



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Federal Department of Economic Affairs,  
Education and Research EAER

**Agroscope**

# Dealing with the problem of missing or incomplete off-farm income data

**Swetlana Renner**

3rd October 2017

[www.agroscope.ch](http://www.agroscope.ch) | good food, healthy environment



# Motivation

- Primary objective of the Swiss-FADN is to evaluate the financial situation in the agricultural sector and the income situation of the farm household, including the off-farm income.
- Participation is voluntary, the consent and co-operation of farmer is crucial for the survey.
- Farmers may refuse to provide income information because they consider it too private or sensitive or do not know their income.
- Problem of missing off-farm income information and how to deal with the item non-response



# Data sources

- Agricultural Structure Survey (available for all farms)  
*i.e. utilized agricultural area, livestock units*
- Financial accounting (mandatory information)  
*i.e. farming income, costs, balance sheet, off-farm income*
- Tax return (not mandatory information)  
*i.e. off-farm income, household expenditures*
- Additional information provided by data supplier  
(farmer or/and accounting company)  
*i.e. labour input, estimated private consumption*

# Components of household income/spending

## Household income:

- Farming income
- Off-farm business
- Non-farm employment (wages/salaries of operator/spouse)
- Other off-farm income, including:
  - Securities income
  - Family & child allowance
  - Pensions, sick pay and other state transfers
  - Gains realized by sale of real estate

## Household spending:

- Payments in pension funds and insurances
- Private consumption



# Completeness of the information

*Does the information declared in the survey provide a complete picture of the household financial situation with regard to ...?*

	Incomplete cases		Share of additional information from tax return
	Number	Share	
Off-farm business	113	6%	30%
Non-farm employment	89	5%	39%
Other off-farm income	143	7%	14%
Private consumption	155	8%	5%

# Missing Data or Zero

Information from financial accounting: complete (mandatory)

Tax return: differentiate between a zero and missing data

Complete cases: share of zero values

Incomplete cases: share of missing values (might be zero)

	Financial accounting		Tax return	
	Complete cases	Incomplete cases	Complete cases	Incomplete cases
Off-farm business	97%	100%	98%	99% (112)
Non-farm employment	40%	66%	72%	85% ( 76)
Other off-farm income	28%	36%	52%	80% (114)

  
Treat as missing values



# How to deal with missing data?

## 1) Strict: Do not accept incomplete data

Negative impact on response rate, introduces additional bias

## 2) Radical: Delete incomplete data from the sample

Reduces net sample size, increases costs per survey unit, increases variance, introduces additional bias

## 3) Ignorant: Treat missing data as Zero-Values

Systematically underestimated means, magnitude of bias should be estimated and communicated

## 4) Appropriate: Handle by imputation methods

Choice of the procedure to “fill in” the missing values



# Considered imputation methods

## 1) Unconditional Mean (Median) Imputation

Missing value = mean (or median) of the recorded values

Off-farm income is underestimated due to high portion of Zero-values, distorts the distribution of the variable, influenced by extreme values

-> not recommended (Little and Rubin, 2002)

## 2) Regression Imputation (REG)

Normality assumption is violated due to high portion of Zero-Values.

Considering only non-zero values is not appropriate due to the skewed distribution (leads to negative imputed values) (Paul & Hippel, 2013)

## 3) Predictive Mean Matching (PMM) (Rubin 1987, Little 1988)

Imputations are restricted to the observed values, semi-parametric

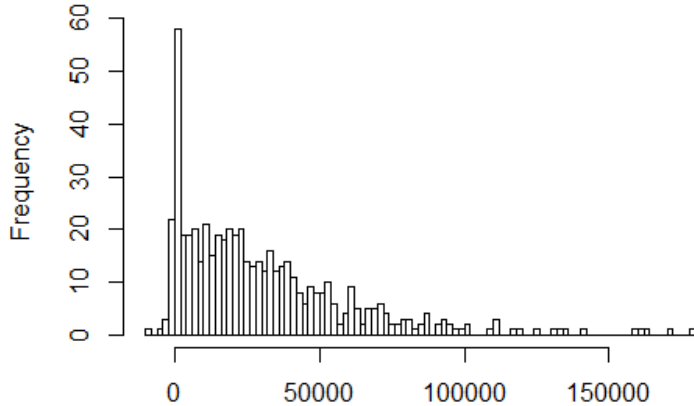
Measure distance between units based on the predicted value from the specified regression model. Imputed value comes from responding (complete) units close to the unit with the missing value (few donors?).





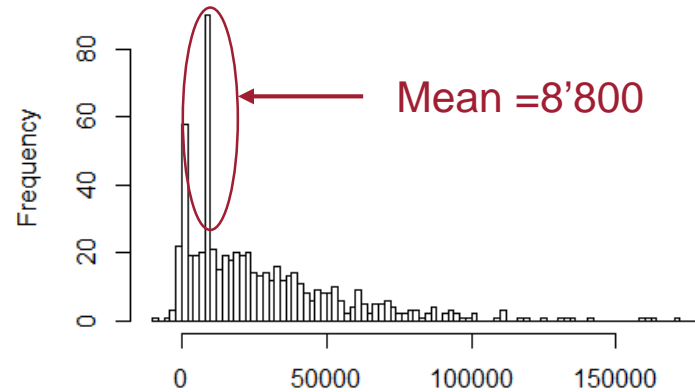
# Distribution (non-farm employment)

Before imputation



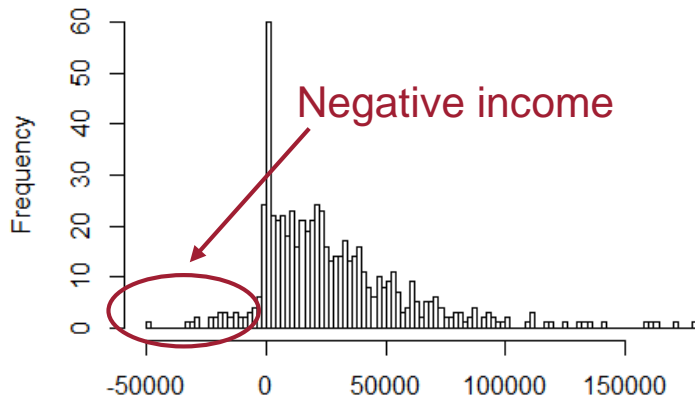
Income from non-farm employment

Mean imputation



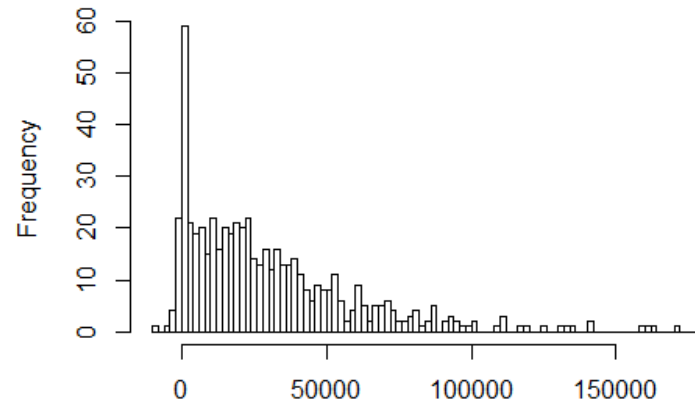
Income from non-farm employment

Regression imputation



Income from non-farm employment

PMM imputation



Income from non-farm employment

# Model specification

Off-farm business & non-farm employment: Only complete cases with non-zero-values are used as predictors for imputation of missing values

Other off-farm income: All complete cases are used for imputation

Variables used for imputation models	REG	PMM
Strata variables (region, type of farming)		X
Off-farm-income from financial accounts		X
Farm income, assets	X	X
Family size (consumer units, family members total and without income)	X	X
Working Days (on-farm, off-farm, employment)	X	X
Age & Qualification (agricultural and general education of farm manager and spouse)	X	X

Software: R package MICE



# Comparison of Results

Off-farm income in 2016 (weighted average in CHF) estimated by different imputation methods

Imputation method	Off-farm business	Non-farm employment	Other off-farm income	Total off-farm income	
Median (NULL)	1'908	20'846	7'884	30'638	status quo
Mean	1'948	21'191	7'816	30'955	1.0%
Mean (non-zero)	3'780	21'535	7'834	33'150	8.2%
Regression	1'991	21'017	7'836	30'844	0.7%
Regression (non-zero)	4'039	21'912	8'072	34'022	11.0%
PMM	1'940	20'987	7'863	30'789	0.5%
PMM (non-zero)	5'234	21'720	7'716	34'669	13.2%
<b>Combined</b>	<b>3'780</b>	<b>21'720</b>	<b>7'863</b>	<b>33'363</b>	8.9%

# Conclusions

## **Pro Imputation:**

- More precise estimation of the off-farm income and total household income (?)
- Use multiple imputation and interval estimates to evaluate sensitivity of the imputation methods

## **Contra Imputation:**

- Additional bias might be introduced (if only few responding cases available)
- Communication of methods in standard publications (data users think that off-farm income data was complete)
- Issues of providing farm-level data to the external users

# Current solution and future ressearch

## Information provided in standard publications:

“Some farms do not fully compile the data on the off-farm income. This concerns the supplementary information on budget income from the tax declaration as well as the estimate of budget expenditure (private consumption). Although the proportion of incomplete data is relatively small, this leads to a lower estimate of total income and private consumption. The extent of the distortion is about -X% for non-agricultural income.”

## Alternative procedure for future:

Income estimation from the off-farm business and off-farm employment based on the available information about working days and estimated daily rates / wages



**Thank you for your attention**

**Swetlana Renner**

swetlana.renner@agroscope.admin.ch

**Agroscope** good food, healthy environment

www.agroscope.admin.ch





# Literature

Little RJA, Rubn DB (2002). *Statistical Analysis with Missing Data*, 2<sup>nd</sup> edition. John Wiley & Sons, New York

Little RJA (1988). Missing Data Adjustments in Large Surveys, in *Journal of business Economics and Statistics*, 6, 287-301

von Hippel PT (2013) Should a Normal Imputation Model Be Modified to Impute Skewed Variables? In *Sociological Methods and Research*, 2013, 42(1), 105-138

Rubin DB (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York

Van Buuren S, Groothuis-Oudshoorn K (2011). Mice: Multivariate Imputation by Chained Equations in R, in *Journal of Statistical Software*, 45 (3)