UNECE

WORLD BANK GROUP

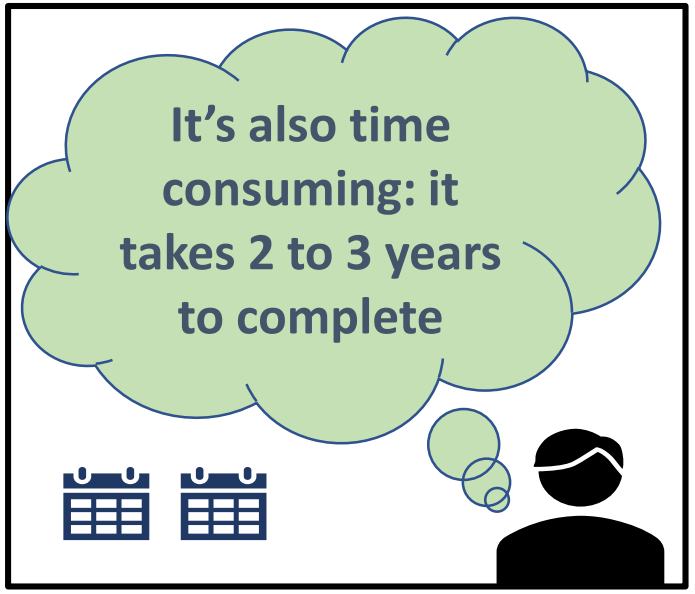
SWIFT

Survey of
Well-being via
Instant and
Frequent Tracking

Project Lead:
Nobuo Yoshida,
Lead Economist,
Poverty & Equity Global Practice

SWIFT
Survey of
Well-being via
Instant and
Frequent Tracking

Constraints to “no-one left behind” commitment

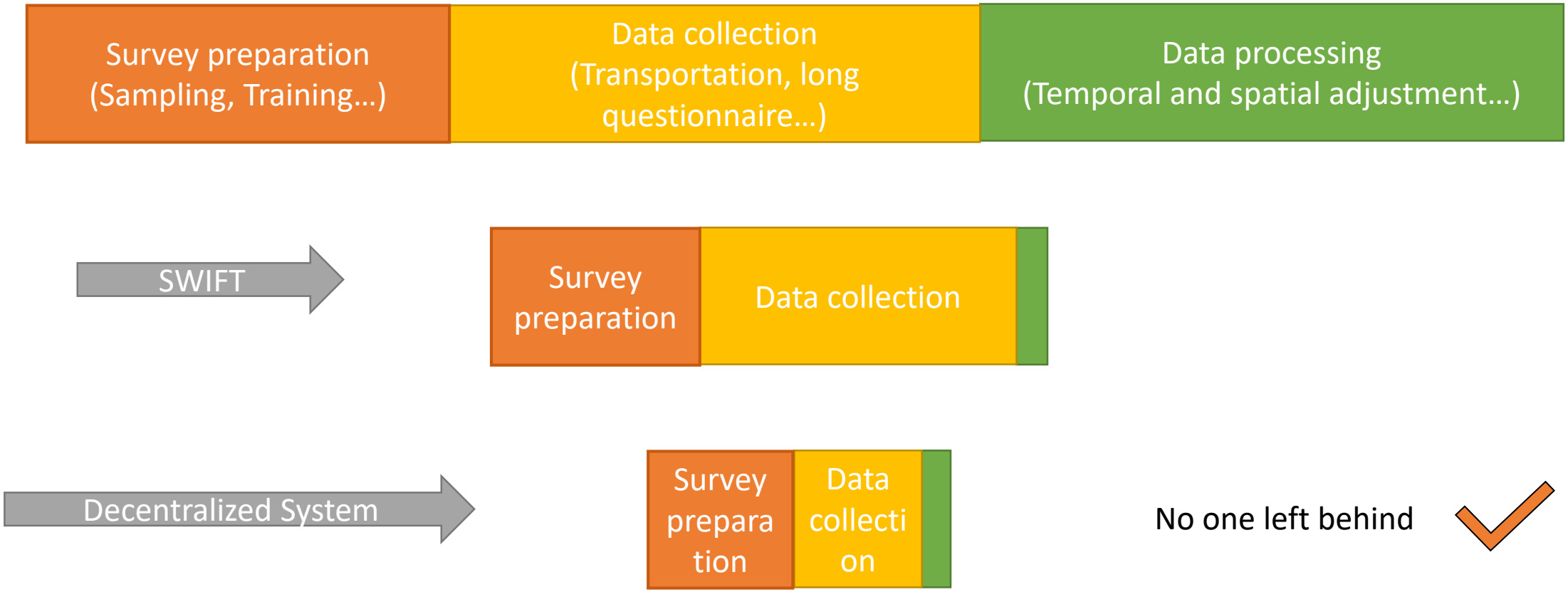


More than 60 countries cannot monitor poverty and shared prosperity

WORLD BANK GROUP

SWIFT Survey of Well-being via Instant and Frequent Tracking

How can we help?

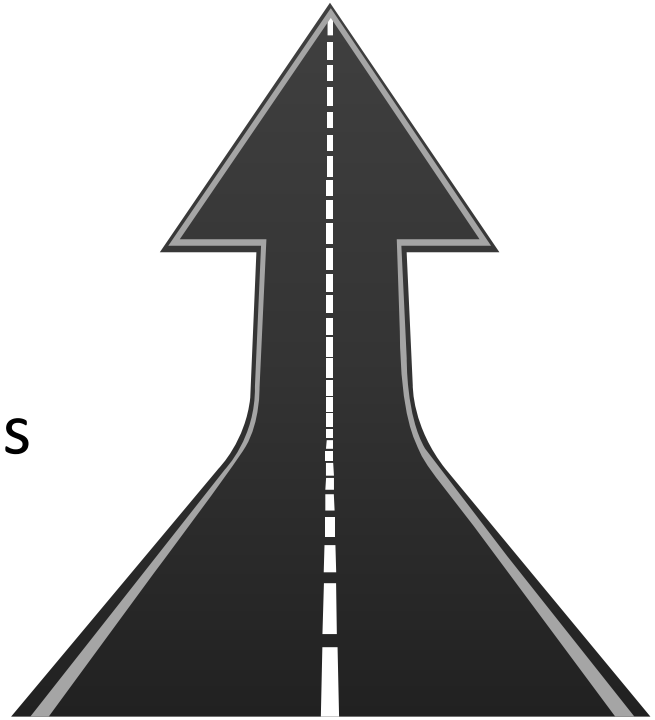


WORLD BANK GROUP

SWIFT Survey of Well-being via Instant and Frequent Tracking

Sequence of the presentation

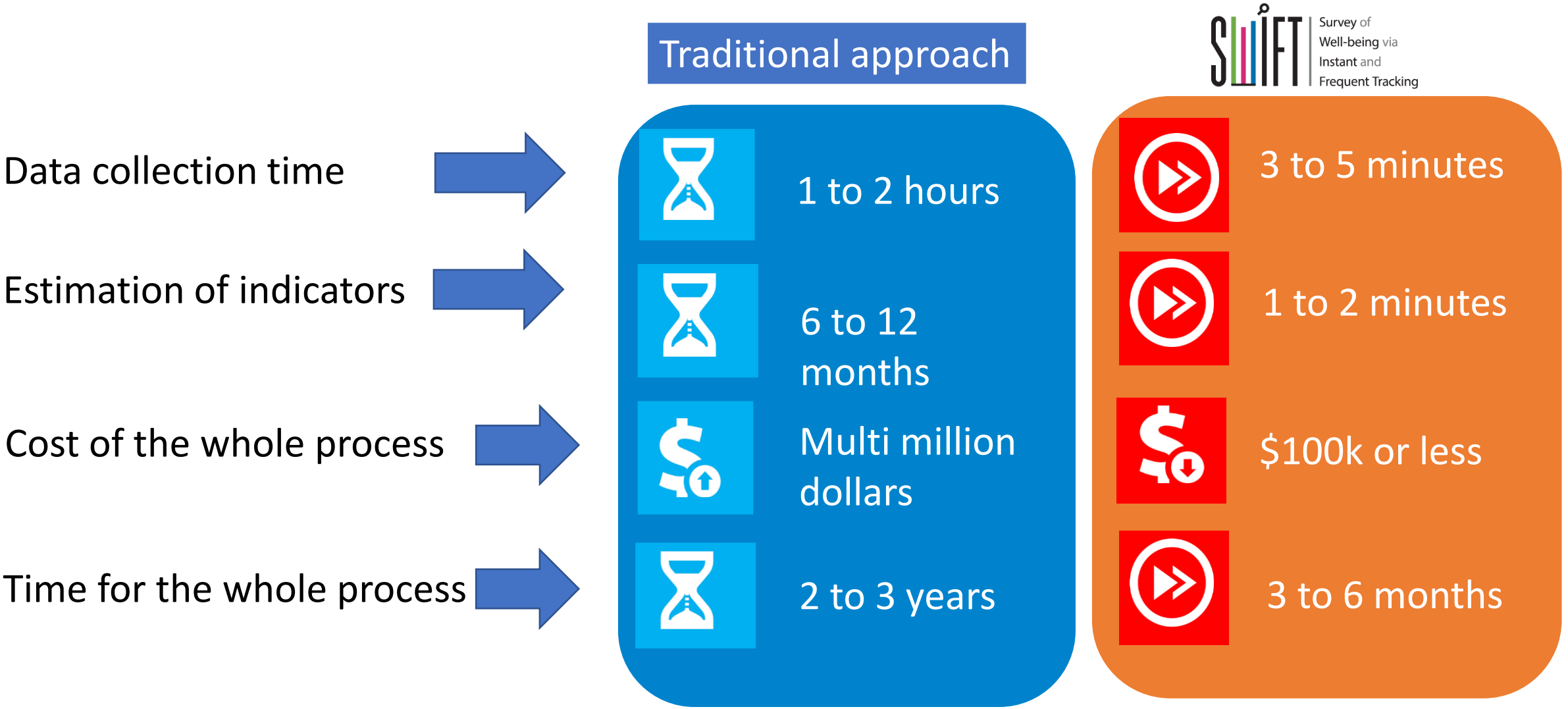
- SWIFT Imputation
  1. Standard SWIFT
  2. Consumption Dummy Based SWIFT Imputation
  3. SWIFT 2.0: ultimate solution to stability concerns
- Systematic Revolution
  - Decentralized Survey System



WORLD BANK GROUP

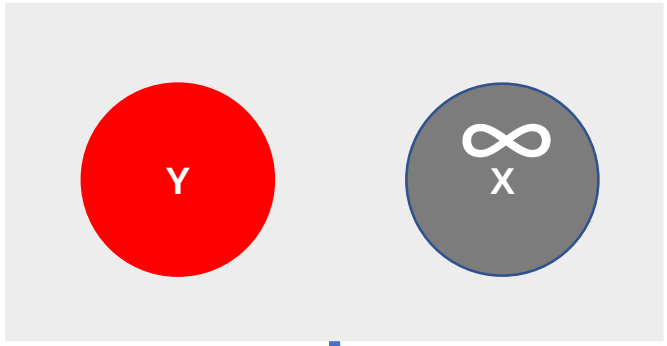
SWIFT Survey of Well-being via Instant and Frequent Tracking

Difference between SWIFT and traditional approach



Standard SWIFT

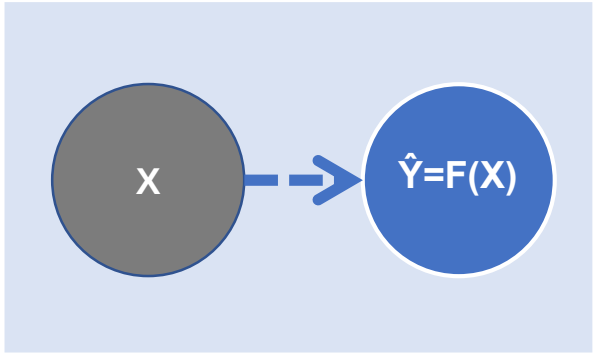
Baseline Survey: Most recent HBS



Use *Machine Learning techniques* to find a **formula** that connect consumption with limited number of non-consumption variables

Y=F(X)

Endline Survey



SWIFT imputation utilizes the existing HBS, extracts 10-20 questions that can be merged into any follow up surveys.

Data collection time and cost reduced significantly.

Imputation to get final poverty estimates is done in real time.

Y: Consumption
X: Identified non-consumption variables (e.g. housing condition, asset ownerships)
Ŷ = F(X): Projected Consumption data

Standard SWIFT

Modeling

- Create a model by running regressions

$$lny_h = \alpha + \beta_1 * x_{1h} + \beta_2 * x_{2h} + \cdots + \beta_k * x_{kh} + \varepsilon_h$$

Left hand side variable:
log of household expenditure per capita/per adult equivalence

Right hand side variables:
Stepwise selection from candidates pool

Typical Categories	Examples
Household characteristics	Household size, female ratio, dependency ratio, Education level, age, gender, employment, marital status…
Housing condition	Wall material, water source, toilet facility, electricity…
Livestock and agriculture	Cattle, chicken, donkey, agriculture land…
Asset ownership	Bicycle, radio, mobile phone, television…
Location and seasonality	Region, season

SWIFT Survey of Well-being via Instant and Frequent Tracking

Standard SWIFT

Issues

- Over-fitting
  - A model performs well in current dataset but might not project well using new data
- Multi-collinearity
  - Stepwise regression is vulnerable to multi-collinearity
- Stability of coefficients over time
  - Models developed in current round might not hold for other round of survey
- Misspecification of error structure
  - Error distributions can be very complex

Standard SWIFT

Solutions

1. Cross Validation to minimize risk of overfitting
2. Multicollinearity checks
3. Stability test using Backward/Forward Imputation
4. Addressing distributional issues with PMM, SAE or State-of-art Machine Learning techniques

Standard SWIFT

Cross Validation

Backward/Forward Imputation during different rounds of survey

Coefficients checks

PMM
SAE
State-of-art Machine Learning techniques

Standard SWIFT

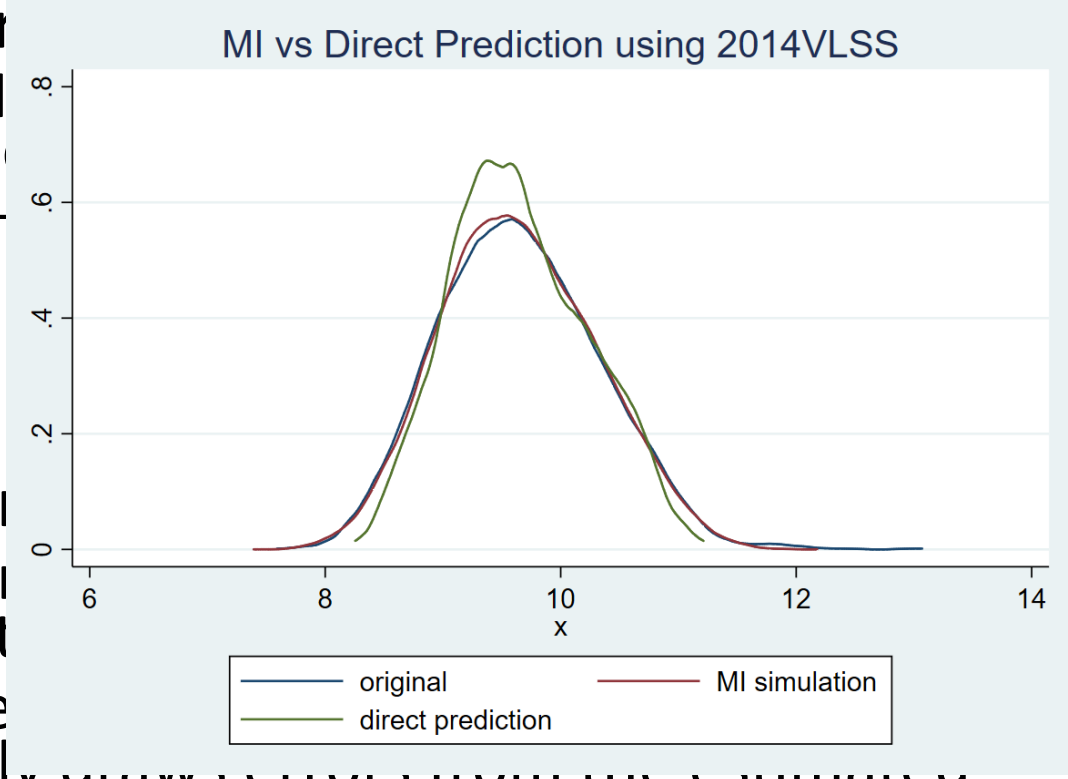
Cross-Validation -- > Stepwise Regression

- Cross-Validation is used to see the out-of-sample performance rather than within-sample performance
- The risk of overfitting problem rises as more variables are included
- Using the cross-validation approach, we try to find the optimal number of variables
- The optimal p-value for the stepwise regression set the threshold for including more variables, which find the balance between in-sample accuracy and out-of-sample projection

Standard SWIFT

Simulation stage – Multiple Imputations

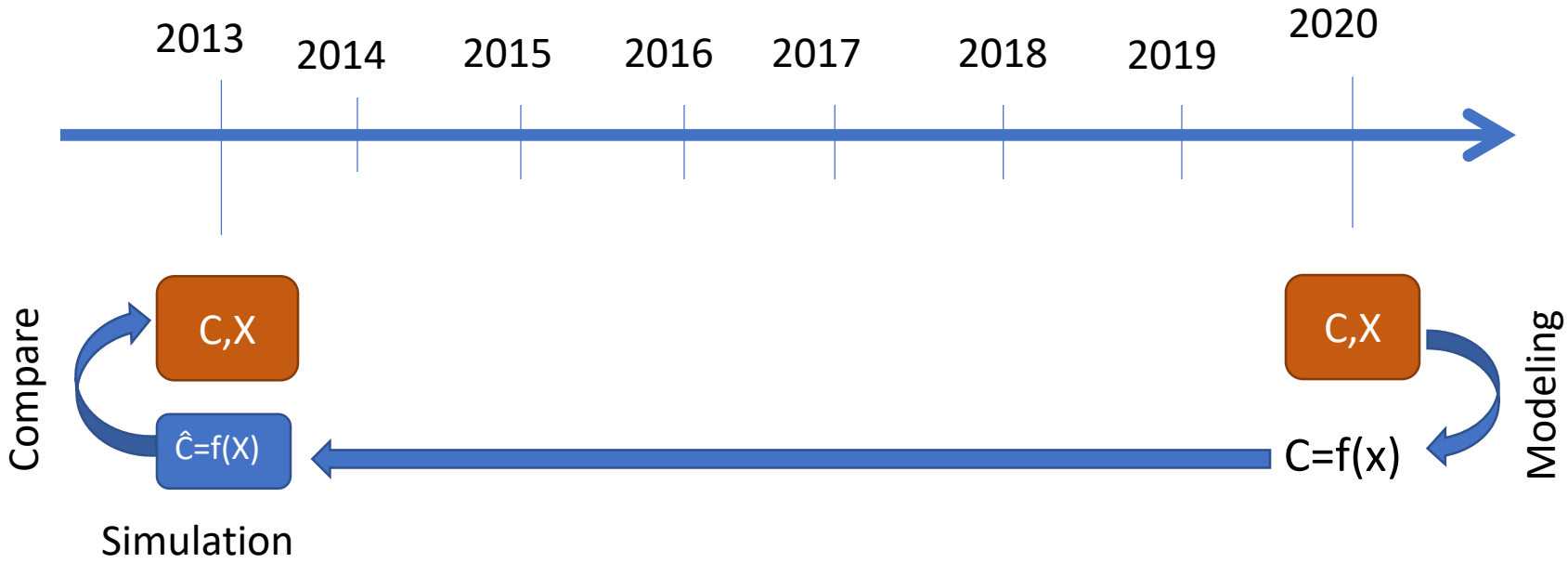
- Simulate HH expenditure for each household in SWIFT Su
  - Randoml
    estimate
    $lny_h = \alpha +$ ... h
- Simulatio
  - Direct pr ... ed around t
  - Multiple ... use it randomly draws errors from the estimated distribution



SWIFT Survey of Well-being via Instant and Frequent Tracking

Standard SWIFT

Stability Test



C,X	Two rounds of Household Budget Survey datasets
Ĉ=f(X)	Projected consumption data using SWIFT model

SWIFT Survey of Well-being via Instant and Frequent Tracking

Standard SWIFT

Steps for SWIFT modeling & simulation

1. Cross-validation to decide the optimal p-value
2. Run the stepwise regression using the optimal p-value
3. Check the coefficients of the final model
4. Simulate household expenditures using Multiple Imputation
5. Backward imputation analysis to check model stability

SWIFT Survey of Well-being via Instant and Frequent Tracking

Standard SWIFT

Standard SWIFT can fail -- A real case in Afghanistan

	Original	Standard SWIFT Projection using 2011 model
2011	38.3%	
2016	54.5%	44.8%

44.8% is far from original 54.5%. Why standard SWIFT fails to capture the change?

Household assets are accumulating

Household report worsened economic condition, less satisfied with police work, and much insecure

Standard SWIFT is Asset Ownership based. In Afghanistan, asset ownerships are not reliable welfare indicators.

Revised SWIFT

Consumption Based SWIFT Imputation

Standard SWIFT

- Asset Ownership Approach
  - Asset ownership *irreversibility*
    - Can people trade in their useless assets for money?
    - Do they have fair second hand market?

Replaced

Revised SWIFT

- Consumption Dummy Approach
  - Food and non-food dummies are more sensitive to welfare change

Revised SWIFT

Standard SWIFT

82% asset increased

Replaced

Revised SWIFT

73% consumption decreased

Asset Ownership Dummy	2011/12	2013/14	2016/17
Refrigerator	12.0%	16.2%	16.3%
Stove	64.7%	75.8%	78.9%
Sewing Machine	63.7%	63.6%	64.6%
Iron	56.7%	58.1%	63.7%
Radio	57.1%	43.3%	30.9%
TV	34.9%	39.4%	42.3%
Motorcycle	23.8%	28.0%	32.1%
Car	12.5%	13.6%	14.6%
Tractor	1.9%	3.4%	2.6%
Carpets	17.8%	31.4%	17.2%
Blankets	99.6%	96.7%	99.6%

No fair second hand market

Consumption Item Dummy	2011/12	2016/17	Consumption Item Dummy	2011/12	2016/17
Cigarettes	17.5%	15.1%	Women shoe	96.8%	93.6%
Tobacco	35.0%	37.1%	Child shoe	88.3%	87.8%
Matches	98.4%	96.9%	Wedding	7.8%	8.0%
Cleaning	95.6%	96.7%	Celebration	44.1%	32.4%
Soap	82.9%	81.1%	Hospital	3.5%	3.3%
Shampoo	94.3%	93.6%	high quality rice	43.6%	37.6%
Toothpaste	60.0%	54.9%	low quality rice	75.4%	68.3%
Grooming	61.8%	60.9%	Purchased Nan	10.9%	12.4%
Bath out	16.3%	10.7%	Beans	53.2%	50.7%
Laundry	24.8%	12.9%	Pasta	26.0%	25.0%
Transport	52.5%	57.6%	Beef	28.0%	12.2%
Fuel	25.4%	33.1%	Chicken	35.8%	33.8%
Tax	5.0%	6.4%	Milk	45.2%	31.5%
Construction	27.6%	17.4%	Yogurt	49.0%	44.2%
Education	11.0%	11.2%	Dough	24.1%	20.6%
Uniform	42.2%	39.3%	Egg	38.9%	27.2%
Test book	20.4%	22.0%	Potato	91.9%	84.9%
Pen	49.8%	57.9%	Onion	92.2%	93.3%
Repair	16.7%	24.3%	Tomato	39.7%	46.5%
Airfare	1.8%	1.5%	Fruit	61.6%	52.3%
Men cloth	96.5%	95.1%	White sugar	86.0%	82.9%
Women cloth	96.9%	95.5%	Chocolate	58.6%	52.9%
Child cloth	90.4%	89.9%	Black tea	45.1%	34.0%
Men shoe	96.9%	93.5%	Green tea	84.0%	87.0%

Tight budget. Less consumption

Revised SWIFT

Consumption Based SWIFT Imputation

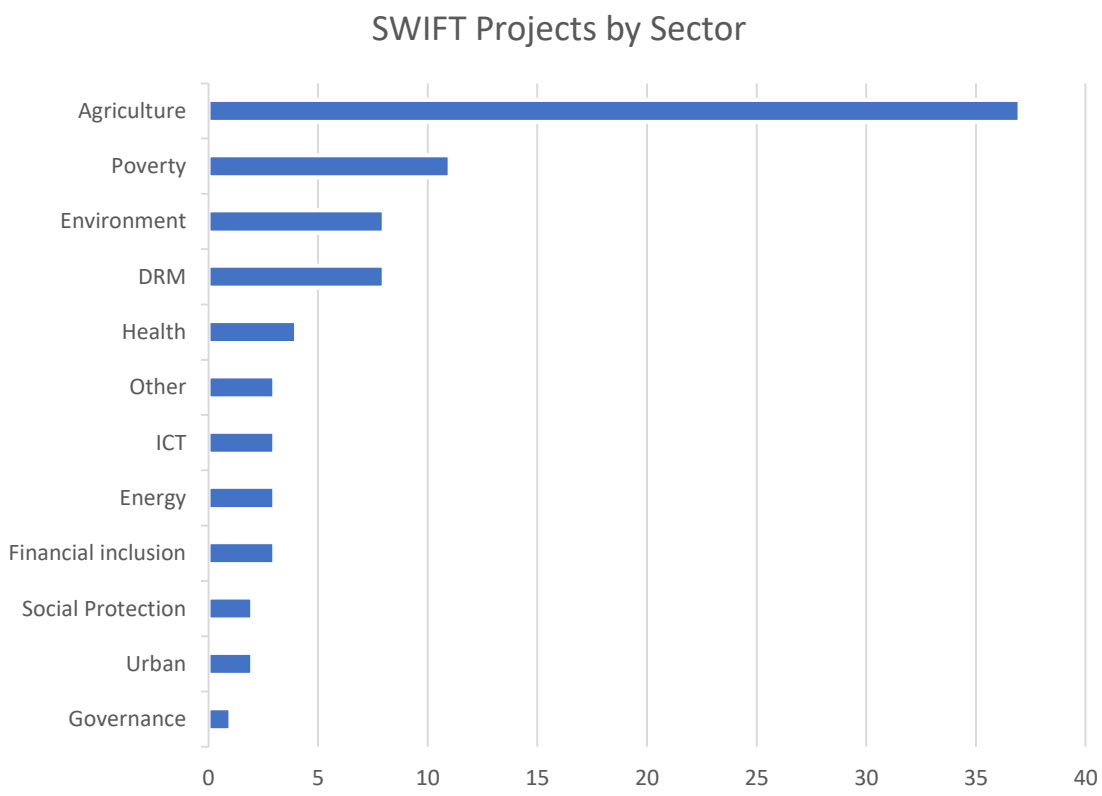
Simulation Results

	Original	Standard SWIFT Imputation	Consumption Based SWIFT
2011	38.3%		
2016	54.5%	44.8%	53.5%

Significant improvement

Current status of SWIFT projects

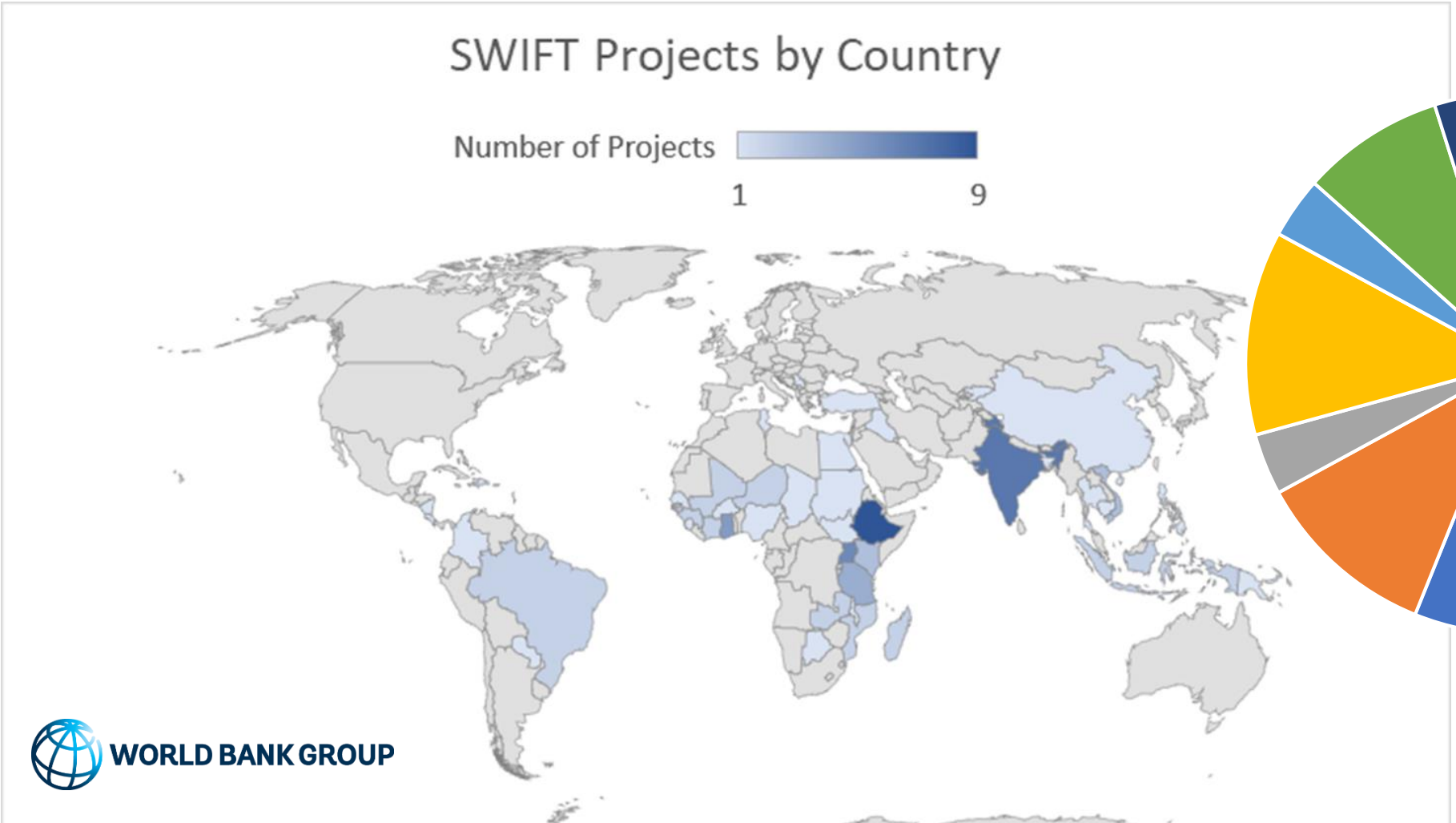
- SWIFT project was launched in June 2015
- Over 100 SWIFT surveys in 52 countries are under implementation or completed
- Varieties of field including DRM, ICT, health, agriculture, energy, etc.

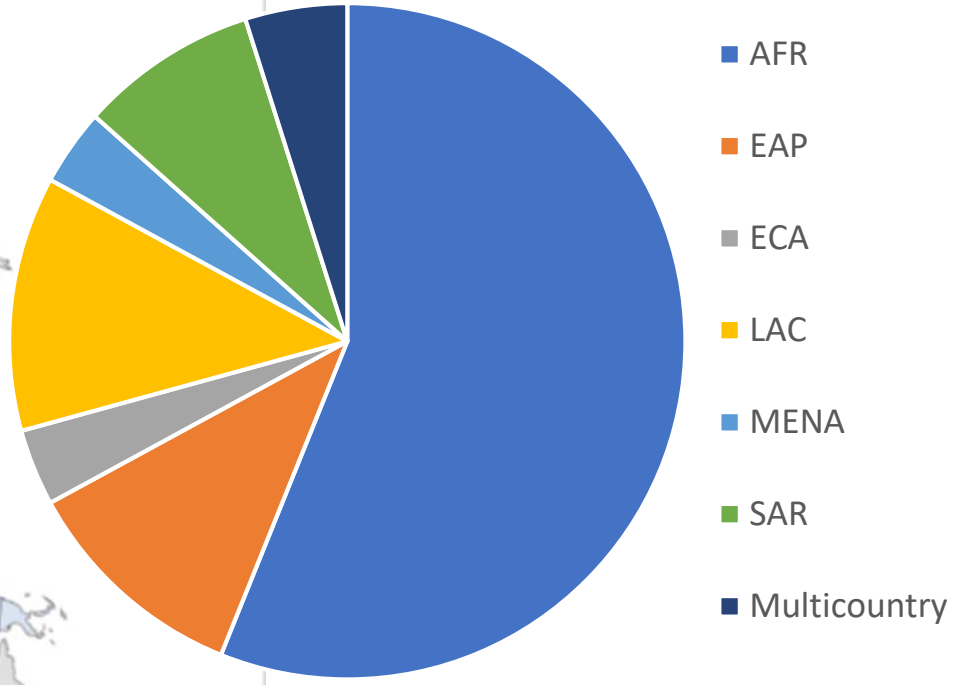


WORLD BANK GROUP

SWIFT Survey of Well-being via Instant and Frequent Tracking

Current status of SWIFT projects



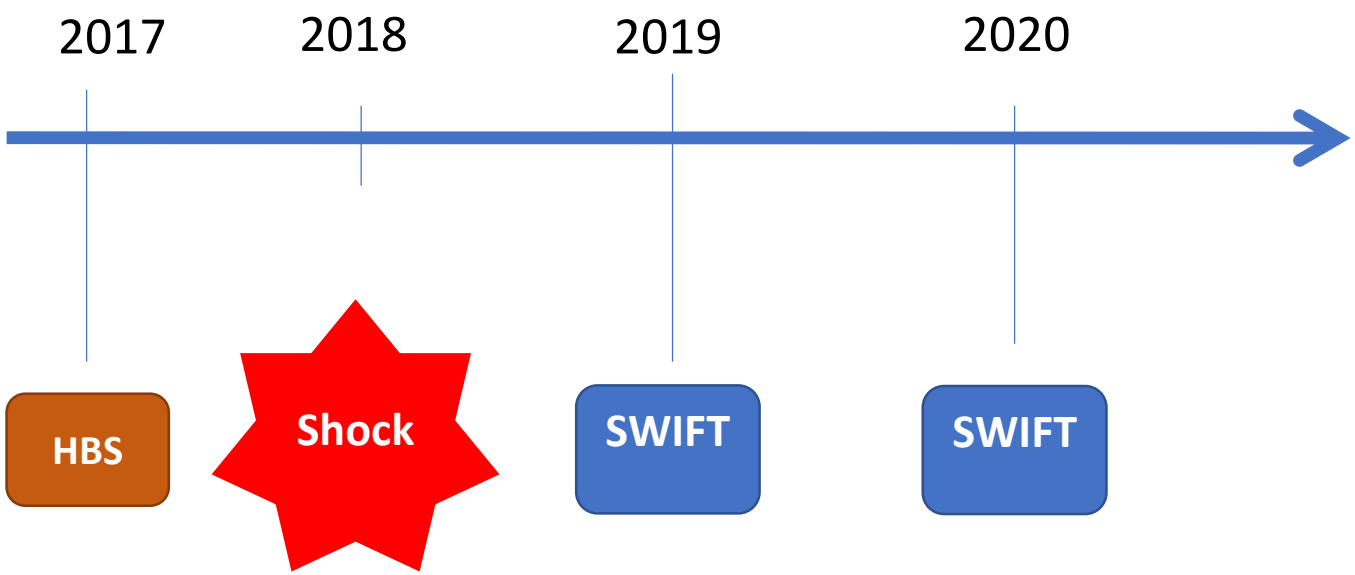


WORLD BANK GROUP

SWIFT Survey of Well-being via Instant and Frequent Tracking

SWIFT 2.0

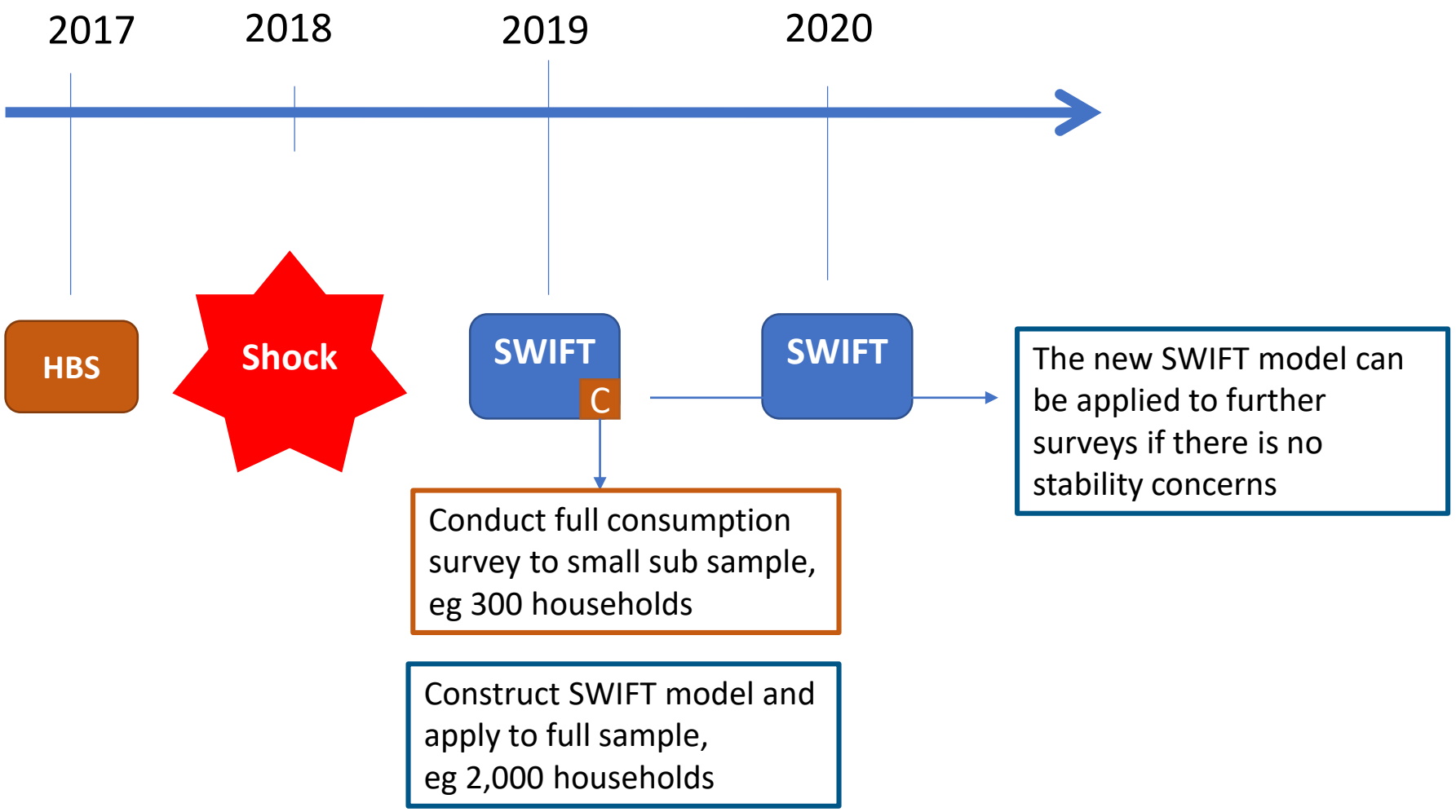
A model might not be stable over time



SWIFT 2.0

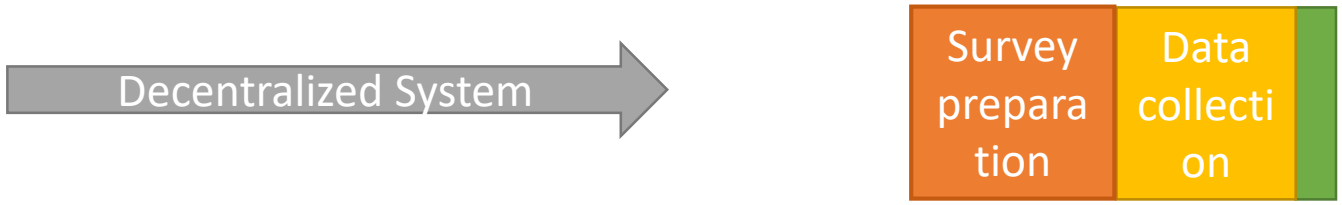
SWIFT 2.0: Ultimate solution to Stability concerns

- Collect consumption data from a small sub-sample in each round of hybrid surveys
- Use them to recreate imputation models for each round



But even with SWIFT, preparation of household survey is painful, costly, time-consuming…

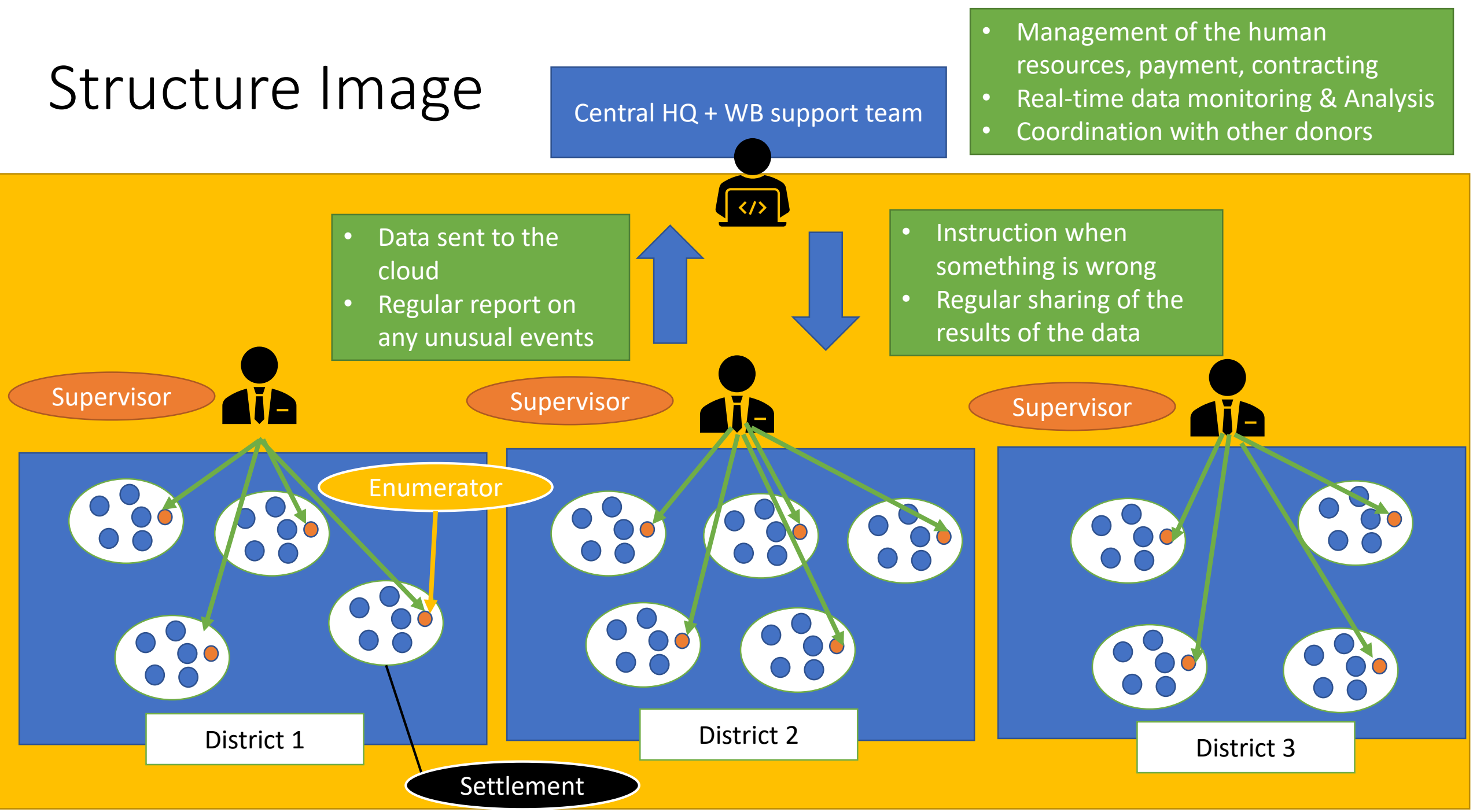
- Survey preparation includes many activities:
  - Design sampling with help of National Statistics Office or using satellite images
  - Design questionnaires; hire/train enumerators; monitoring of data collection
- As a result, the data collection with SWIFT costs $50k - $100k and requires 2 – 3 months of preparation



WORLD BANK GROUP

SWIFT Survey of Well-being via Instant and Frequent Tracking

Structure Image



Furthermore – Similar survey setup are redone by multiple agents eg Malawi's three rapid monitoring surveys

USAID

- Health survey
- Survey preparation and data collection

IFPRI and DFID

- Vulnerability assessment
- Survey preparation and data collection

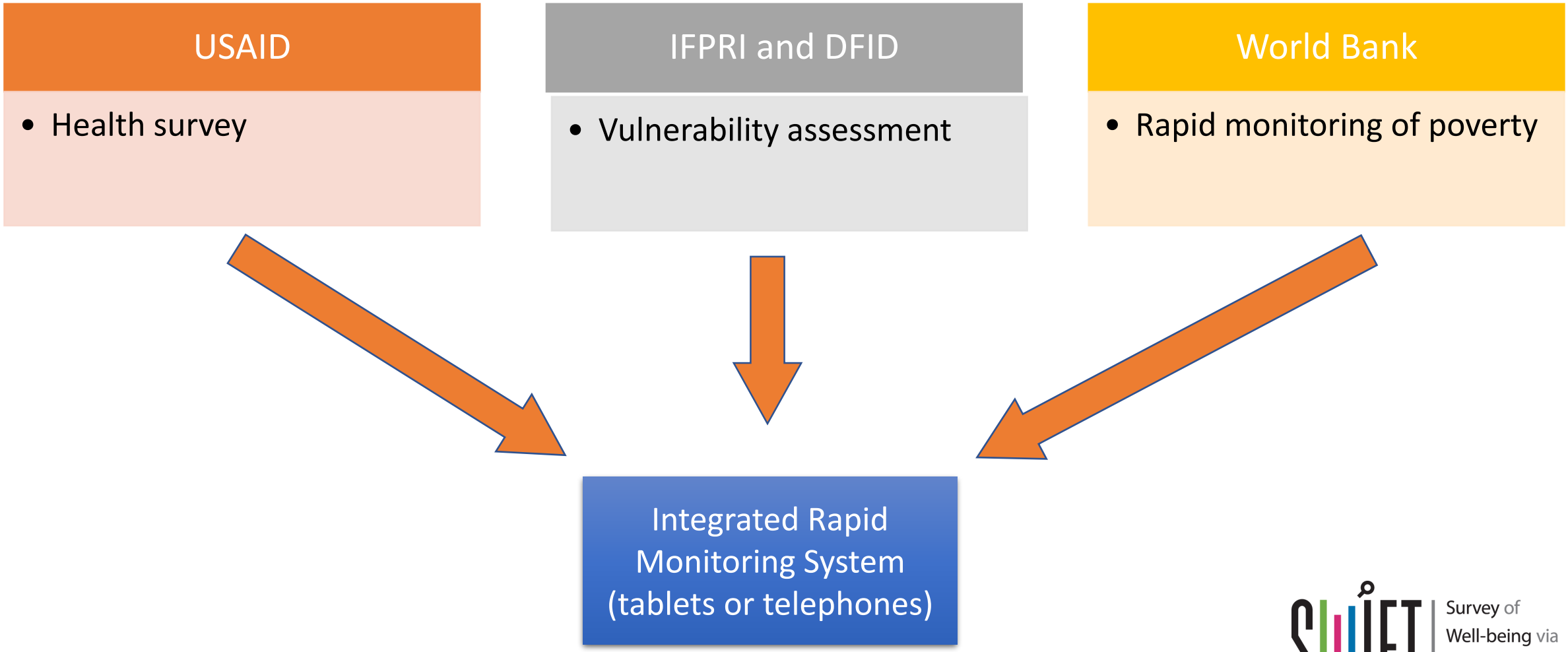
World Bank

- Rapid monitoring of poverty
- Survey preparation and data collection

WORLD BANK GROUP

SWIFT Survey of Well-being via Instant and Frequent Tracking

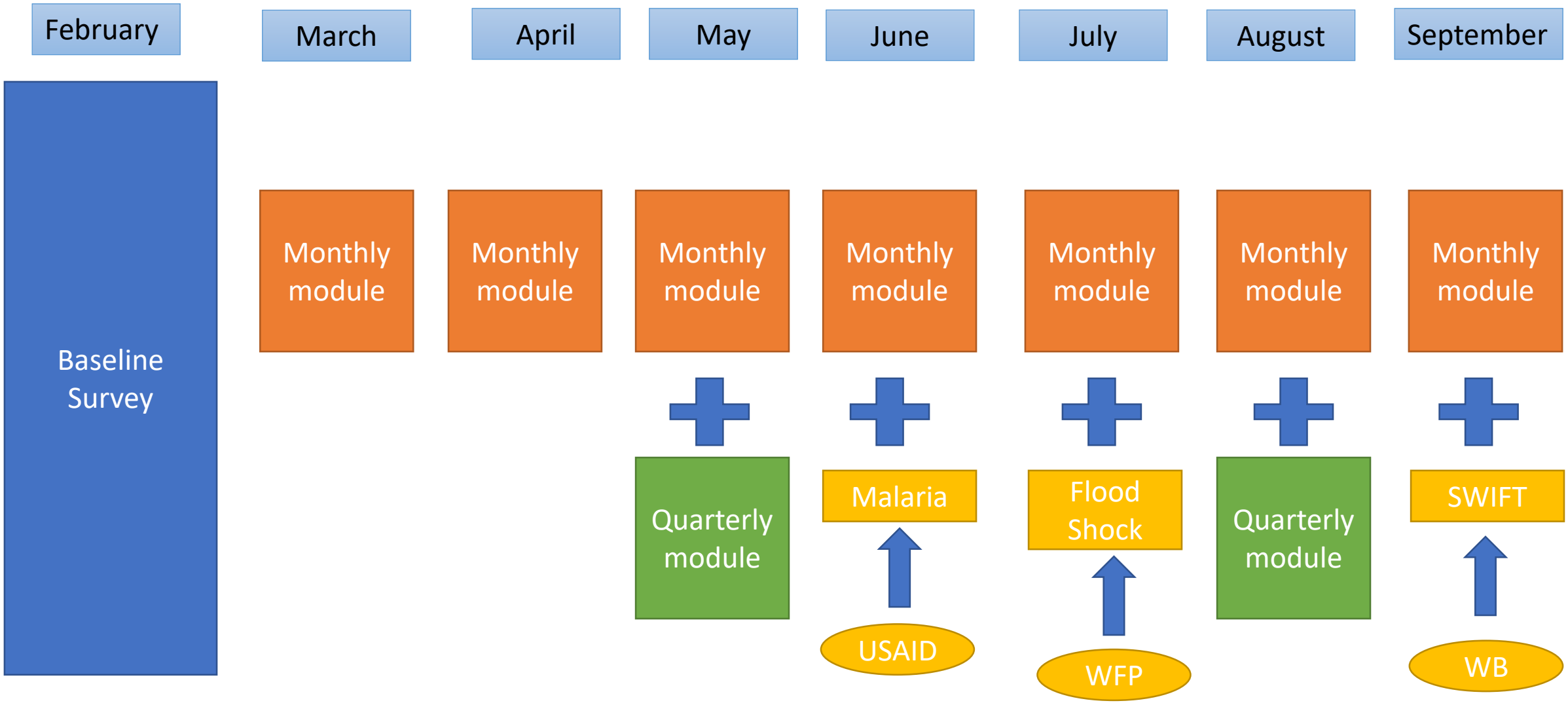
We propose an Integrated Rapid Monitoring System



WORLD BANK GROUP

SWIFT Survey of Well-being via Instant and Frequent Tracking

Structure of the Questionnaire: Example



Thank you!

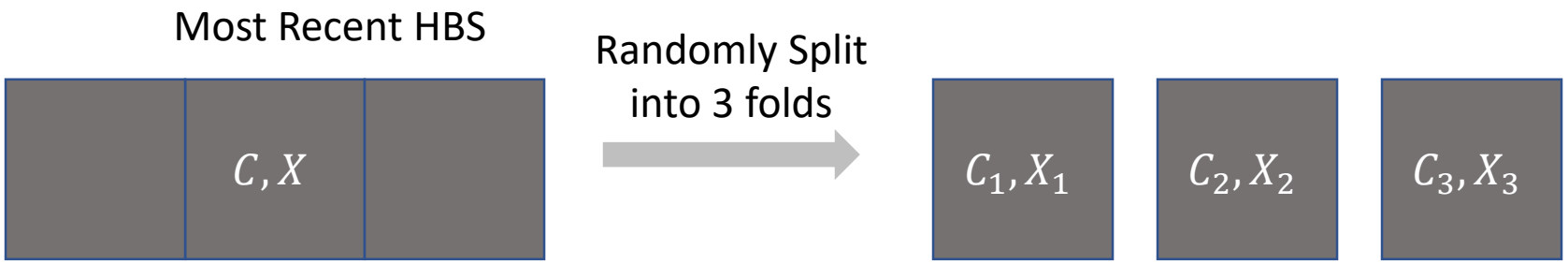
Xiaomeng Chen
xchen8@worldbank.org

Nobuo Yoshida
nyoshida@worldbank.org

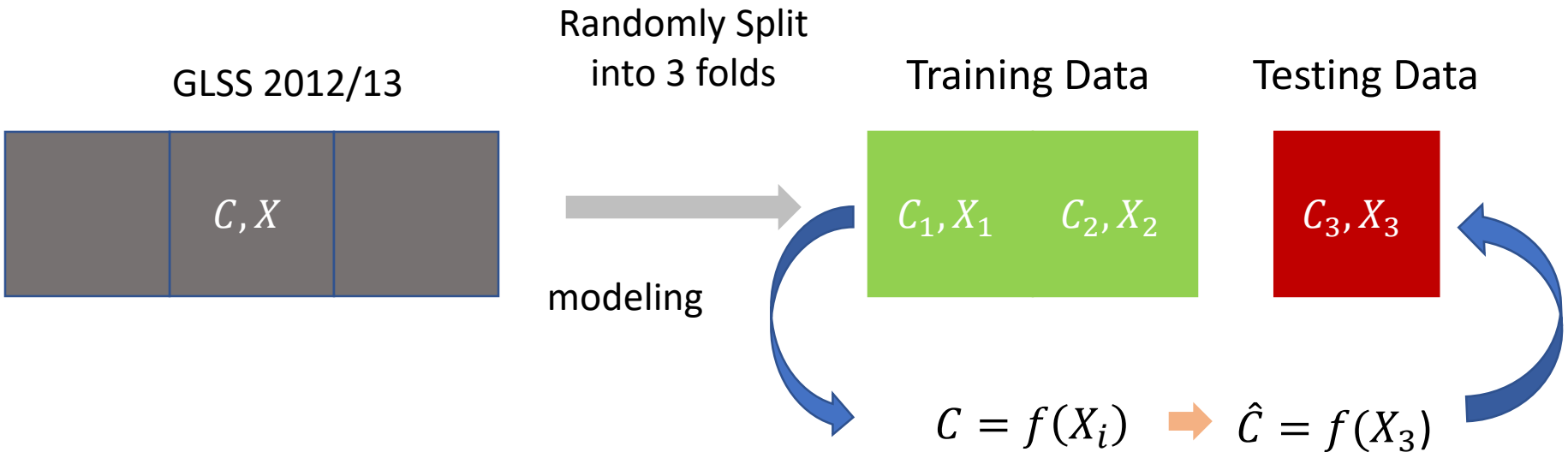
SWIFT | Survey of Well-being via Instant and Frequent Tracking

Annex

Cross-Validation: Step 1

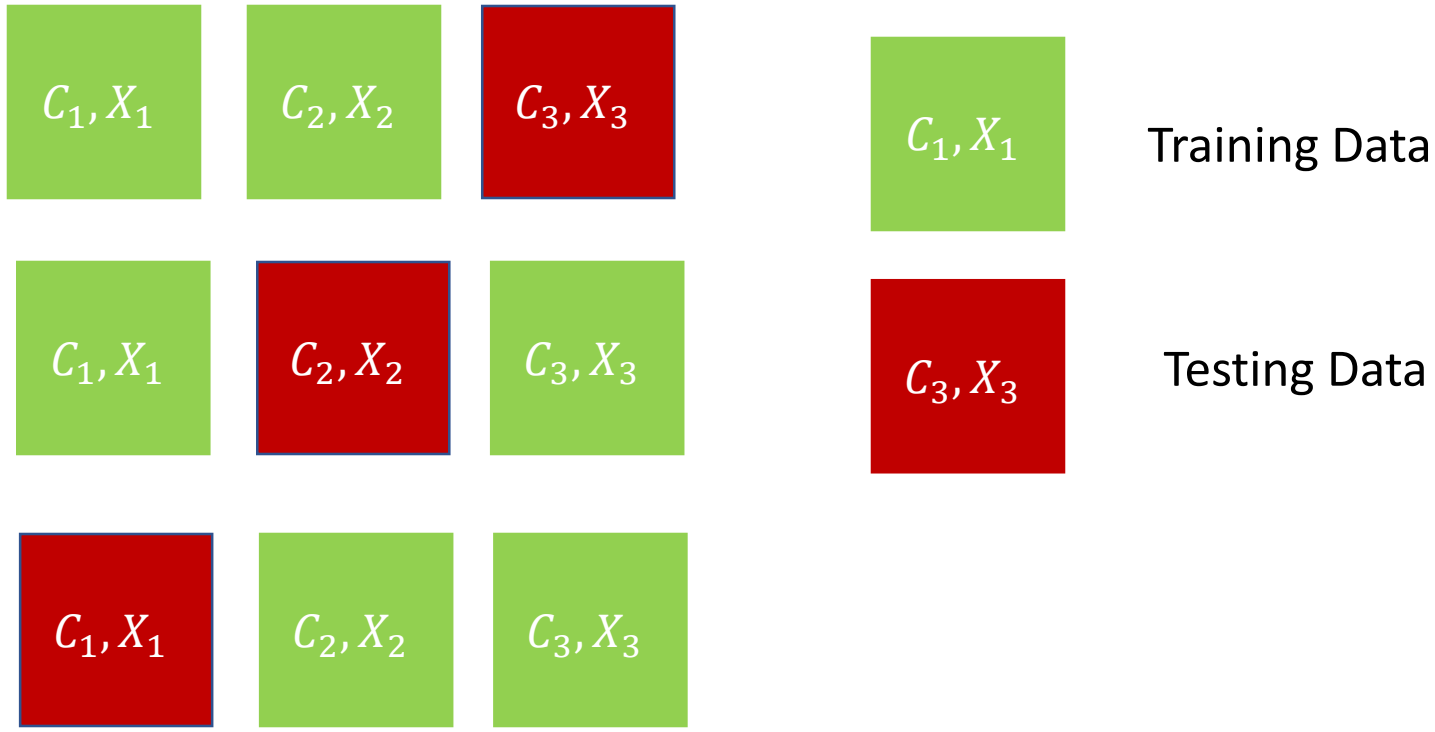


Cross-Validation: Step 2



Compare *Mean Square Error* and *Absolute difference* between original and simulations

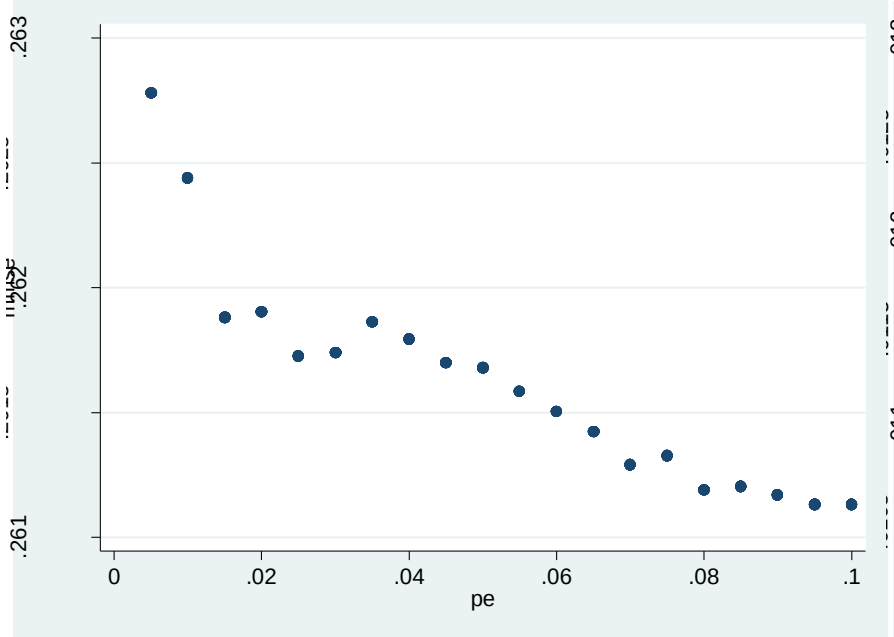
Cross-Validation: Step 3



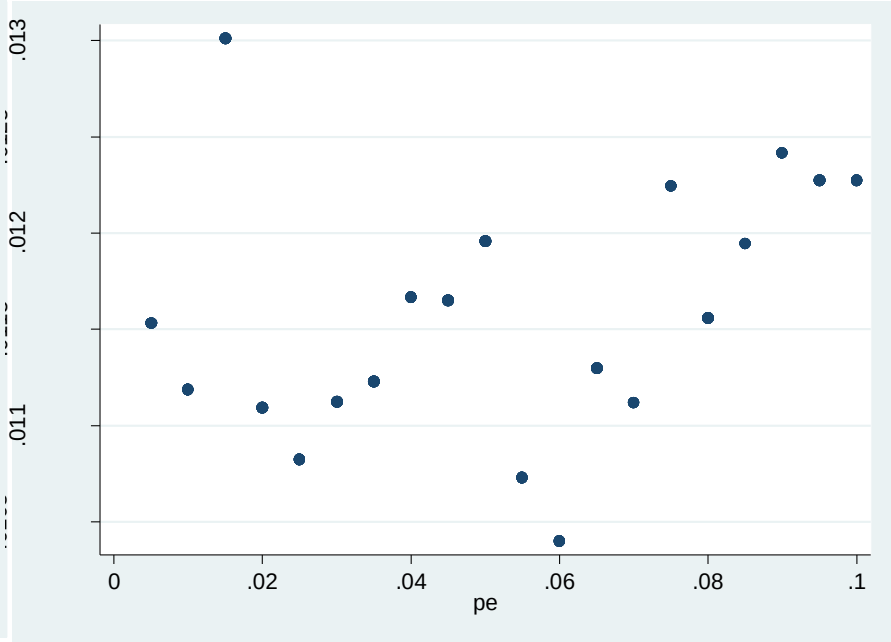
SWIFT Survey of Well-being via Instant and Frequent Tracking

Cross-Validation: Select the best p-value for stepwise regression

P-value = 0.06 is the optimal p-value



Mean Squared Errors



Average of absolute differences between actual and projected poverty rates