# Ex-ante Harmonization – Innovations in collection of welfare aggregates

The World Bank

The Poverty and Equity Global Practice

Global Solutions Group on Welfare Measurement and Statistical Capacity


João Pedro Wagner De Azevedo

Paul Andres Corral Rodas

# Short consumption modules – Why?

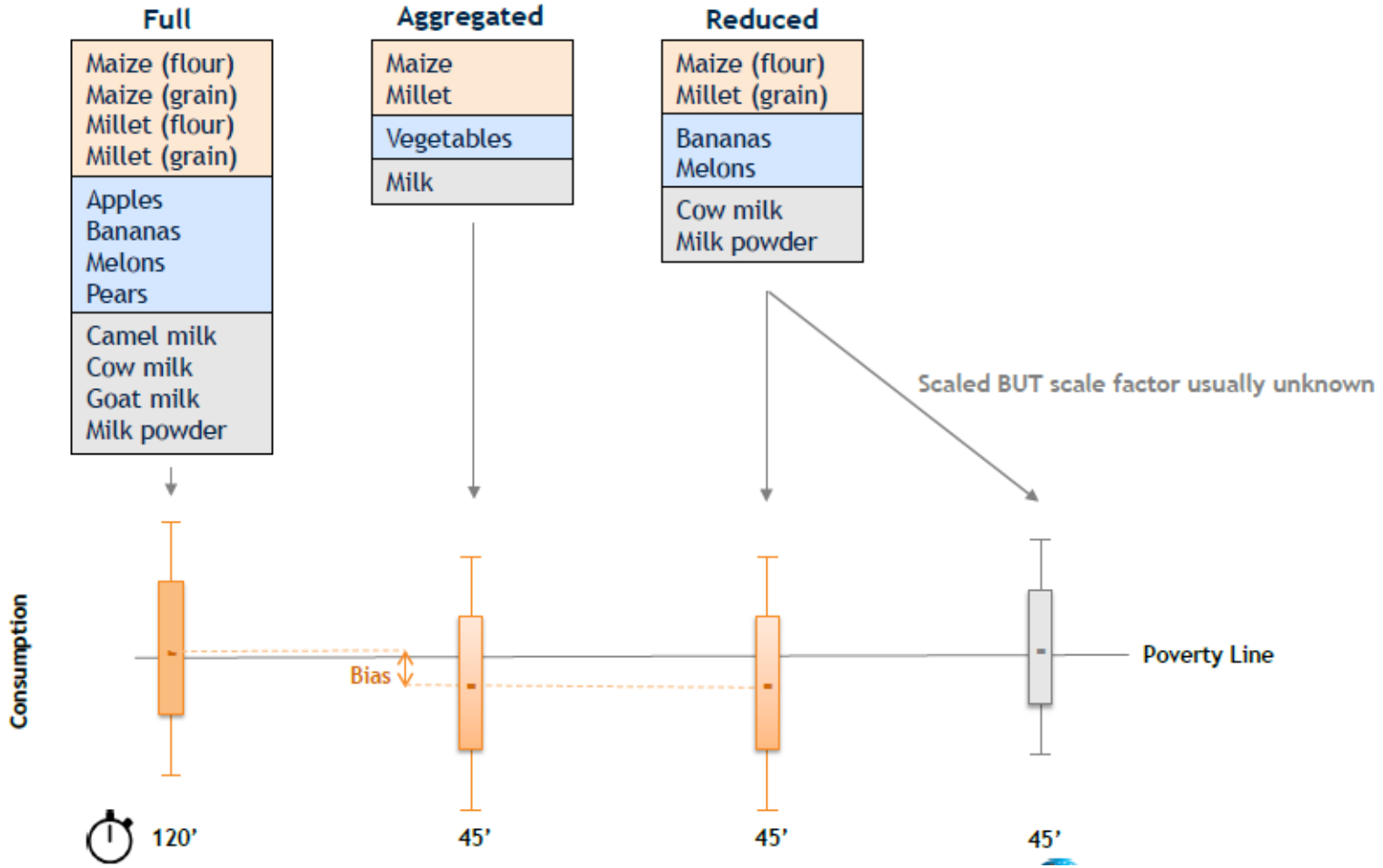
- In many cases it is not feasible to include consumption in one survey for every respondent in one survey
  - Security concerns
  - Cost
  - Time
  - Increases respondent fatigue
- A full aggregate is necessary for monitoring poverty, or program evaluation
- The need for an aggregate database which includes income and consumption
    - EUROMOD uses this for simulating direct and indirect tax policies

# Rapid consumption surveys

- Ask few items per household, but all items across households
- Use statistical methods to obtain full aggregate for all households
- Ex-ante simulations show virtually unbiased results at aggregated levels
  - Successfully implemented in Somalia, South Sudan, and Kenya

# How does it work?

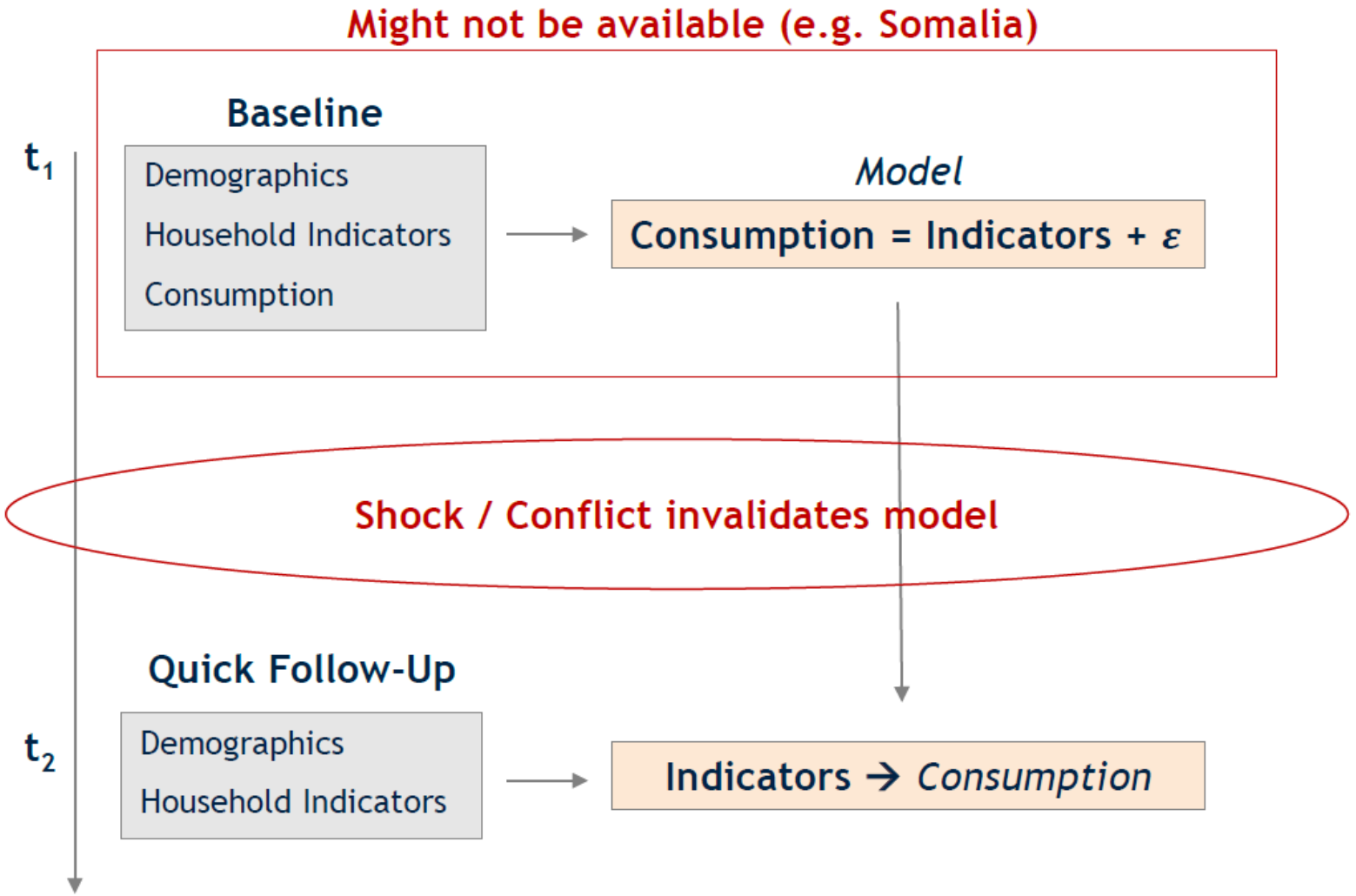
## Reducing the number of items under-estimates consumption





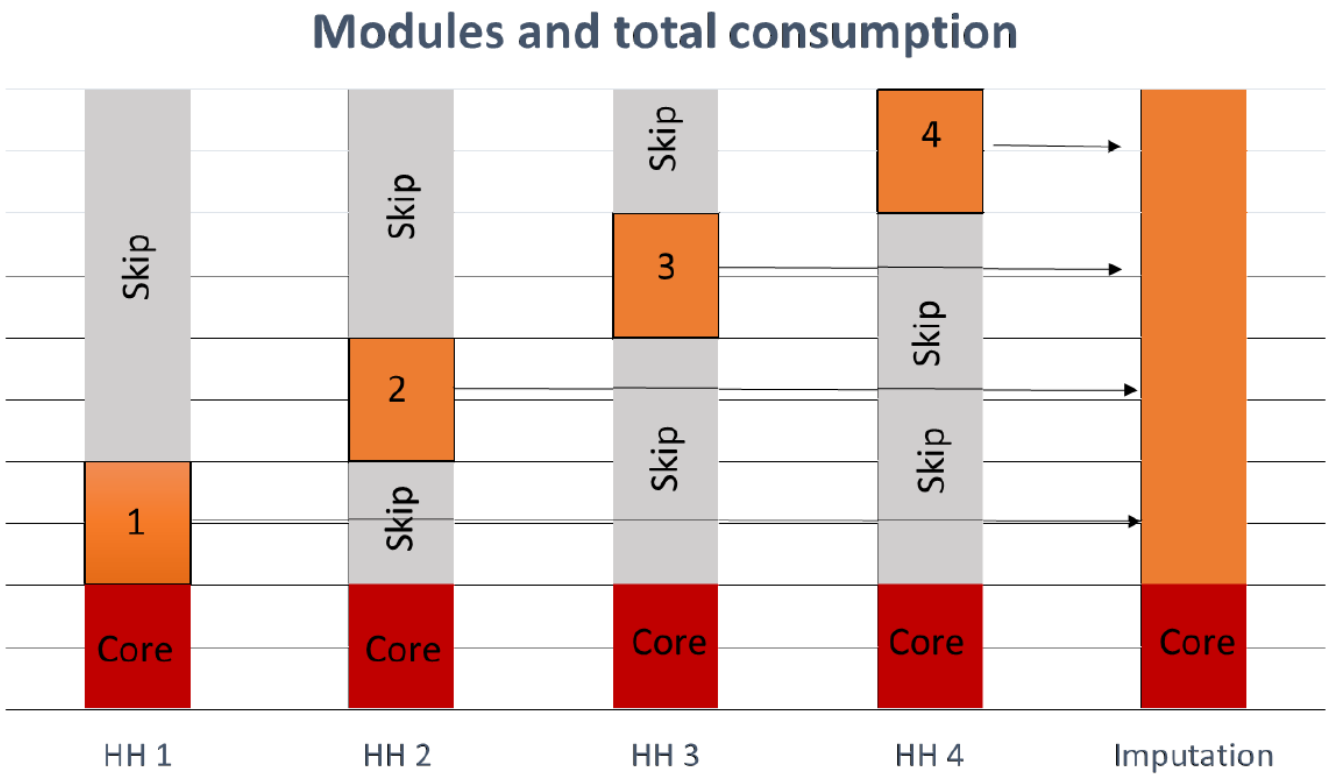

WORLD BANK GROUP

# Why not use cross survey imputation?




- When the source survey was conducted matters
- If the survey is too old, it is unlikely the parameters obtained may be applicable for valid imputation

# The basics

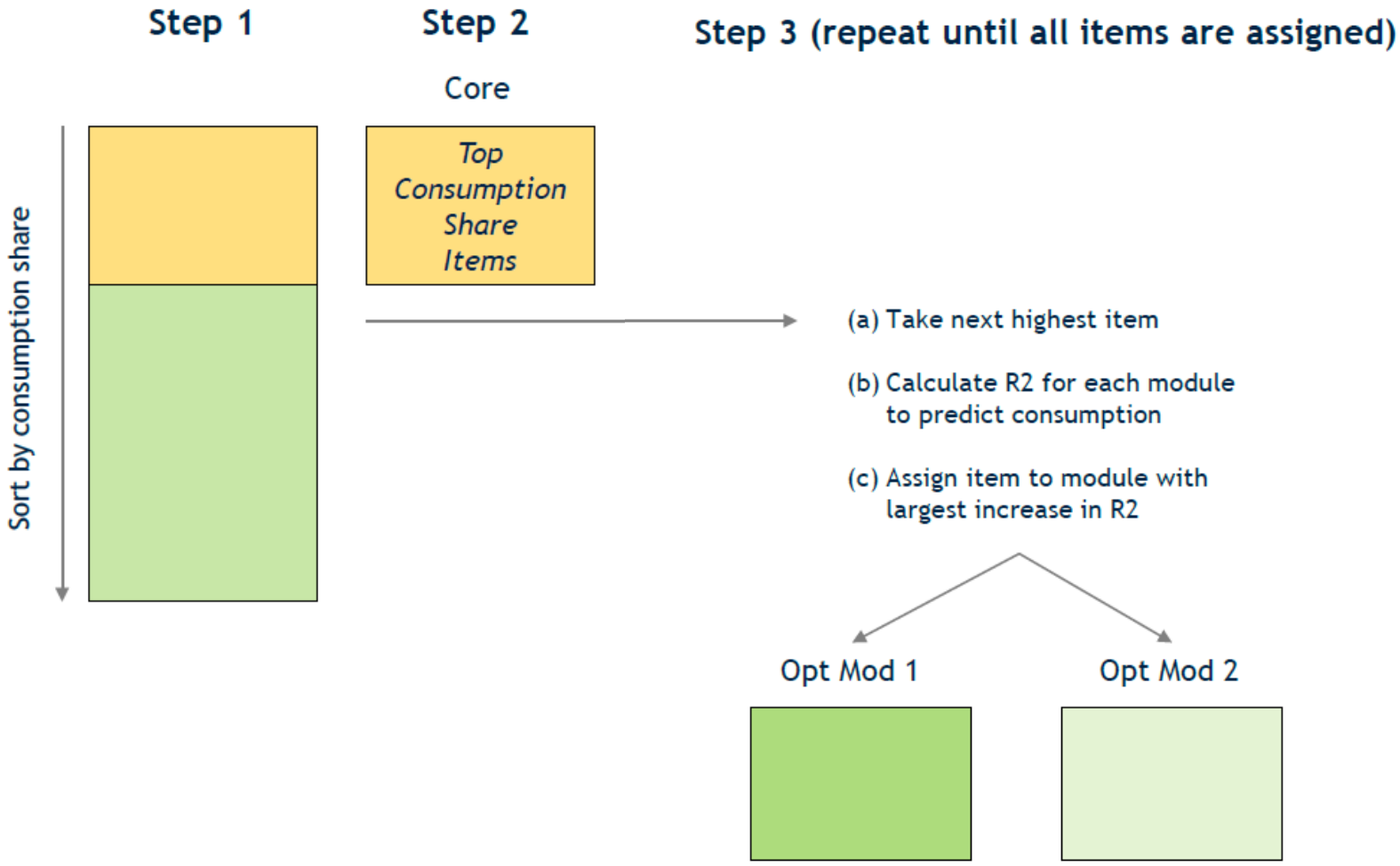



- All households are asked the core questions, and one additional module
- Ask a few items per household, but all items across households
- An imputed complete aggregate is obtained for every household
- Different methods are available for imputing (OLS, Tobit, MICE, averages and medians)



WORLD BANK GROUP

# The basics

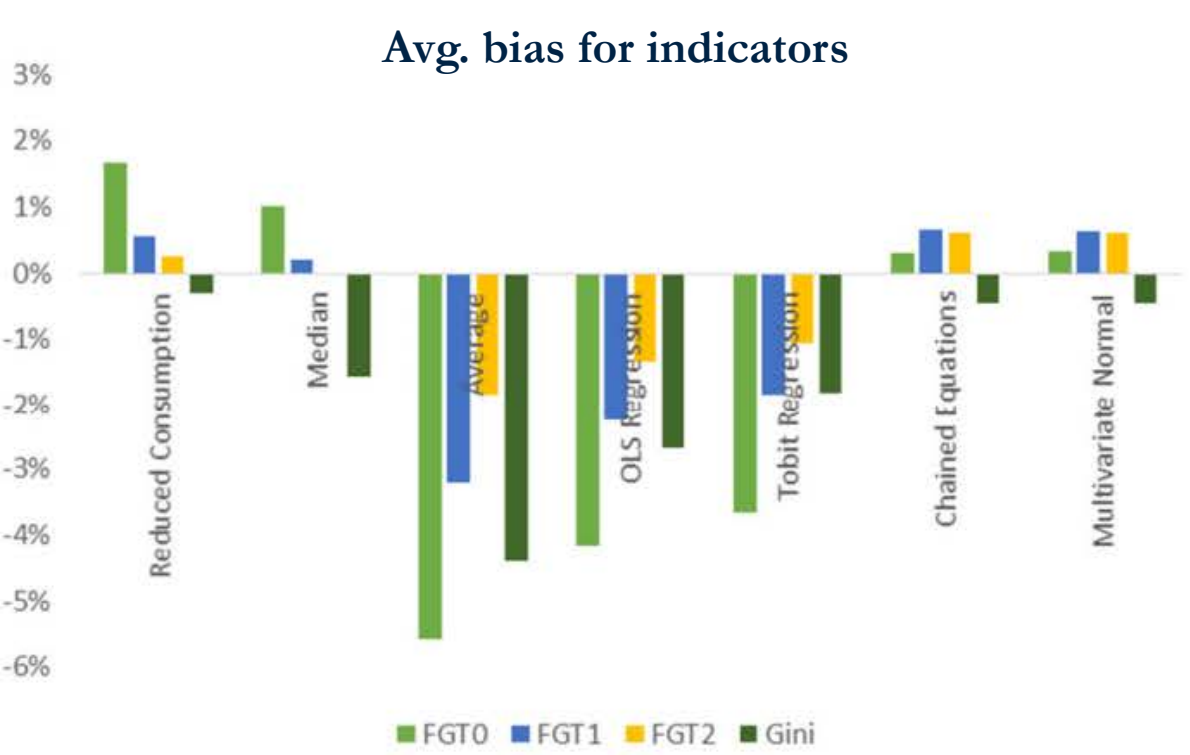


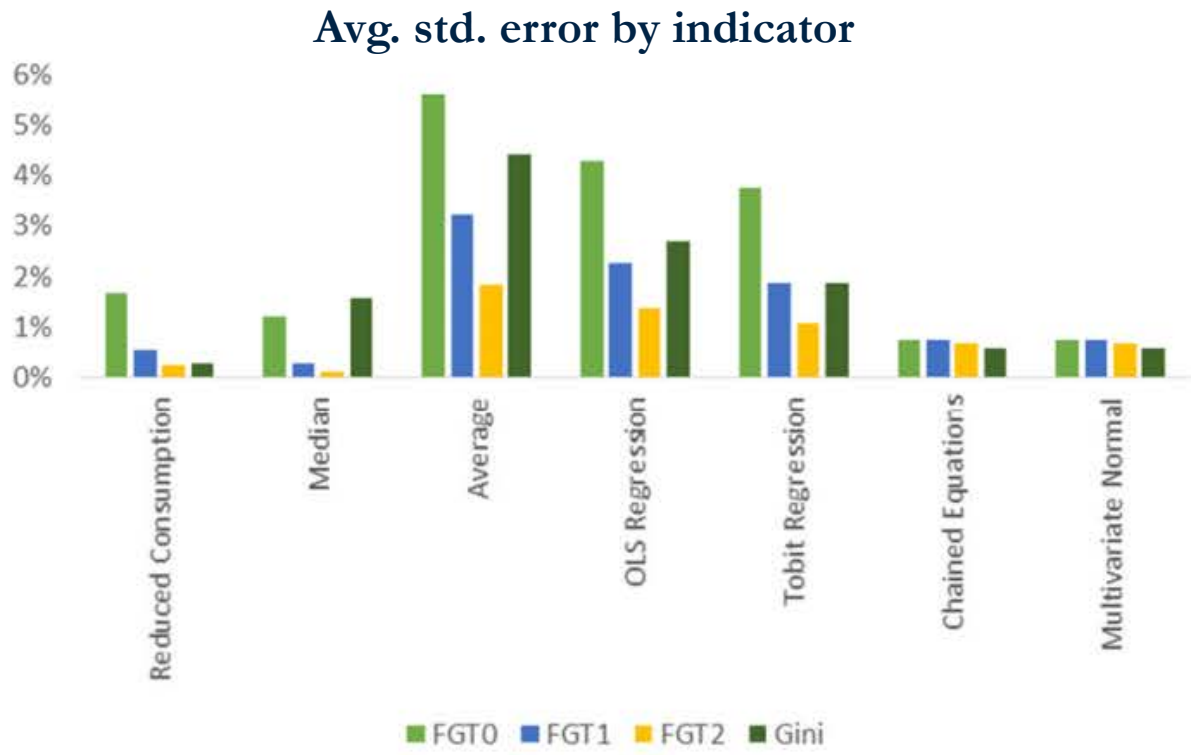

- Module partitions are done on previous consumption data for a similar country
- This improves estimation precision, but does not introduce bias is previous data is poorly correlated with current consumption



WORLD BANK GROUP

# How well does it work?




- Relative bias is avg. error
- Relative std. error is the standard deviation of the relative error
- MI methods work best



WORLD BANK GROUP

# An alternative?

**Common approaches for imputation rely on:**

- Collecting data from samples which are representative of the underlying population of interest
  - Use welfare correlates on a sub-sample to estimate poverty for all the sample
  - Correlates are usually demographic and socio-economic characteristics

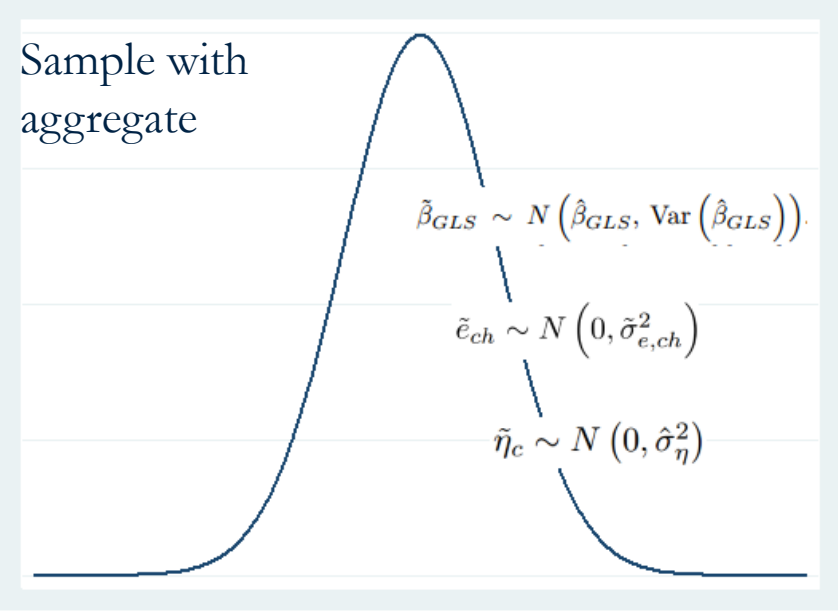
**However**:

- Household consumption patterns can be a great predictor of overall consumption levels
  - A limited set of questions on whether or not a household has consumed certain COICOP level 4 item can greatly assist in obtaining a comparable consumption aggregates

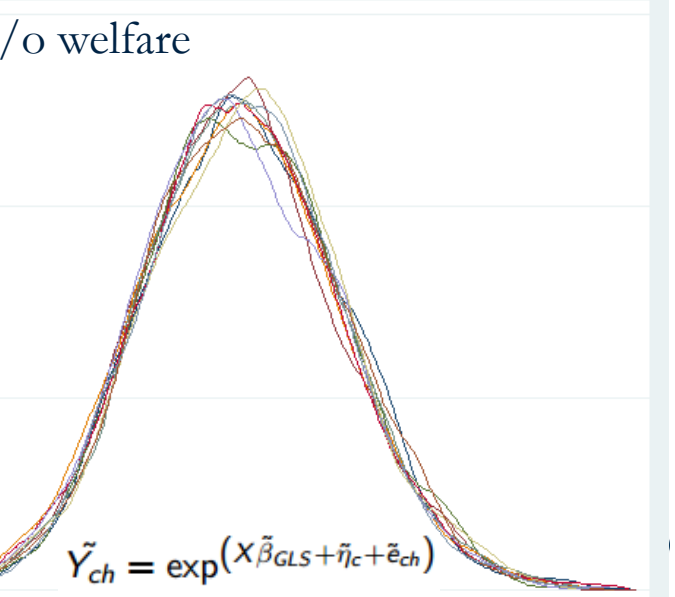
# How ? – Selecting the candidate questions on consumption items

- COICOP 4 questions are selected depending on how well these predict overall consumption
- The pool of possible questions is filtered by requiring a proportion of households reporting consumption between 0.10 and 0.90
- Correlates are chosen based on overall model fit (adjusted R2 and F-stat), as well as the significance of the correlate
  - All correlates with high variance inflation factors are removed
  - Since we are simulating welfare onto a different sample, multicollinearity and overfitting are issues we want to avoid

# Monte Carlo Simulation




- The goal is to simulate a sufficiently large number of vectors of welfare to allow for reliable estimates of poverty (usually 100)
- From the first stage parameters it is possible to take random drawings from the assumed distributions
  - Alternatively it is possible to get bootstrapped samples of the survey data to yield all parameters needed for simulating welfare vectors




- Take the drawn parameters and apply these to the *X* matrix of characteristics in the sample w/o welfare, and simulate the residuals
- This yields *R* simulated vectors in the sample data
- From these vectors we get *R* poverty rates

# How good are our results?

- Household consumption patterns can be a great predictor of overall consumption level
- A limited set of 35 questions if a household has consumed certain COICOP level 4 items can go a long way towards obtaining a comparable consumption aggregate

**Table 1: ELL simulation of adult equivalent consumption for Greece**

| | Sim (1) | Sim (2) | Sub sample 2 sim (1) | Sub sample 2 sim (2) |
|---|---|---|---|---|
| *Model details:* | | | | |
| Observations | 5,857 | 5,857 | 2,860 | 2,860 |
| Regressors | 35 | 35 | 35 | 35 |
| Adjusted R2 | 0.67 | 0.67 | 0.66 | 0.66 |
| Max. VIF | 3.63 | 3.63 | 3.59 | 3.59 |
| F-Stat | 345.80 | 345.80 | 249.40 | 161.51 |
| Het. Adj. R2 | | 2.E-03 | | 1.E-03 |
| Het. F-Stat | | 6.50 | | 2.52 |

Source: Greek 2014 HBS, own estimates

**Table 2: Simulation exercise for Greece**

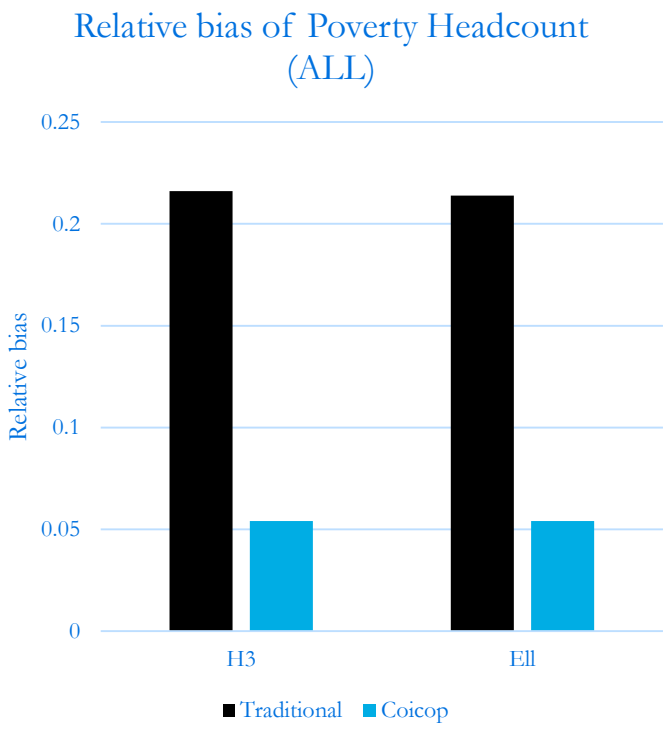
| | Direct estimate | Full sample Sim (1) | Full sample Sim (2) | Sub-sample (50%) Sim (2) | Sub-sample (50%) Sim (2) |
|---|---|---|---|---|---|
| **Children** | | | | | |
| *Headcount* | 21.5 | 25.1 | 20.3 | 25.1 | 20.0 |
| *Gap* | 4.1 | 8.4 | 5.2 | 8.3 | 5.2 |
| *Severity* | 1.4 | 3.9 | 1.9 | 3.9 | 2.0 |
| **Elderly** | | | | | |
| *Headcount* | 27.3 | 31.9 | 26.8 | 32.3 | 27.4 |
| *Gap* | 6.0 | 10.6 | 6.5 | 10.7 | 6.6 |
| *Severity* | 2.1 | 4.9 | 2.3 | 4.9 | 2.3 |
| **National** | | | | | |
| *Headcount* | 20.9 | 25.2 | 19.9 | 25.5 | 20.2 |
| *Gap* | 4.8 | 8.3 | 4.9 | 8.4 | 5.0 |
| *Severity* | 1.8 | 3.8 | 1.8 | 3.8 | 1.8 |
| ***Gini*** | 34.6 | 42.7 | 34.7 | 42.7 | 34.6 |
| ***GE 0*** | 20.0 | 31.5 | 19.8 | 31.6 | 19.7 |
| ***GE1 (Theil)*** | 21.1 | 32.1 | 20.3 | 32.0 | 20.2 |
| ***GE2*** | 28.2 | 45.7 | 25.6 | 45.3 | 25.2 |

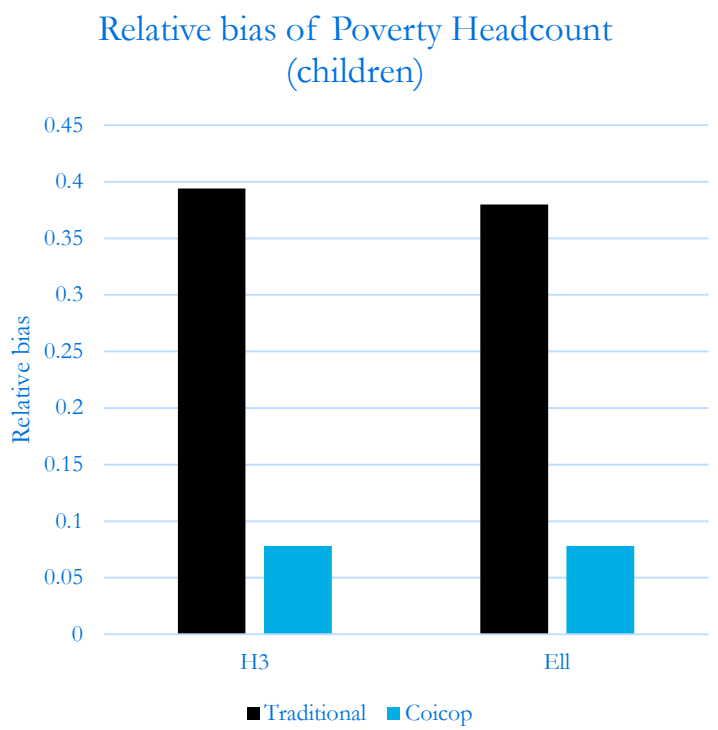
**Source: Greek HBS 2014, Own estimates.**

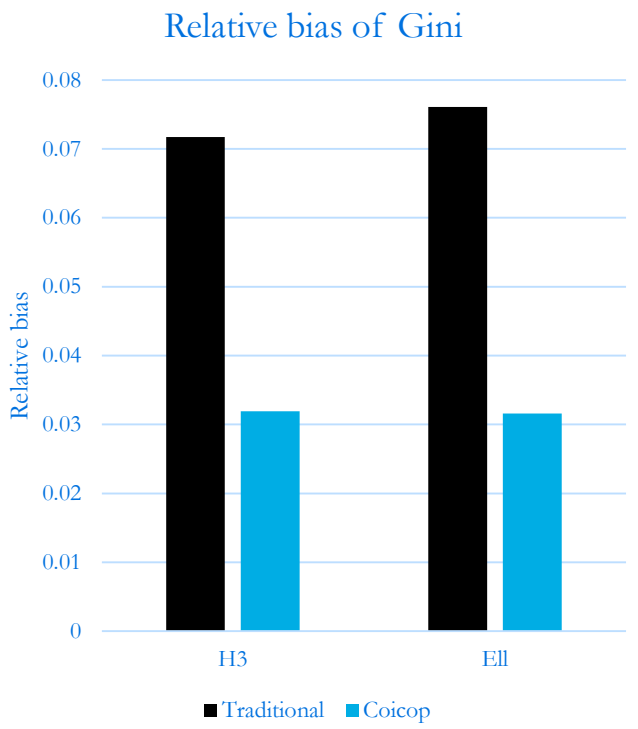
Note: Relative line (60% of median adult equivalent consumption). Sim 1: Only consumption dummies, and nat log. of hosuehold size. Sim 2: Only consumption dummies, heteroskedasticity modeled with share children and elderly. Simulations done using ELL methodology. Full sample (5,888 hh).

# How does it compare to other approaches?

- It does considerably better than imputations which use demographics and socio-economic characteristics



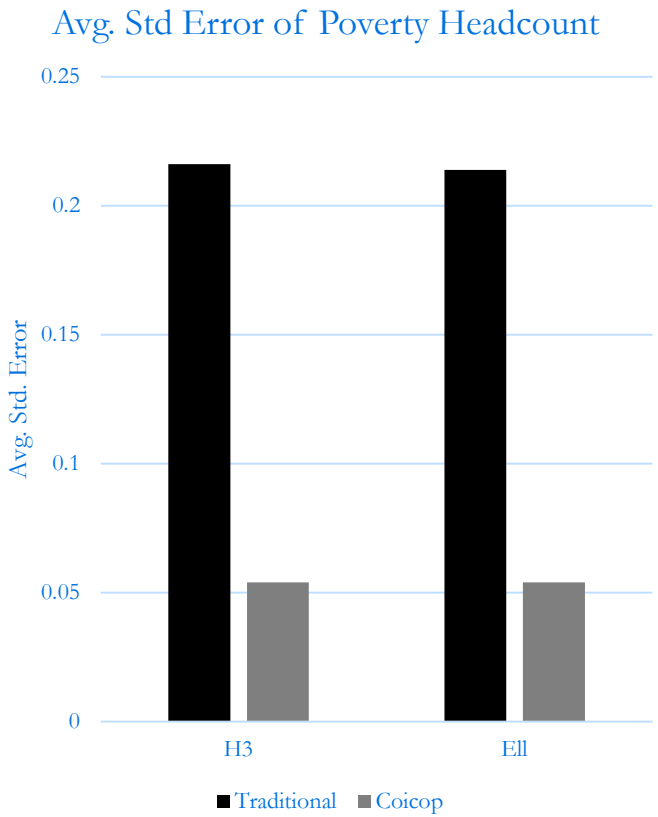



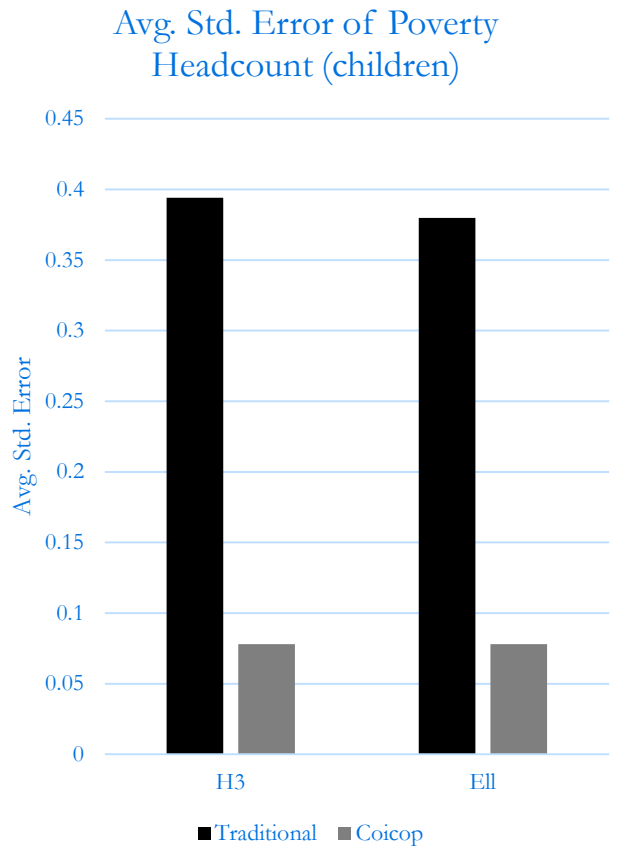


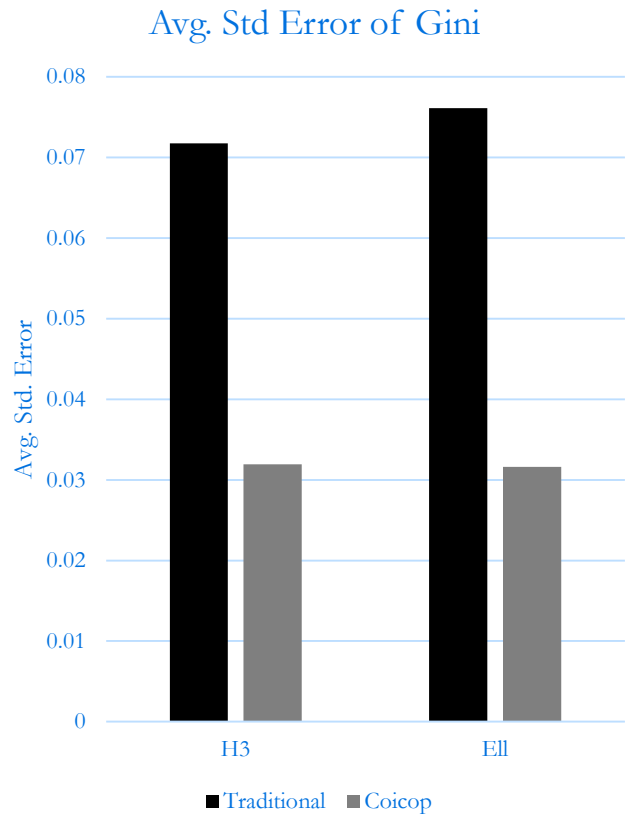

WORLD BANK GROUP

# How does it compare to other approaches?

- Also yields considerably lower standard average errors






WORLD BANK GROUP

# Next Steps & Conclusions

- Short consumption modules show plenty of promise
  - Possible to estimate consumption and poverty
  - Accurately simulate distribution sensitive parameters
  - Accurately simulate estimates for sub-groups of the population
- Not only valid, but a cost effective approach for obtaining consumption aggregates