Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für Wirtschaft,
Bildung und Forschung WBF

**Agroscope**

# Sources of Error in Swiss FADN Survey and Adjustment for Nonresponse

## Swetlana Renner

29.10.2015



Agroscope

# Content

## 1. Types and sources of error

- ✓Coverage (frame) error
- ✓Sampling error
- ✓Nonresponse error

## 2. Weighting methods

- ✓Weighting class adjustment
- ✓Poststratification
- ✓Response propensity model
- ✓Calibration

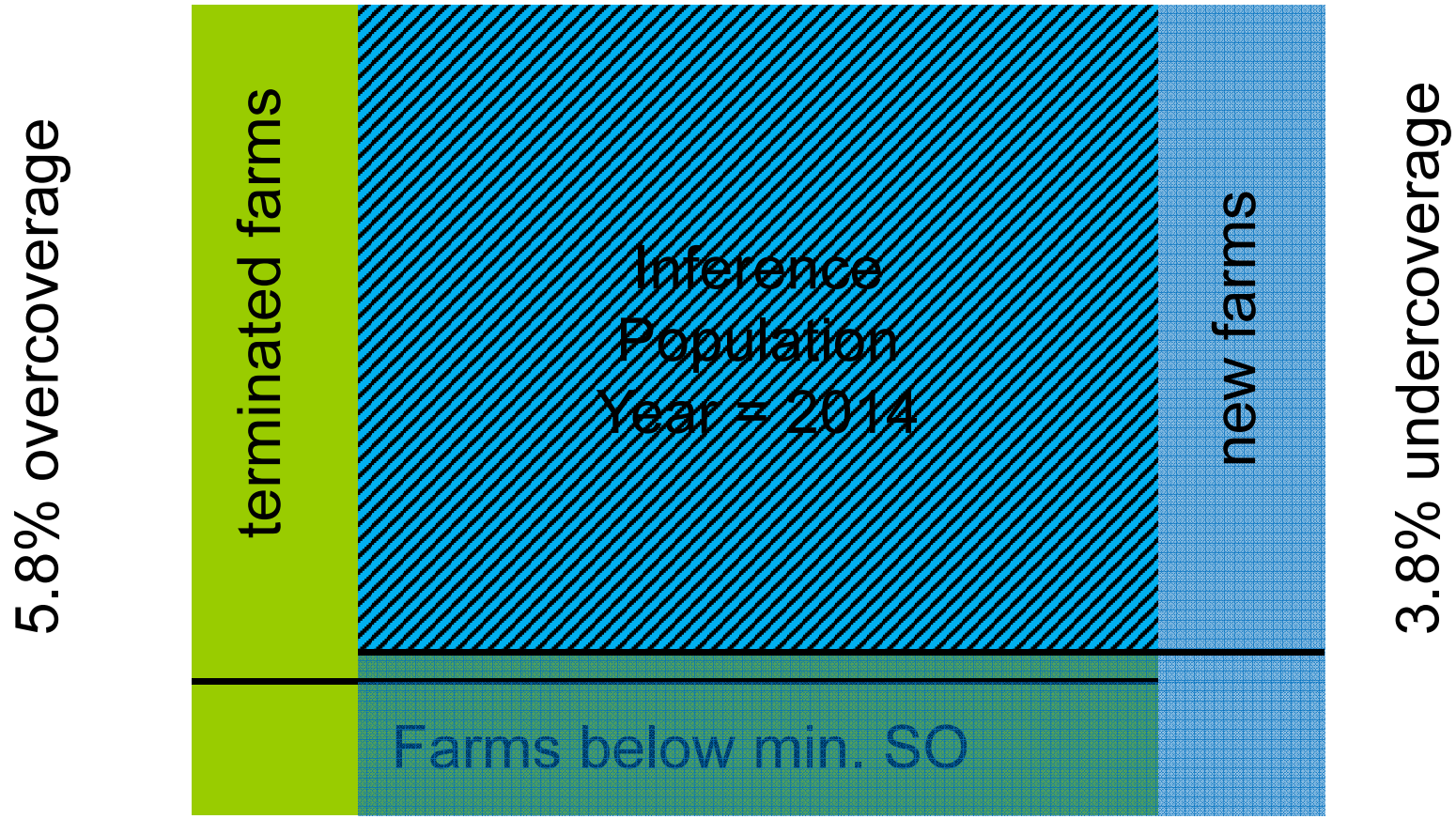
## 3. Results

Agroscope

# Main types and sources of error

1. **Coverage (frame) error**
   - ✓Quality of the sampling frame and its completeness for the target population
2. **Sampling error**
   - ✓Caused by observing a sample instead of the whole population, can be estimated for the random sample
3. **Nonresponse error**
   - ✓Failure to measure some of the units in the selected sample (e.g. person is not reached or refuses to participate the survey)

# Coverage (frame) error




Covarage rate = 94.4%

# Sampling error

Refers to the **expected variation in estimates** due to the random selection scheme used to select a sample.

**Probability sampling** enables the estimation of the sampling error. Formula of the **variance** of the estimator for **stratified random samples**:

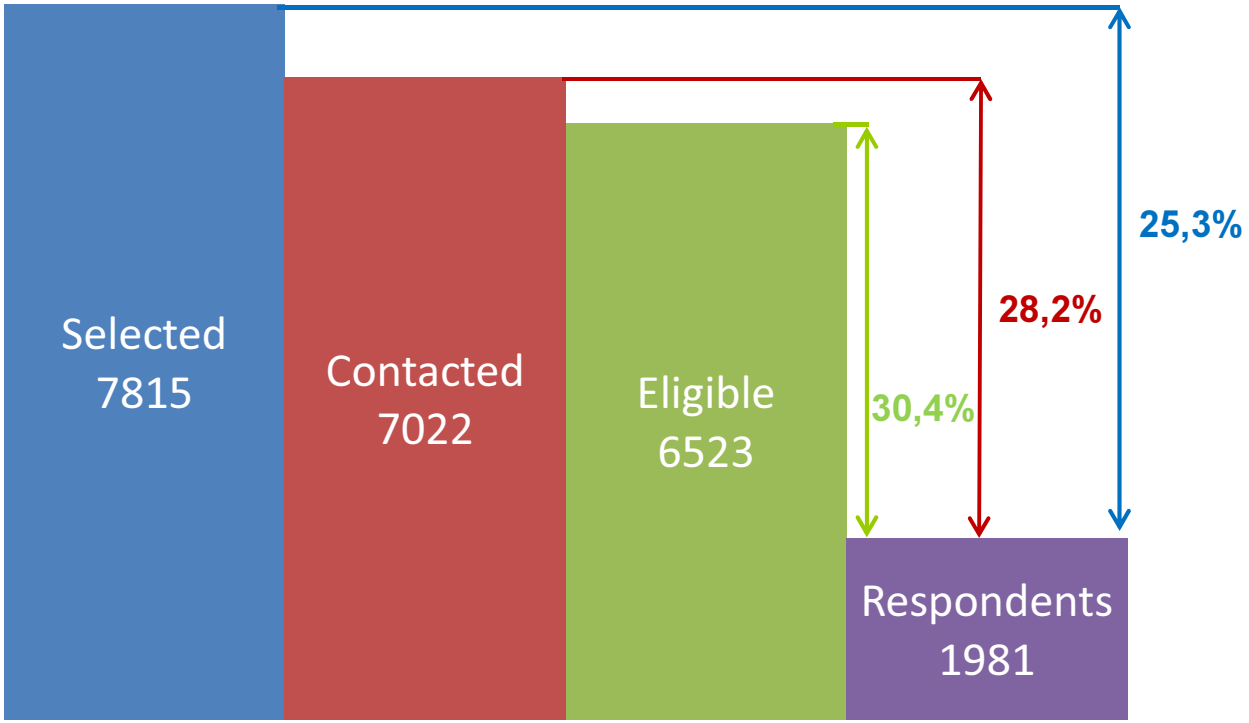
$$\mathrm{Var}(\bar{y}_{str}) = \sum_{h=1}^{H} \left(1 - \frac{n_h}{N_h}\right) \left(\frac{N_h}{N}\right)^2 \frac{s_h^2}{n_h}$$

$\bar{y}_{str}$ – estimated mean, $N_h$ – Number of units in $h$th stratum, $n_h$ – Number of sampling units in $h$th stratum, $s_h^2$ – sample variance in $h$th stratum

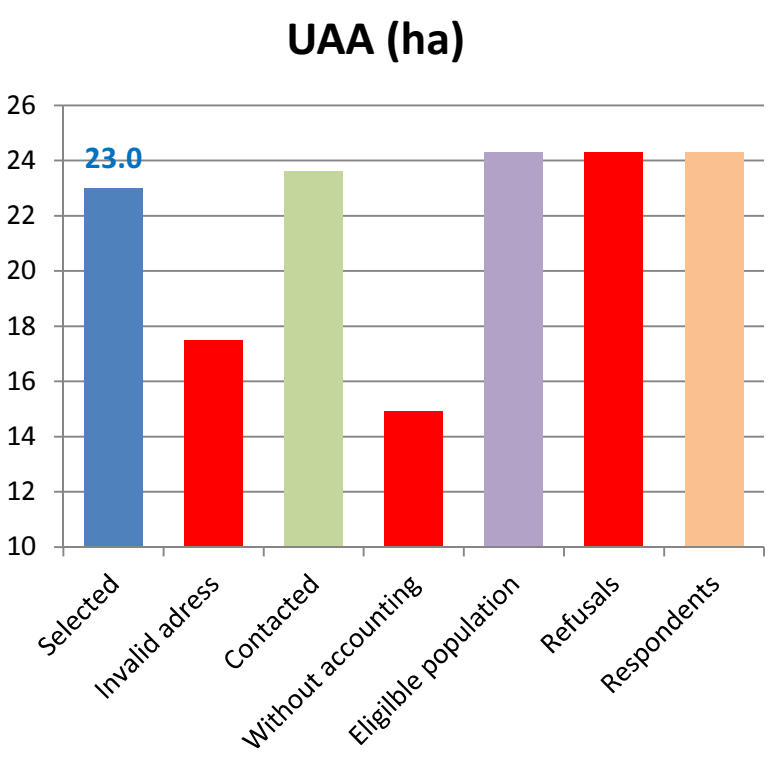
For more complex surveys: Taylor-series approximation or replication methods of variance estimation

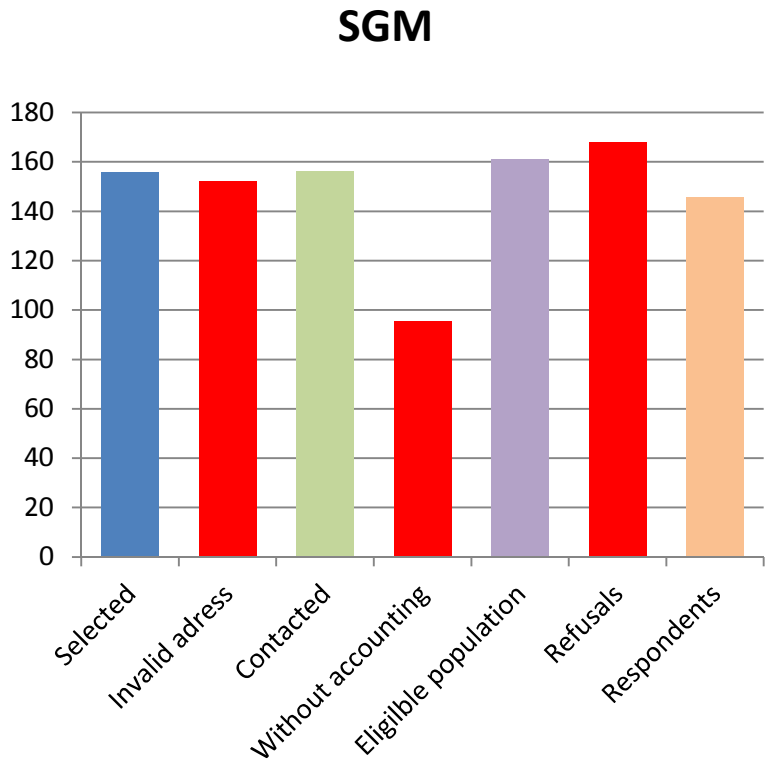
# Response Rate

$$\text{Response Rate} = \frac{\text{Number of respondents}}{\text{Total number of contacted farms}} = 28{,}2\%$$



# Estimated mean differences



# Weighting procedures

**Idea**: increase the weights of the respondent farms to represent the nonrespondents using some additional information and assumptions

**Adjusted Horvitz-Thompson Estimator:**

$$\bar{y}_{HT} = \sum_{j \in R} \tilde{w}_j y_j \cdot \left(\sum_{j \in R} \tilde{w}_j\right)^{-1} \text{ with } \tilde{w}_j = \frac{w_j}{\hat{\rho}_j} = \frac{1}{\pi_j \hat{\rho}_j}$$

$w_j$ - sampling weight

$\hat{\rho}_j$ - estimated response probability (usually unknown)

-> Use of the auxiliary information

# Poststratification and weighting class adjustment

- Classify respondents and nonrespondents into adjustment cells defined by auxiliary variable
- Auxiliary variable should be correlated with the response rate and the variable of interest
- Weight for respondents in the cell: inverse of the response rate

## Weighting Class Adjustment (WCA)

| WCA | language class | | | Total |
|---|---|---|---|---|
| | GER | FR | IT | |
| sum (w) | 27133 | 9046 | 912 | 37091 |
| sum($w_R$) | 9891 | 1808 | 89 | 11788 |
| **Resp. Prob.** | **0.365** | **0.200** | **0.098** | **0.318** |

Using the information collected during the survey

## Poststratification (PostStr)

| PostStr | language class | | | Total |
|---|---|---|---|---|
| | GER | FR | IT | |
| Number TP | 28759 | 7701 | 631 | 37091 |
| sum($w_R$) | 9891 | 1808 | 89 | 11788 |
| **Resp. Prob.** | **0.344** | **0.235** | **0.142** | **0.318** |

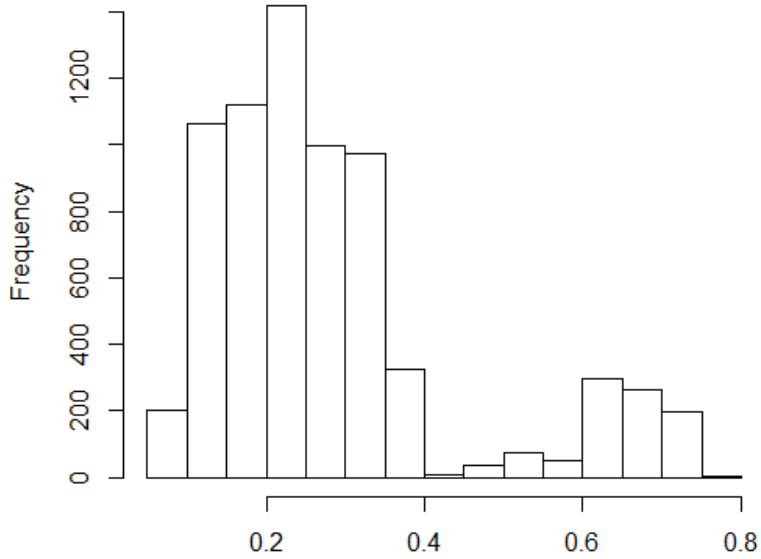
Using the known distribution of the population

**Assumption:** same response probability for all farms in the class

# Response propensity modeling

Regression parametric model (Logit): Indicator of unit non-response is regressed on different variables (language, size, type of farming, panel group)

| | Estimate | Std. Error | Pr(>\|z\|) | |
|---|---|---|---|---|
| (Intercept) | 0.44 | 0.15 | 0.00 | *** |
| FR | -0.62 | 0.07 | 0.00 | *** |
| IT | -1.38 | 0.27 | 0.00 | *** |
| SGMgroup2 | 0.27 | 0.08 | 0.00 | *** |
| SGMgroup3 | 0.39 | 0.09 | 0.00 | *** |
| SGMgroup4 | 0.24 | 0.09 | 0.01 | ** |
| FATTYP_12 | -0.27 | 0.14 | 0.05 | * |
| FATTYP_21 | 0.04 | 0.15 | 0.79 | |
| FATTYP_22 | 0.24 | 0.24 | 0.32 | |
| FATTYP_23 | 0.09 | 0.20 | 0.64 | |
| FATTYP_31 | -0.29 | 0.21 | 0.16 | |
| FATTYP_41 | -0.21 | 0.16 | 0.18 | |
| FATTYP_51 | 0.07 | 0.16 | 0.65 | |
| FATTYP_52 | 0.30 | 0.21 | 0.15 | |
| FATTYP_53 | 0.24 | 0.14 | 0.10 | * |
| FATTYP_54 | -0.01 | 0.15 | 0.92 | |
| Group | -1.66 | 0.08 | 0.00 | *** |



# Calibration

Average distance: $$\sum_{i \in R} \frac{(\tilde{w}_i - w_{0i})^2}{2w_{0i}} \underset{\tilde{w}_i}{\rightarrow} min$$

Minimize the distance between the original weight $w_{0i}$ and the new weight $\tilde{w}_i$ subject to:

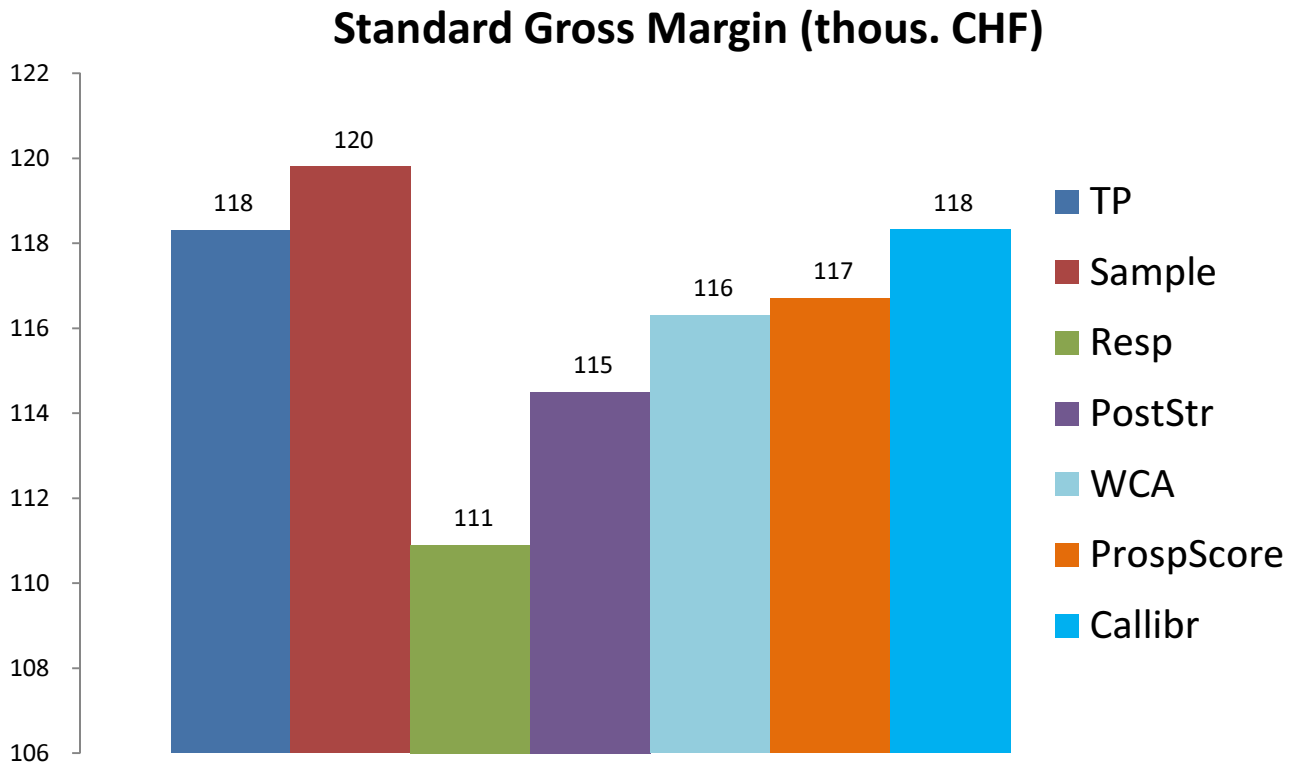
$$\sum_{i \in R} \tilde{w}_{Ci} \cdot x_{Ci} \left( \sum_{i \in R} \tilde{w}_{Ci} \right)^{-1} = \sum_{i \in TP} x_{Ci} \cdot N^{-1}$$

The estimator is calibrated to the known mean of $X$ (can be a set of variables) of the target population (at different levels)

# Calibration

- **Initial weights $w_{0j}$:** sampling weight $w_{hj} = \frac{N_{hj}}{n_{hj}}$
- **Auxiliary/calibration variables:**
  - ✓ Number of farms
  - ✓ UAA (agricultural area)
  - ✓ LU (livestock units)
  - ✓ SO (standard output)
  - ✓ SGM (standard gross margin)
  - ✓ Language structure
- **Levels for calibration:**
  - ✓ Country(1)
  - ✓ Regions (3)
  - ✓ Type of farming (11)
- **Other options:** Specification of the distance function, boundaries for weights (e.g. between 0 and 100)

# Comparison of weighting methods






Agroscope

# Summary

- ✓ Non-response is the major and the most problematic source of error leading to the biased estimates
- ✓ Weighting procedures are used to adjust for non-response
- ✓ The non-response bias can be reduced by using the available auxiliary variables from the data frame or collected during the survey
- ✓ Calibration is the flexible approach making more use of the auxiliary information and providing good performance

**Next steps:**

- ✓ Modifications of the calibration model
- ✓ Calculation and comparison of the variance estimates

Agroscope

# Thank you for your attention

**Agroscope** gutes Essen, gesunde Umwelt