Food and Agriculture Organization of the United Nations

SUSTAINABLE DEVELOPMENT GOALS

An indirect estimation approach for SDG indicators using survey data:

Case study based on SDG Indicator 5.a.1 on equal tenure rights for women on agricultural land

Outline Uros Conference 2022 - indirect estimation approaches

Outline

1. Overview
2. Projector Estimator
3. Implementation
4. Conclusions

1. Overview (1/4)

SDG Indicator 5

GOAL 5

Achieve gender equality and empower all women and girls

TARGET 5.1	TARGET 5.2	TARGET 5.3	TARGET 5.4	TARGET 5.5	TARGET 5.6	TARGET 5.A	TARGET 5.B	TARGET 5.C

Undertake reforms to give women equal rights to economic resources as well as access to ownership and control over land and other forms of property, financial services, inheritance and natural resources, in accordance with national laws

1. Overview (2/4)

Indicator 5.a.1: sub-indicators A and B

INDICATOR 5.a.1 has two sub-indicators:

a	Percentage of total agricultural population with ownership or secure rights over agricultural land, by sex;	measures how prevalent ownership / secure rights over agricultural land is in the reference population (agricultural population), by sex
b	Share of women among owners or rights-bearers of agricultural land, by type of tenure	monitors the share of women in ag households with ownership or secure rights over agricultural land over the total individuals with ownership / secure rights

1. Overview (3/4)

- The presentation aims to present the use of R in statistical analysis to develop and implement indirect methods to produce more accurate proxy estimates through the use of the so-called ***"projector estimator"*** proposed by Kim and Rao (2012)

- As a case study, we have considered the Nigeria Household Survey Panel (LSMS) 2019, which follows the **recommended methodology** for generating estimates for SDG indicator 5.a.1. **On the contrary,** the Demographic and Health Survey (DHS) 2018 has some limitation with regards to the recommended methodology, ***leading to a generation of proxy estimates***

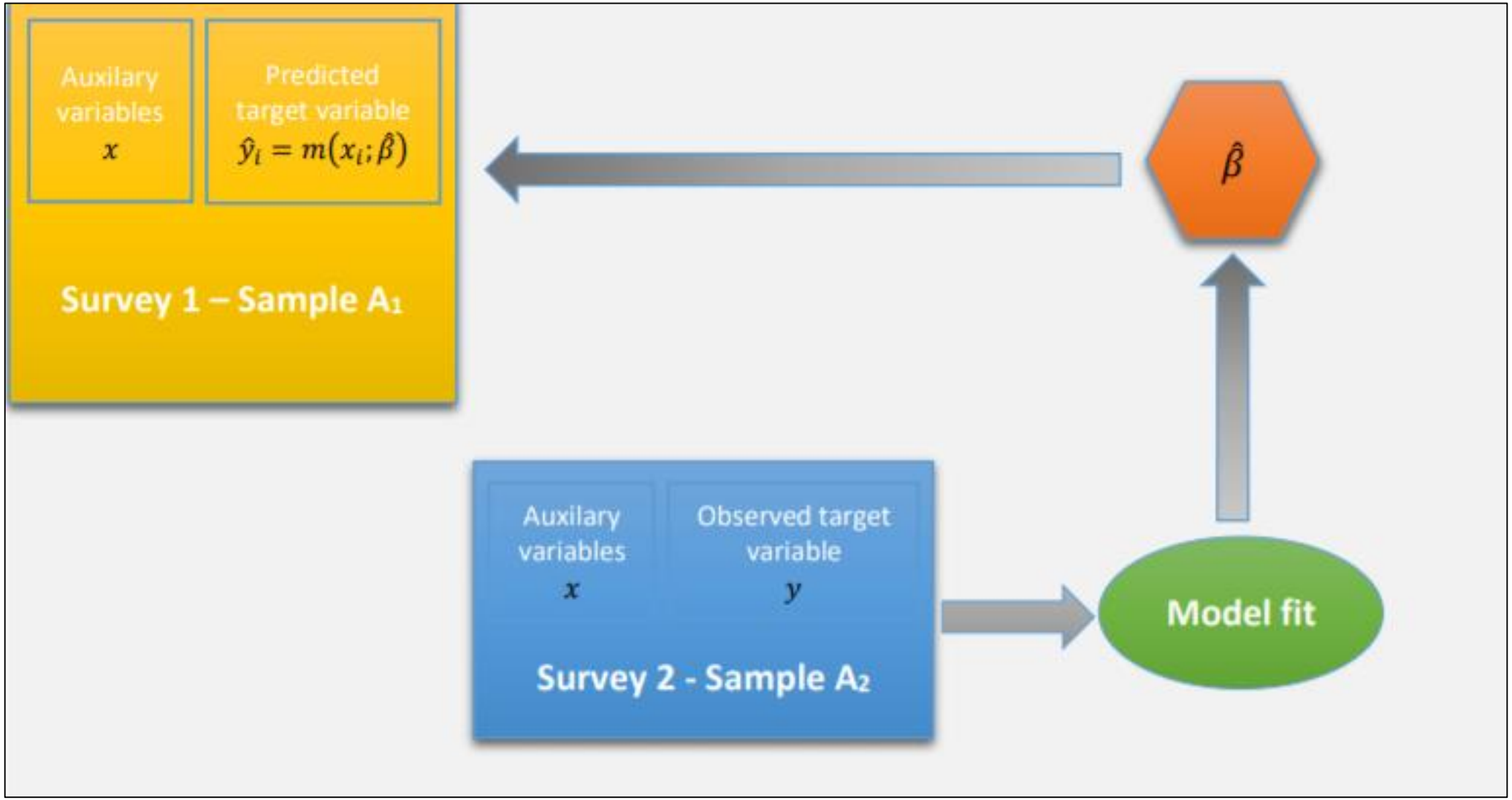
1. Overview (4/4)

Differences across the LSMS and DHS' survey

	DHS	LSMS
Population covered	Women 15-49, Men 15-49/59	Whole population
Exclude wage workers	Man questionnaire does not have a specific question on self-employment	Yes
Agricultural land	Yes	Yes
Ownership of land	Data on ogricultural land ownership refers to household ownership. Individual ownership is asked for any type of land	Yes
Agriculture population	No specific provision for sampling agricultural households	Specific provision for sampling agricultural households
Reference period	12 months	12 months
Alienation rights	Missing	Available

2. Projector estimator

Explanation



Steps for the implementation

1) **Identifying and recoding auxiliary variables**
   - basic R commands for recoding and harmonizing have been used for this task
2) Selection of variables to be included in the model
   - Boruta feature selection method of Kursa and Rudnicki (2010), implemented in the **Boruta** package
   - Dominance Analysis for relative score importance: *domin()* function in the **domir** package
3) Definition of the working model and estimation of projection parameters
   - Weighted multinomial logistic regression: *svyglm()* function in the **survey** package
   - Assess model performance: pseudo-rsquared statistics (*psrsq*() function, **survey** package) , ROC (*roc*() function, **pROC** package)
4) Computation of synthetic values and estimates
   - Get synthetic values using the *predict()* function and estimates using the **ReGenesees** package

First step (1/2)

Identifying and recoding auxiliary variables

LSMS	DHS	*Note*
What is [NAME]'s relationship to the head of household?	What is the relationship of (NAME) to the head of the household?	Answers grouped by: Head; Wife; Child; Other relatives; No relatives
What is the sex of [NAME]?	Is (NAME) male or female?	
Age	How old is (NAME)?	Answers group by: -Inf, 24; 25,34; 35,44; 45,59; 60,+Inf
What is [NAME]'s marital status?	What is (NAME)'s current marital status?	Answers grouped by: Married; Ex married; Never married
Has [NAME] ever attended school?	Has (NAME) ever attended school?	
Household size (retrieved)	Number of household members	
State	State	
Zone	Zone (retrieved)	
Urban/Rural	Urban/Rural	

First step (2/2)

Identifying and recoding auxiliary variables

LSMS	DHS	*Note*
What is the main source of drinking water for the household? (Dry or rainy season)	What is the main source of drinking water for members of your household?	
What is the main source of water used by members of your household for other purposes, such as cooking and hand washing?	What is the main source of water used by your household for other purposes such as cooking and handwashing?	
Do you do anything to the water to make it safer to drink?	Do you do anything to the water to make it safer to drink?	
Does [NAME] have access to a mobile phone?	Do you own a mobile phone?	
Have you worked in the last 7 days? (retrieved)	Aside from your own housework, have you done any work in the last seven days?	Employment in RuLIS (and ILO) refers to the previous 7 days

Steps for the implementation

1) Identifying and recoding auxiliary variables

- basic R commands for recoding and harmonizing have been used for this task

2) Selection of variables to be included in the model

- Boruta feature selection method of Kursa and Rudnicki (2010), implemented in the **Boruta** package
- Dominance Analysis for relative score importance: *domin()* function in the **domir** package

3) Definition of the working model and estimation of projection parameters

- Weighted multinomial logistic regression: *svyglm()* function in the **survey** package
- Assess model performance: pseudo-rsquared statistics (*psrsq*() function, **survey** package) , ROC (*roc*() function, **pROC** package)

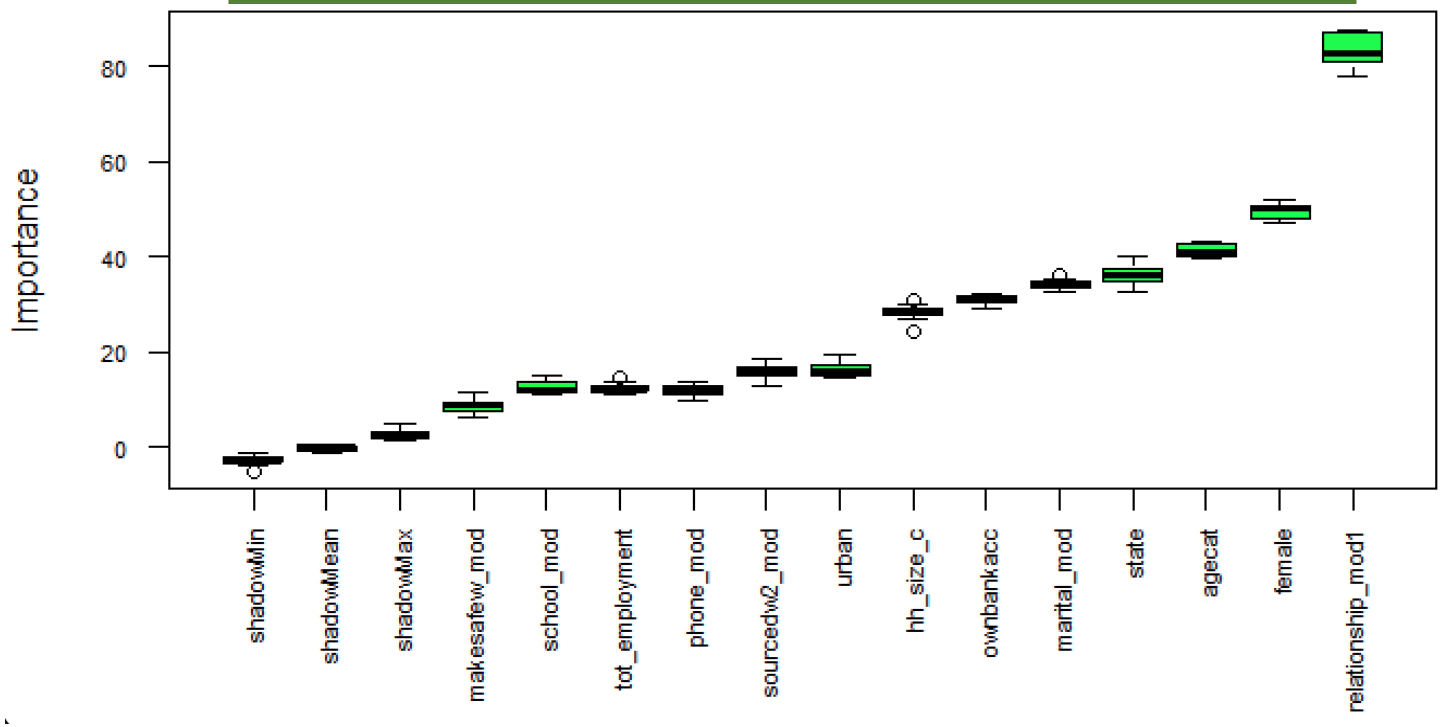
4) Computation of synthetic values and estimates

- Get synthetic values using the *predict()* function and estimates using the **ReGenesees** package

Second step

Selection of variables to be included in the model

BORUTA Model



Score relative (%)	
Relationship w/ hhead	41%
Sex	19%
Age	11%
State	10%
Marital Status	6,5%

Note: kept variables that score above the minimum threshold of 5%

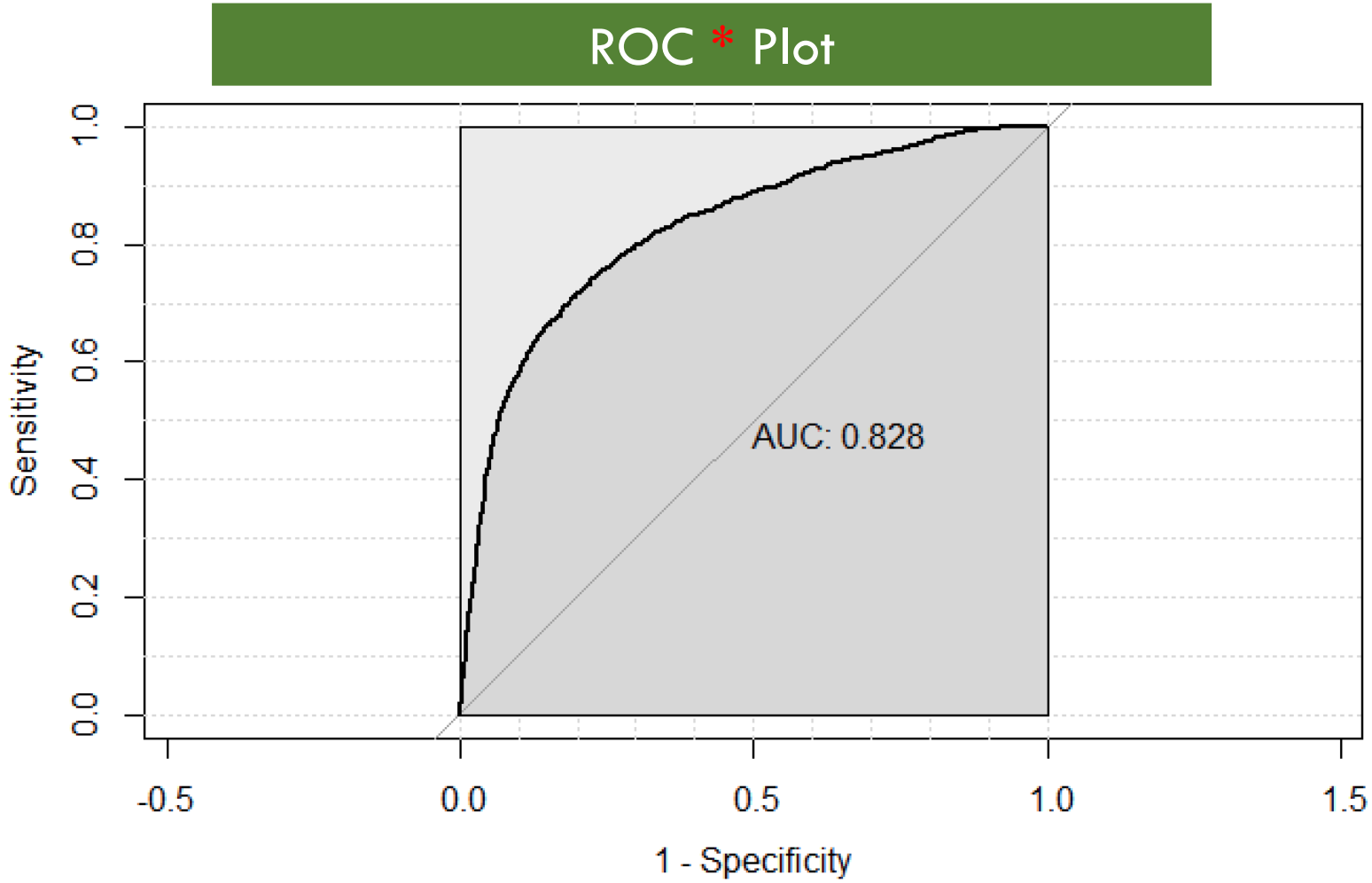
- Eventually use a different selection method such as LASSO, **glmnet** package

Steps for the implementation

1) Identifying and recoding auxiliary variables
   - basic R commands for recoding and harmonizing have been used for this task
2) Selection of variables to be included in the model
   - Boruta feature selection method of Kursa and Rudnicki (2010), implemented in the **Boruta** package
   - Dominance Analysis for relative score importance: *domin()* function in the **domir** package
3) **Definition of the working model and estimation of projection parameters**
   - Weighted multinomial logistic regression: *svyglm()* function in the **survey** package
   - Assess model performance: pseudo-rsquared statistics (*psrsq*() function, **survey** package) , ROC (*roc*() function, **pROC** package)
4) Computation of synthetic values and estimates
   - Get synthetic values using the *predict()* function and estimates using the **ReGenesees** package

Third step

Goodness of fit of the working model



* An ROC curve (receiver operating characteristic curve) is a graph showing **the performance of a classification model at all classification thresholds**

Steps for the implementation

1) Identifying and recoding auxiliary variables
   - basic R commands for recoding and harmonizing have been used for this task
2) Selection of variables to be included in the model
   - Boruta feature selection method of Kursa and Rudnicki (2010), implemented in the **Boruta** package
   - Dominance Analysis for relative score importance: *domin()* function in the **domir** package
3) Definition of the working model and estimation of projection parameters
   - Weighted multinomial logistic regression: *svyglm()* function in the **survey** package
   - Assess model performance: pseudo-rsquared statistics (*psrsq*() function, **survey** package) , ROC (*roc*() function, **pROC** package)
4) **Computation of synthetic values and estimates**
   - Get synthetic values using the *predict()* function and estimates using the **ReGenesees** package

Fourth step - Results

Computation of synthetic values and estimates

	5.a.1(a) Total	5.a.1(a) Woman	5.a.1(a) Male	5.a.1(b)	Note
LSMS 2019	39,0%	23,5%	56,1%	31,6%	Formal ownership + alienation rights
DHS 2018 indirect estimates *(calibrated)*	*44,2%*	*25,3%*	*65,1%*	*30,1%*	Formal ownership + alienation rights
DHS 2018	*38,6%*	*24,4%*	*59,1%*	*37,2%*	*Reported ownership**

- The sex ratio of the DHS agricultural population has been calibrated based on the The Nigeria Household Survey Panel (LSMS) agricultural population

*Reported ownership is a relatively less reliable than documented ownership. However, it has been used because the Nigeria DHS' survey has scarce data on formal documentation along with missing information on alienation rights.

Conclusions

Possible next steps and conclusions:

1. Use the indirect estimation approach in surveys to produce time series estimations with respect to country's law evolution on land ownership
2. Produce new data point to compute cross and integrated analyses between different SDG indicators
3. Scale up activities with countries

What do you think?

Thank you!

Food and Agriculture Organization of the United Nations

SUSTAINABLE DEVELOPMENT GOALS

Stefano Di Candia
FAO OCS
Stefano.DiCandia@fao.org

Raymond J Shama
FAO OCS
raymond.shama@fao.org

http://www.fao.org/sustainable-development-goals/indicators/5.a.1/en/