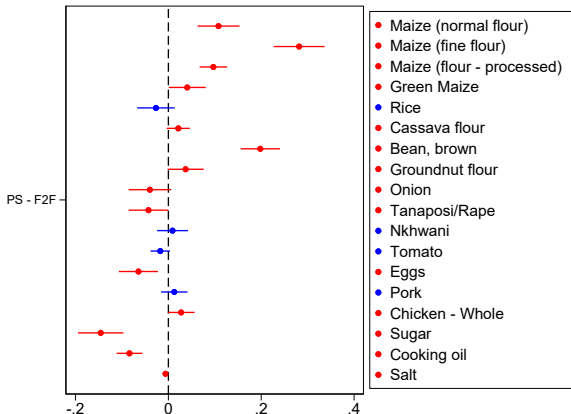
▶ PS Q2 vs F2F Q1

# Mode Differences - Assets



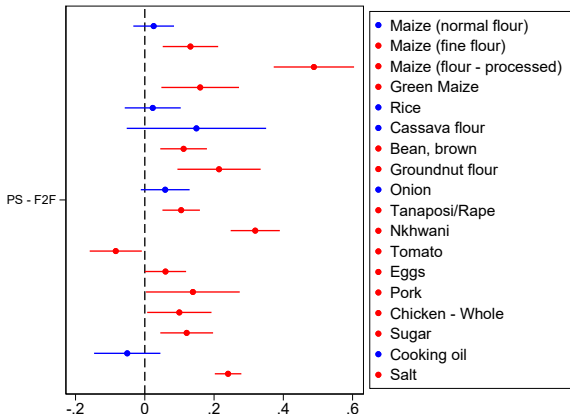
**Figure:** Differences assessed via weighted regression comparing PS report in Q2 to F2F (Q2-Q4). F2F Sample is restricted to phone survey eligible households. 95% confidence intervals of the coefficients are plotted. Coefficients are red for significance level  $p < 0.1$

# Mode Differences - Likelihood of consuming food items



**Figure:** Differences assessed via pooled weighted regression comparing PS report in (Q2-Q4) to F2F (Q2-Q4) with quarter dummies. F2F Sample is restricted to phone survey eligible households. 95% confidence intervals of the coefficients are plotted. Coefficients are red for significance level  $p < 0.1$

# Mode Differences - PC log value of food items consumed



**Figure:** Differences assessed via pooled weighted regression comparing PS report in (Q2-Q4) to F2F (Q2-Q4) with quarter dummies. F2F Sample is restricted to phone survey eligible households. 95% confidence intervals of the coefficients are plotted. Coefficients are red for significance level  $p < 0.1$

# Mode Differences - Summary

## ● Household and dwelling characteristics

- Updating roster affects household size significantly
  - Households appear to be more likely to add members
- Dwelling characteristics are different and less likely to be improved
  - Enumerator in F2F observe more improved dwelling conditions
- Assets are not significantly different overall
  - One notable item is car with large over-reporting difference
- More differences for items that can be observed in F2F

## ● Food consumption

- Households are more likely to report consuming food items with large differences for the first items (more calorific)
- Value over reported over the phone for most items
- Opposite direction of what is found in Abate et al. (2023)

- **Parameter stability risk**

- F2F vs. phone changes measurement of key predictors (roster, dwelling, assets), threatening stable welfare–predictor relationships.

- **Sampling and nonresponse bias**

- Phone surveys over-represent connected/better-off households; reweighting helps but is not sufficient alone.

- **Questionnaire/measurement differences**

- Shorter phone modules and self-reporting distort consumption and observational items.

# Imputation Across Survey Modes



# Imputation Exercise

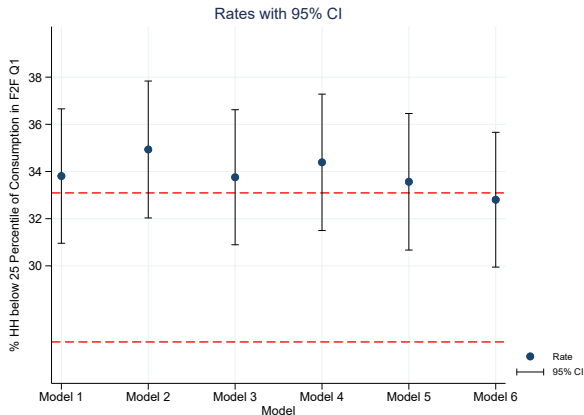
- Train model on F2F Q1 and predict PS Q2
- Models:
  - Model 1: Household characteristics
  - Model 2: Model 1 + dwelling characteristics
  - Model 3: Model 1 + assets
  - Model 4: Model 2 + assets
  - Model 5: Model 4 + PC Food
  - Model 6: Model 3 + PC Food

# Preliminary Imputation Results

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
ols	34.41 (1.45)	35.40 (1.47)	34.55 (1.46)	34.97 (1.47)	34.55 (1.49)	33.95 (1.48)
emp	33.81 (1.45)	34.93 (1.48)	33.76 (1.46)	34.39 (1.48)	33.56 (1.48)	32.80 (1.46)
"truth"	29.93 (1.61)	29.93 (1.61)	29.93 (1.61)	29.93 (1.61)	29.93 (1.61)	29.93 (1.61)
$R^2$	0.32	0.39	0.42	0.44	0.60	0.58
N	2152	2152	2152	2152	2152	2152

Notes: The "poverty" line is defined as the 25th percentile of per-capita expenditure in quarter 1 F2F. We train model on F2F Q1 and predict PS Q2. We model log of per-capita total expenditure. ols assumes normal distribution of errors. emp uses the empirical distribution of residuals to adjust predictions and uncertainty non-parametrically. Truth is the rate of households below the "poverty" line using F2F quarter 2. We restrict our estimations to phone owning households. Standard errors are bootstrapped with 1000 replications. Model 1: Household characteristics; Model 2: Model 1 + dwelling characteristics; Model 3: Model 1 + assets; Model 4: Model 2 + assets; Model 5: Model 4 + PC Food; Model 6: Model 3 + PC Food

# Preliminary Imputation Results



- Notes: The red dashed lines represent the 95% confidence interval of the true "poverty" rate. The "poverty" line is defined as the 25th percentile of per-capita expenditure in quarter 1 F2F. We train model on F2F Q1 and predict PS Q2. We model log of per-capita total expenditure. ols assumes normal distribution of errors. emp uses the empirical distribution of residuals to adjust predictions and uncertainty non-parametrically. Truth is the rate of households below the "poverty" line using F2F quarter 2. We restrict our estimations to phone owning households. Standard errors are bootstrapped with 1000 replications. Model 1: Household characteristics; Model 2: Model 1 + dwelling characteristics; Model 3: Model 1 + assets; Model 4: Model 2 + assets; Model 5: Model 4 + PC Food; Model 6: Model 3 + PC Food

# Conclusion



- **Mode effects affect key predictors of imputation models**
  - Household roster inflation
  - Mis-alignment of reported dwelling characteristics
  - Assets are overall consistent
  - Over-reporting of food consumption
- **However imputation exercise shows promising results**
  - Judicious imputation modeling achieves rate within 95% confidence of ground truth rate
- **Implication for S2S Modeling**
  - Build a parsimonious, mode-robust predictor set
  - Standardize distributions of mode-sensitive variables before modeling
- **Next Steps**
  - Uncover the mechanisms behind the mode differences
  - More imputation modeling including with ML

# Bibliography

- Abate, G. T., de Brauw, A., Hirvonen, K., and Wolle, A. (2023). Measuring consumption over the phone: Evidence from a survey experiment in urban ethiopia. *Journal of Development Economics*, 161:103026.
- Corral Rodas, P. A., Ham, A., Lanjouw, P., Lucchetti, L. R., and Stemmler, H. (2025). Stress testing survey to survey imputation: Understanding when poverty predictions can fail. Policy Research Working Paper Series 11192, The World Bank.
- Dang, H.-A., Jolliffe, D., and Carletto, C. (2019). Data gaps, data incomparability, and data imputation: A review of poverty measurement methods for data-scarce environments. *Journal of Economic Surveys*, 33(3):757–797.
- Dang, H.-A. H. and Carletto, C. (2022). Recall bias revisited: Measure farm labor using mixed-mode surveys and multiple imputation. GLO Discussion Paper Series 1020, Global Labor Organization (GLO).
- Dang, H.-A. H., Kilic, T., Abanokova, K., and Carletto, C. (2025). Poverty imputation in contexts without consumption data: A revisit with further refinements. *Review of Income and Wealth*, 71(1):e12714.
- Dang, H.-A. H. and Lanjouw, P. F. (2023). Regression-based imputation for poverty measurement in data-scarce settings.
- Mathiassen, A. and Wold, B. K. G. (2021). Predicting poverty trends by survey-to-survey imputation: the challenge of comparability. *Oxford Economic Papers*, 73(3):1153–1174.

# Annex



# Balance Tests - Household Characteristics

**Table:** Mean Differences Balance Test - HH X

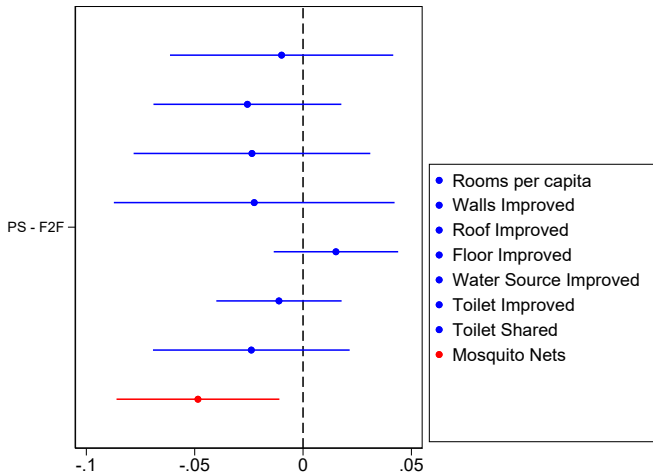
	(1)	(2)	(3)	(4)	(5)	(6)
	Age	Female	Education	Work	HH Size	Urban
PS - F2F	0.690 (0.439)	0.018 (0.015)	-0.066 (0.042)	0.001 (0.005)	0.200*** (0.070)	0.050 (0.042)
Mean Dep. Var.	41.076	0.255	2.360	0.975	4.376	1.780
N PS	2332	2332	2332	2332	2332	2332
Observations	9088	9088	9088	9088	9088	9088

*Notes: Weighted robust standard errors clustered at the EA level are reported in parentheses. Sample is restricted to phone survey eligible households for the cross section. We compare F2F PS sample from Q1 to eligible households in Q2-Q4. Significance levels: \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$*

▶ Study Design

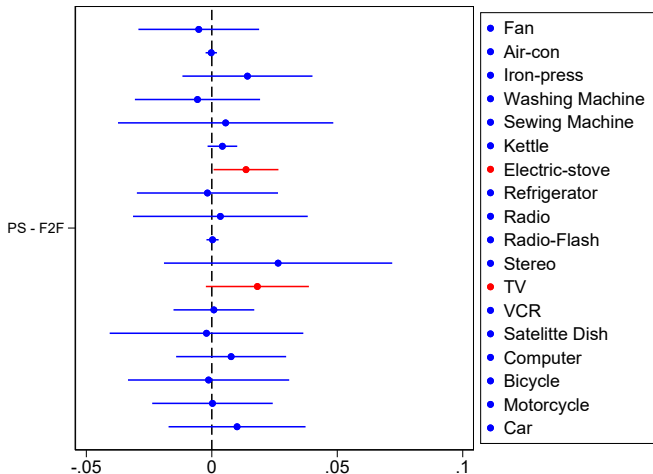
# Balance Tests - Dwelling Characteristics

## ▶ Study Design



# Balance Tests - Assets

## ▶ Study Design



# Monthly Interviews Count by Survey Mode

## ▶ Study Design

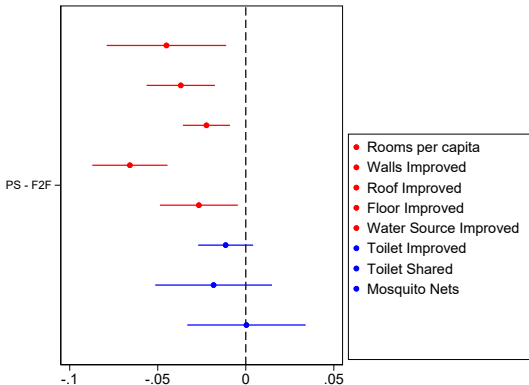
Year/Month	Survey Mode				Total
	F2F	PS Q2	PS Q3	PS Q4	
2024/07	259	0	0	0	259
2024/08	912	0	0	0	912
2024/09	697	0	0	0	697
2024/10	426	0	0	0	426
2024/11	921	360	0	0	1,281
2024/12	752	886	0	0	1,638
2025/01	658	868	0	0	1,526
2025/02	654	45	0	0	699
2025/03	622	0	824	0	1,446
2025/04	897	0	992	0	1,889
2025/05	966	0	238	708	1,912
2025/06	919	0	0	0	919
2025/07	406	0	0	1,220	1,626
2025/08	0	0	0	5	5
<b>Total</b>	<b>9,089</b>	<b>2,159</b>	<b>2,054</b>	<b>1,933</b>	<b>15,235</b>

# Mode differences - Household size

	(1)	(2)	(3)
	hhsiz Q2	hhsiz Q3	hhsiz Q4
PS - F2F	0.301*** (0.083)	0.768*** (0.079)	0.735*** (0.082)
Mean hhsiz F2F	4.658	4.526	4.617
N PS	2158	2053	1932
Observations	4480	4221	4198

Notes: Weighted robust standard errors clustered at the EA level are reported in parentheses. Sample is restricted to phone survey eligible households for the cross section. Significance levels: \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

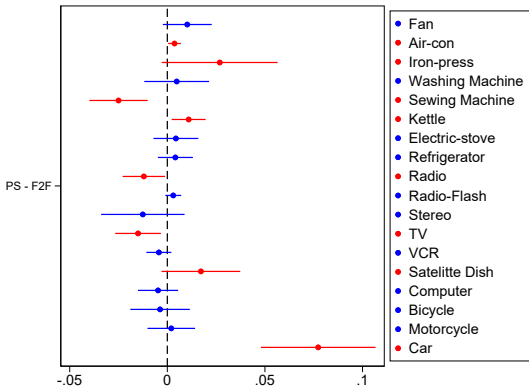
# Mode Differences - Dwelling Characteristics $PS_{Q2} - F2F_{Q1}$



**Figure:** Differences assessed via weighted regression comparing PS report in Q2 to Same household interviewed in Q1 F2F. Sample is restricted to phone survey eligible households. 95% confidence intervals of the coefficients are plotted. Coefficients are red for significance level  $p < 0.1$

► PS Q2 vs F2F (q2-q4)

# Mode Differences - Assets $PS_{Q2} - F2F_{Q1}$



**Figure:** Differences assessed via weighted regression comparing PS report in Q2 to Same household interviewed in Q1 F2F. Sample is restricted to phone survey eligible households. 95% confidence intervals of the coefficients are plotted. Coefficients are red for significance level  $p < 0.1$

▶ PS Q2 vs F2F (q2-q4)