Dealing with non-response in survey data: on the usage of weights

Emmanuel Rousseaux

Institute for Demographic and Life Course Studies

University of Geneva
1211 Geneva 4, Switzerland

emmanuel.rousseaux@unige.ch

LIVES
Swiss National Centre of Competence in Research

Outline

Outline

Outline

Motivation

- We want to test a theory on a population
- But we can only reach a sample of the population
- Hypothesis testing allows to escape from sampling hazard
- But is our sample representative of the population studied?

Motivation

- We want to test a theory on a population
- But we can only reach a sample of the population
- Hypothesis testing allows to escape from sampling hazard
- But is our sample representative of the population studied?

Motivation

- We want to test a theory on a population
- But we can only reach a sample of the population
- Hypothesis testing allows to escape from sampling hazard
- But is our sample representative of the population studied?

Motivation

- We want to test a theory on a population
- But we can only reach a sample of the population
- Hypothesis testing allows to escape from sampling hazard
- But is our sample representative of the population studied?

Motivation

- We want to test a theory on a population
- But we can only reach a sample of the population
- Hypothesis testing allows to escape from sampling hazard
- But is our sample representative of the population studied?

Motivation

- Sampling: generally complex, *at random* is assumed
- Surveying: some individuals don't respond
- Randomly?
- Intuitively no: health problems, family difficulties, problems at work, ...
- Database manager check if marginal distributions differ
- If significative differences are found, weights are provided

Motivation

- Sampling: generally complex, *at random* is assumed
- Surveying: some individuals don't respond
- Randomly?
- Intuitively no: health problems, family difficulties, problems at work, ...
- Database manager check if marginal distributions differ
- If significative differences are found, weights are provided

Motivation

- Sampling: generally complex, *at random* is assumed
- Surveying: some individuals don't respond
- Randomly?
- Intuitively no: health problems, family difficulties, problems at work, ...
- Database manager check if marginal distributions differ
- If significative differences are found, weights are provided

Motivation

- Sampling: generally complex, *at random* is assumed
- Surveying: some individuals don't respond
- Randomly?
- Intuitively no: health problems, family difficulties, problems at work, ...
- Database manager check if marginal distributions differ
- If significative differences are found, weights are provided

Motivation

- Sampling: generally complex, *at random* is assumed
- Surveying: some individuals don't respond
- Randomly?
- Intuitively no: health problems, family difficulties, problems at work, ...
- Database manager check if marginal distributions differ
- If significative differences are found, weights are provided

Motivation

- Sampling: generally complex, *at random* is assumed
- Surveying: some individuals don't respond
- Randomly?
- Intuitively no: health problems, family difficulties, problems at work, ...
- Database manager check if marginal distributions differ
- If significative differences are found, weights are provided

Motivation

- Sampling: generally complex, *at random* is assumed
- Surveying: some individuals don't respond
- Randomly?
- Intuitively no: health problems, family difficulties, problems at work, ...
- Database manager check if marginal distributions differ
- If significative differences are found, weights are provided

Motivation

- Is it mandatory to use weights?
- Do the results really change in my analysis?
- It is bad if I don't use weights?
- Is it worst to use weights than not?

Unfortunately, there is no 'YES' or 'NO' answer

Motivation

- Is it mandatory to use weights?
- Do the results really change in my analysis?
- It is bad if I don't use weights?
- Is it worst to use weights than not?

Unfortunately, there is no 'YES' or 'NO' answer

Motivation

- Is it mandatory to use weights?
- Do the results really change in my analysis?
- It is bad if I don't use weights?
- Is it worst to use weights than not?

Unfortunately, there is no 'YES' or 'NO' answer

Motivation

- Is it mandatory to use weights?
- Do the results really change in my analysis?
- It is bad if I don't use weights?
- Is it worst to use weights than not?

Unfortunately, there is no 'YES' or 'NO' answer

Motivation

- Is it mandatory to use weights?
- Do the results really change in my analysis?
- It is bad if I don't use weights?
- Is it worst to use weights than not?

Unfortunately, there is no 'YES' or 'NO' answer

Motivation

- Is it mandatory to use weights?
- Do the results really change in my analysis?
- It is bad if I don't use weights?
- Is it worst to use weights than not?

Unfortunately, there is no 'YES' or 'NO' answer

Outline

Scope of this presentation

- Is it important to use weights?
- Discussing side effects of weighting
- Assessing the advisability of weighting
- An example :usage of weights in the SHP
- Bonus :-)

Scope of this presentation

- Is it important to use weights?
- Discussing side effects of weighting
- Assessing the advisability of weighting
- An example :usage of weights in the SHP
- Bonus :-)

Scope of this presentation

- Is it important to use weights?
- Discussing side effects of weighting
- Assessing the advisability of weighting
- An example :usage of weights in the SHP
- Bonus :-)

Scope of this presentation

- Is it important to use weights?
- Discussing side effects of weighting
- Assessing the advisability of weighting
- An example :usage of weights in the SHP
- Bonus :-)

Scope of this presentation

- Is it important to use weights?
- Discussing side effects of weighting
- Assessing the advisability of weighting
- An example :usage of weights in the SHP
- Bonus :-)

Scope of this presentation

- Is it important to use weights?
- Discussing side effects of weighting
- Assessing the advisability of weighting
- An example :usage of weights in the SHP
- Bonus :-)

Type of non-response

- Total non-response
- Partial non-response

Type of non-response

- Total non-response
- Partial non-response

Type of non-response

- Total non-response
- Partial non-response

Outline

In the context of weighting it is useful to distinguish two purposes of estimation

- To estimate population descriptive statistics
- To estimate covariable effects

In the context of weighting it is useful to distinguish two purposes of estimation

- To estimate population descriptive statistics
- To estimate covariable effects

In the context of weighting it is useful to distinguish two purposes of estimation

- To estimate population descriptive statistics
- To estimate covariable effects

Outline

Estimating population descriptive statistics: an example

- Poverty rate for the USA in 1967
- Officially measure as 13%
- (Current Population Survey, U.S. Bureau of the Census, 1968)

Estimating population descriptive statistics: an example

- Poverty rate for the USA in 1967
- Officially measure as 13%
- (Current Population Survey, U.S. Bureau of the Census, 1968)

Estimating population descriptive statistics: an example

- Poverty rate for the USA in 1967
- Officially measure as 13%
- (Current Population Survey, U.S. Bureau of the Census, 1968)

Estimating population descriptive statistics: an example

- Poverty rate for the USA in 1967
- Officially measure as 13%
- (Current Population Survey, U.S. Bureau of the Census, 1968)

Estimating population descriptive statistics: an example

- Now we estimate this rate with the PSID survey
- PSID: Panel Study of Income Dynamics, started in 1967
- This survey purposefully overrepresented low-income households
- Value of the poverty rate without weighting: 26%
- Value of the poverty rate with weighting: 12%

Estimating population descriptive statistics: an example

- Now we estimate this rate with the PSID survey
- PSID: Panel Study of Income Dynamics, started in 1967
- This survey purposefully overrepresented low-income households
- Value of the poverty rate without weighting: 26%
- Value of the poverty rate with weighting: 12%

Estimating population descriptive statistics: an example

- Now we estimate this rate with the PSID survey
- PSID: Panel Study of Income Dynamics, started in 1967
- This survey purposefully overrepresented low-income households
- Value of the poverty rate without weighting: 26%
- Value of the poverty rate with weighting: 12%

Estimating population descriptive statistics: an example

- Now we estimate this rate with the PSID survey
- PSID: Panel Study of Income Dynamics, started in 1967
- This survey purposefully overrepresented low-income households
- Value of the poverty rate without weighting: 26%
- Value of the poverty rate with weighting: 12%

LIVES
Swiss National Centre of Competence in Research

Estimating population descriptive statistics: an example

- Now we estimate this rate with the PSID survey
- PSID: Panel Study of Income Dynamics, started in 1967
- This survey purposefully overrepresented low-income households
- Value of the poverty rate without weighting: 26%
- Value of the poverty rate with weighting: 12%

Estimating population descriptive statistics: an example

- Now we estimate this rate with the PSID survey
- PSID: Panel Study of Income Dynamics, started in 1967
- This survey purposefully overrepresented low-income households
- Value of the poverty rate without weighting: 26%
- Value of the poverty rate with weighting: 12%

LIVES
Swiss National Centre of Competence in Research

Outline

The conclusion is more nuanced

Model $Y = \beta X + e$

- Weights correct for representativness of the sample
- Weights correct for heteroskedasticity
  $\Rightarrow$ Weights allows to achieve more precise estimates
- Weights can be a function of independant variables
  $\Rightarrow$ Introduce linear dependency of X in the model
- Weights can be a function of the dependant variable
  $\Rightarrow$ Introduce a correlation of X with e
- Traditional softwares assumed a straightforward random sampling plan
  $\Rightarrow$ Formulas used for estimating variance are wrong

The conclusion is more nuanced

Model $Y = \beta X + e$

- Weights correct for representativness of the sample
- Weights correct for heteroskedasticity
  $\Rightarrow$ Weights allows to achieve more precise estimates
- Weights can be a function of independant variables
  $\Rightarrow$ Introduce linear dependency of X in the model
- Weights can be a function of the dependant variable
  $\Rightarrow$ Introduce a correlation of X with e
- Traditional softwares assumed a straightforward random sampling plan
  $\Rightarrow$ Formulas used for estimating variance are wrong

The conclusion is more nuanced

Model $Y = \beta X + e$

- Weights correct for representativness of the sample
- Weights correct for heteroskedasticity
  $\Rightarrow$ Weights allows to achieve more precise estimates
- Weights can be a function of independant variables
  $\Rightarrow$ Introduce linear dependency of X in the model
- Weights can be a function of the dependant variable
  $\Rightarrow$ Introduce a correlation of X with e
- Traditional softwares assumed a straightforward random sampling plan
  $\Rightarrow$ Formulas used for estimating variance are wrong

The conclusion is more nuanced

Model $Y = \beta X + e$

- Weights correct for representativness of the sample
- Weights correct for heteroskedasticity
  $\Rightarrow$ Weights allows to achieve more precise estimates
- Weights can be a function of independant variables
  $\Rightarrow$ Introduce linear dependency of X in the model
- Weights can be a function of the dependant variable
  $\Rightarrow$ Introduce a correlation of X with e
- Traditional softwares assumed a straightforward random sampling plan
  $\Rightarrow$ Formulas used for estimating variance are wrong

The conclusion is more nuanced

Model $Y = \beta X + e$

- Weights correct for representativness of the sample
- Weights correct for heteroskedasticity
  $\Rightarrow$ Weights allows to achieve more precise estimates
- Weights can be a function of independant variables
  $\Rightarrow$ Introduce linear dependency of X in the model
- Weights can be a function of the dependant variable
  $\Rightarrow$ Introduce a correlation of X with e
- Traditional softwares assumed a straightforward random sampling plan
  $\Rightarrow$ Formulas used for estimating variance are wrong

LIVES
Swiss National Centre of Competence in Research

The conclusion is more nuanced

Model $Y = \beta X + e$

- Weights correct for representativness of the sample
- Weights correct for heteroskedasticity
  $\Rightarrow$ Weights allows to achieve more precise estimates
- Weights can be a function of independant variables
  $\Rightarrow$ Introduce linear dependency of X in the model
- Weights can be a function of the dependant variable
  $\Rightarrow$ Introduce a correlation of X with e
- Traditional softwares assumed a straightforward random sampling plan
  $\Rightarrow$ Formulas used for estimating variance are wrong

The conclusion is more nuanced

Model $Y = \beta X + e$

- Weights correct for representativness of the sample
- Weights correct for heteroskedasticity
  $\Rightarrow$ Weights allows to achieve more precise estimates
- Weights can be a function of independant variables
  $\Rightarrow$ Introduce linear dependency of X in the model
- Weights can be a function of the dependant variable
  $\Rightarrow$ Introduce a correlation of X with e
- Traditional softwares assumed a straightforward random sampling plan
  $\Rightarrow$ Formulas used for estimating variance are wrong

Hypothesis testing: partial conclusion

1. Check if the software provide methods for handling the sampling design of the survey we work on
   - 1.1 YES: use weights
   - 1.2 NO: you probably should use weights, but be careful and set out your arguments

Hypothesis testing: partial conclusion

1. Check if the software provide methods for handling the sampling design of the survey we work on
   - 1.1 YES: use weights
   - 1.2 NO: you probably shoud use weights, but be carefull and set out your arguments

Hypothesis testing: partial conclusion

1. Check if the software provide methods for handling the sampling design of the survey we work on
   - 1.1 YES: use weights
   - 1.2 NO: you probably shoud use weights, but be carefull and set out your arguments

Hypothesis testing: partial conclusion

1. Check if the software provide methods for handling the sampling design of the survey we work on
   - 1.1 YES: use weights
   - 1.2 NO: you probably shoud use weights, but be carefull and set out your arguments

Correctly taking weights into account in statistical softwares

R: Not native, need the `survey` package

SPSS: Not native, need the `complex sample` module

SAS: Not native, need the `survey` module

Correctly taking weights into account in statistical softwares

R: Not native, need the `survey` package

SPSS: Not native, need the `complex sample` module

SAS: Not native, need the `survey` module

Correctly taking weights into account in statistical softwares

R: Not native, need the `survey` package

SPSS: Not native, need the `complex sample` module

SAS: Not native, need the `survey` module

Correctly taking weights into account in statistical softwares

R: Not native, need the `survey` package

SPSS: Not native, need the `complex sample` module

SAS: Not native, need the `survey` module

Outline

Quick overview of the SHP

- Survey permanent resident households and individuals
- From 1999 to 2020 (and more?)
- Simple panel data, but will become rotative
- Main goal: to observe social change, especially in life conditions, in Switzerland
- Largest longitudinal survey in social sciences in Switzerland
- Is part of the Cross-National Equivalent File (CNEF)
- Age limit: 14 years old

LIVES
Swiss National Centre of Competence in Research

Quick overview of the SHP

- Survey permanent resident households and individuals
- From 1999 to 2020 (and more?)
- Simple panel data, but will become rotative
- Main goal: to observe social change, especially in life conditions, in Switzerland
- Largest longitudinal survey in social sciences in Switzerland
- Is part of the Cross-National Equivalent File (CNEF)
- Age limit: 14 years old

LIVES
Swiss National Centre of Competence in Research

Quick overview of the SHP

- Survey permanent resident households and individuals
- From 1999 to 2020 (and more?)
- Simple panel data, but will become rotative
- Main goal: to observe social change, especially in life conditions, in Switzerland
- Largest longitudinal survey in social sciences in Switzerland
- Is part of the Cross-National Equivalent File (CNEF)
- Age limit: 14 years old

LIVES
Swiss National Centre of Competence in Research

Quick overview of the SHP

- Survey permanent resident households and individuals
- From 1999 to 2020 (and more?)
- Simple panel data, but will become rotative
- Main goal: to observe social change, especially in life conditions, in Switzerland
- Largest longitudinal survey in social sciences in Switzerland
- Is part of the Cross-National Equivalent File (CNEF)
- Age limit: 14 years old

LIVES
Swiss National Centre of Competence in Research

Quick overview of the SHP

- Survey permanent resident households and individuals
- From 1999 to 2020 (and more?)
- Simple panel data, but will become rotative
- Main goal: to observe social change, especially in life conditions, in Switzerland
- Largest longitudinal survey in social sciences in Switzerland
- Is part of the Cross-National Equivalent File (CNEF)
- Age limit: 14 years old

LIVES
Swiss National Centre of Competence in Research

Quick overview of the SHP

- Survey permanent resident households and individuals
- From 1999 to 2020 (and more?)
- Simple panel data, but will become rotative
- Main goal: to observe social change, especially in life conditions, in Switzerland
- Largest longitudinal survey in social sciences in Switzerland
- Is part of the Cross-National Equivalent File (CNEF)
- Age limit: 14 years old

LIVES
Swiss National Centre of Competence in Research

Quick overview of the SHP

- Survey permanent resident households and individuals
- From 1999 to 2020 (and more?)
- Simple panel data, but will become rotative
- Main goal: to observe social change, especially in life conditions, in Switzerland
- Largest longitudinal survey in social sciences in Switzerland
- Is part of the Cross-National Equivalent File (CNEF)
- Age limit: 14 years old

LIVES
Swiss National Centre of Competence in Research

Quick overview of the SHP

- Survey permanent resident households and individuals
- From 1999 to 2020 (and more?)
- Simple panel data, but will become rotative
- Main goal: to observe social change, especially in life conditions, in Switzerland
- Largest longitudinal survey in social sciences in Switzerland
- Is part of the Cross-National Equivalent File (CNEF)
- Age limit: 14 years old

LIVES
Swiss National Centre of Competence in Research

Main files of the SHP database are

- The SHP1 data for the years 1999 through 2011
- The SHP2 data for the years 2004 through 2011
- The SHP biographical data for 5,560 SHP1 individuals

Main files of the SHP database are

- The SHP1 data for the years 1999 through 2011
- The SHP2 data for the years 2004 through 2011
- The SHP biographical data for 5,560 SHP1 individuals

Main files of the SHP database are

- The SHP1 data for the years 1999 through 2011
- The SHP2 data for the years 2004 through 2011
- The SHP biographical data for 5,560 SHP1 individuals

Main files of the SHP database are

- The SHP1 data for the years 1999 through 2011
- The SHP2 data for the years 2004 through 2011
- The SHP biographical data for 5,560 SHP1 individuals

Weighting variables available

```
weight2011 <- contains(c("weight", "PSM"), and = T, data = shp2011)
```

```
##                                                                                Description
## wp11t1p           PSMI-PSMII transversal individual weight inflating to size of CH-population
## wp11t1s                              PSMI-PSMII transversal individual weight keeping sample size
## wp11lp1p       PSMI longitudinal individual weight inflating to size of CH-population in 1999
## wp11lp1s                                 PSMI longitudinal individual weight keeping sample size
## wp11l1p  PSMI-PSMII longitudinal individual weight inflating to size of CH-population in 2004
## wp11l1s                            PSMI-PSMII longitudinal individual weight keeping sample size
```

Weighting variables available: meaning

wp11t1p PSMI-PSMII transversal individual weight *representative of the 2011 14yo+ CH-population* inflating to size of 14yo+ CH-population

wp11t1s PSMI-PSMII transversal individual weight *representative of the 2011 14yo+ CH-population* keeping sample size *of individuals who responded in 2011*

wp11lp1p PSMI longitudinal individual weight *representative of the 1999 14yo+ CH-population* inflating to size of 14yo+ CH-population in 1999

wp11lp1s PSMI longitudinal individual weight *representative of the 1999 CH-population* keeping sample size *of individuals from 1999 still here in 2011*

wp11l1p PSMI-PSMII longitudinal individual weight *representative of the 2004 14yo+ CH-population* inflating to size of 14yo+ CH-population in 2004

wp11l1s PSMI-PSMII longitudinal individual weight *representative of the 2004 14yo+ CH-population* keeping sample size *of individuals from 1999 or 2004 still here in 2011*

Quick look into the 1999 wave

```
shp1999$weip99ts <- wvar(shp1999$weip99ts)
shp1999w <- shp1999
weighting(shp1999w) <- "weip99ts"
nindividual(shp1999)

## [1] 12931

nindividual(shp1999w)

## [1] 7799

nindividual(shp1999w)/nindividual(shp1999) * 100

## [1] 60.31
```

Quick look into the 2011 wave – cross-sectional setting

```
shp2011$wp11t1s <- wvar(shp2011$wp11t1s)
shp2011w <- shp2011
weighting(shp2011w) <- "wp11t1s"
nindividual(shp2011)

## [1] 11178

nindividual(shp2011w)

## [1] 7459

nindividual(shp2011w)/nindividual(shp2011) * 100

## [1] 66.73
```

Quick look into the 2011 wave – longitudinal setting, from 1999

```
shp2011$wp11lp1s <- wvar(shp2011$wp11lp1s)
shp2011wl1999 <- shp2011
weighting(shp2011wl1999) <- "wp11lp1s"
nindividual(shp2011wl1999)

## [1] 3988

nindividual(shp2011wl1999)/nindividual(shp2011) * 100

## [1] 35.68
```

Quick look into the 2011 wave – longitudinal setting, from 2004

```
shp2011$wp1111s <- wvar(shp2011$wp1111s)
shp2011wl2004 <- shp2011
weighting(shp2011wl2004) <- "wp1111s"
nindividual(shp2011wl2004)

## [1] 6345

nindividual(shp2011wl2004)/nindividual(shp2011) * 100

## [1] 56.76
```

Outline

Swiss Households Panel

Computation of the weights

Who compute PSM weights?

1999 OFS
2000-2004 Statistique Canada
2005 Éric Graf (OFS) and Statistique Canada
2006 Éric Graf (OFS)
2007?-now Martina Rothenbuehler (FORS)

Who compute PSM weights?

1999 OFS

2000-2004 Statistique Canada

2005 Éric Graf (OFS) and Statistique Canada

2006 Éric Graf (OFS)

2007?-now Martina Rothenbuehler (FORS)

Who compute PSM weights?

1999 OFS

2000-2004 Statistique Canada

2005 Éric Graf (OFS) and Statistique Canada

2006 Éric Graf (OFS)

2007?-now Martina Rothenbuehler (FORS)

Who compute PSM weights?

1999	OFS
2000-2004	Statistique Canada
2005	Éric Graf (OFS) and Statistique Canada
2006	Éric Graf (OFS)
2007?-now	Martina Rothenbuehler (FORS)

Who compute PSM weights?

1999	OFS
2000-2004	Statistique Canada
2005	Éric Graf (OFS) and Statistique Canada
2006	Éric Graf (OFS)
2007?-now	Martina Rothenbuehler (FORS)

Who compute PSM weights?

1999	OFS
2000-2004	Statistique Canada
2005	Éric Graf (OFS) and Statistique Canada
2006	Éric Graf (OFS)
2007?-now	Martina Rothenbuehler (FORS)

LIVES
Swiss National Centre of Competence in Research

Quick overview of the sampling design

- PSM1 and PSM2 are stratified according to main geographical areas (NUTS II)
- The sampling is only proportional to the number of households per geographical area
- It doesn't consider the average number of people in household per geographical area
- Sampling was performed on 3 709 215 households for PSM1, 3 436 873 for PSM2

Quick overview of the sampling design

- PSM1 and PSM2 are stratified according to main geographical areas (NUTS II)
- The sampling is only proportional to the number of households per geographical area
- It doesn't consider the average number of people in household per geographical area
- Sampling was performed on 3 709 215 households for PSM1, 3 436 873 for PSM2

Quick overview of the sampling design

- PSM1 and PSM2 are stratified according to main geographical areas (NUTS II)
- The sampling is only proportional to the number of households per geographical area
- It doesn't consider the average number of people in household per geographical area
- Sampling was performed on 3 709 215 households for PSM1, 3 436 873 for PSM2

Quick overview of the sampling design

- PSM1 and PSM2 are stratified according to main geographical areas (NUTS II)
- The sampling is only proportional to the number of households per geographical area
- It doesn't consider the average number of people in household per geographical area
- Sampling was performed on 3 709 215 households for PSM1, 3 436 873 for PSM2

Quick overview of the sampling design

- PSM1 and PSM2 are stratified according to main geographical areas (NUTS II)
- The sampling is only proportional to the number of households per geographical area
- It doesn't consider the average number of people in household per geographical area
- Sampling was performed on 3 709 215 households for PSM1, 3 436 873 for PSM2

Interview

- Method: CATI, landline and mobile phones.
- Possible bias coming from the database used for sampling:
  - Duplicates
  - Under-coverage
  - Over-coverage

Interview

- Method: CATI, landline and mobile phones.
- Possible bias coming from the database used for sampling:
  - Duplicates
  - Under-coverage
  - Over-coverage

Interview

- Method: CATI, landline and mobile phones.
- Possible bias coming from the database used for sampling:
  - Duplicates
  - Under-coverage
  - Over-coverage

Interview

- Method: CATI, landline and mobile phones.
- Possible bias coming from the database used for sampling:
  - Duplicates
  - Under-coverage
  - Over-coverage

Interview

- Method: CATI, landline and mobile phones.
- Possible bias coming from the database used for sampling:
  - Duplicates
  - Under-coverage
  - Over-coverage

Interview

- Method: CATI, landline and mobile phones.
- Possible bias coming from the database used for sampling:
  - Duplicates
  - Under-coverage
  - Over-coverage

Computation of the weights

About 6 steps by weighting variable, but the two mains are:

1. Adjustment for non-response by segmentation analysis
2. Adjustment on margins of: age, sex, nationality, civil status, and 7 major geographical regions (NUTS II)

Computation of the weights

About 6 steps by weighting variable, but the two mains are:

1. Adjustment for non-response by segmentation analysis
2. Adjustment on margins of: age, sex, nationality, civil status, and 7 major geographical regions (NUTS II)

Computation of the weights

About 6 steps by weighting variable, but the two mains are:

1. Adjustment for non-response by segmentation analysis
2. Adjustment on margins of: age, sex, nationality, civil status, and 7 major geographical regions (NUTS II)

Fist step: segmentation analysis

1. If non-response is random within a group, then non-response biais is insignificant
2. So we look for homogeneous groups which efficiently predict non-response
3. For this purpose, segmentation by CHAID method is performed
4. Variables used: from registers, from the interview mode, from previous waves
5. Adjustment applied for a respondent is the inverse rate of non-response of its group

Fist step: segmentation analysis

1. If non-response is random within a group, then non-response biais is insignificant
2. So we look for homogeneous groups which efficiently predict non-response
3. For this purpose, segmentation by CHAID method is performed
4. Variables used: from registers, from the interview mode, from previous waves
5. Adjustment applied for a respondent is the inverse rate of non-response of its group

Fist step: segmentation analysis

1. If non-response is random within a group, then non-response biais is insignificant
2. So we look for homogeneous groups which efficiently predict non-response
3. For this purpose, segmentation by CHAID method is performed
4. Variables used: from registers, from the interview mode, from previous waves
5. Adjustment applied for a respondent is the inverse rate of non-response of its group

Fist step: segmentation analysis

1. If non-response is random within a group, then non-response biais is insignificant
2. So we look for homogeneous groups which efficiently predict non-response
3. For this purpose, segmentation by CHAID method is performed
4. Variables used: from registers, from the interview mode, from previous waves
5. Adjustment applied for a respondent is the inverse rate of non-response of its group

Fist step: segmentation analysis

1. If non-response is random within a group, then non-response biais is insignificant
2. So we look for homogeneous groups which efficiently predict non-response
3. For this purpose, segmentation by CHAID method is performed
4. Variables used: from registers, from the interview mode, from previous waves
5. Adjustment applied for a respondent is the inverse rate of non-response of its group

Fist step: segmentation analysis

1. If non-response is random within a group, then non-response biais is insignificant
2. So we look for homogeneous groups which efficiently predict non-response
3. For this purpose, segmentation by CHAID method is performed
4. Variables used: from registers, from the interview mode, from previous waves
5. Adjustment applied for a respondent is the inverse rate of non-response of its group

<table>
<tr><th>split1</th><th>split2</th><th>split3</th><th>split4</th><th>split5</th><th>split6</th><th>rhg</th></tr>
<tr><td rowspan="29">agevo_1=0
(8387)
45.9%</td><td rowspan="29">FORM_SUP=
0
(5652)
42.8%</td><td rowspan="2">SWISS=0
(762) 33.0%</td><td>APPART=0 (157) 39.6%</td><td></td><td></td><td>1</td></tr>
<tr><td>APPART=1 (605) 31.4%</td><td></td><td></td><td>2</td></tr>
<tr><td rowspan="27">SWISS=1
(4890)
44.5%</td><td rowspan="3">MARIE=0
(1491) 38.1%</td><td rowspan="3">REV_1=0
(1050)
41.3%</td><td rowspan="2">agevo_2=0
(665) 46.0%</td><td>3
4
5</td></tr>
<tr><td>agevo_2=1
(385) 32.8%</td><td>6
7</td></tr>
<tr><td>REV_1=1 (441) 30.7%</td><td></td><td>8</td></tr>
<tr><td rowspan="25">MARIE=1
(3399) 47.3%</td><td rowspan="13">NO_KID=0
(1534)
51.5%</td><td rowspan="12">LOY_1=0
(1281) 49.6%</td><td>9
10
11
12
13
14
15</td></tr>
<tr><td>LOY_1=1 (253) 61.1%</td><td>16</td></tr>
<tr><td rowspan="12">NO_KID=1
(1865)
44.1%</td><td rowspan="7">agevo_5=0
(770) 39.5%</td><td>17
18
19
20
21
22</td></tr>
<tr><td rowspan="7">agevo_5=1
(1095) 47.1%</td><td>23
24
25
26
27
28
29</td></tr>
</table>

Figure: *Segmentation for longitudinal weights, individuals, wave 2006 (extract), (Graf 2008)*

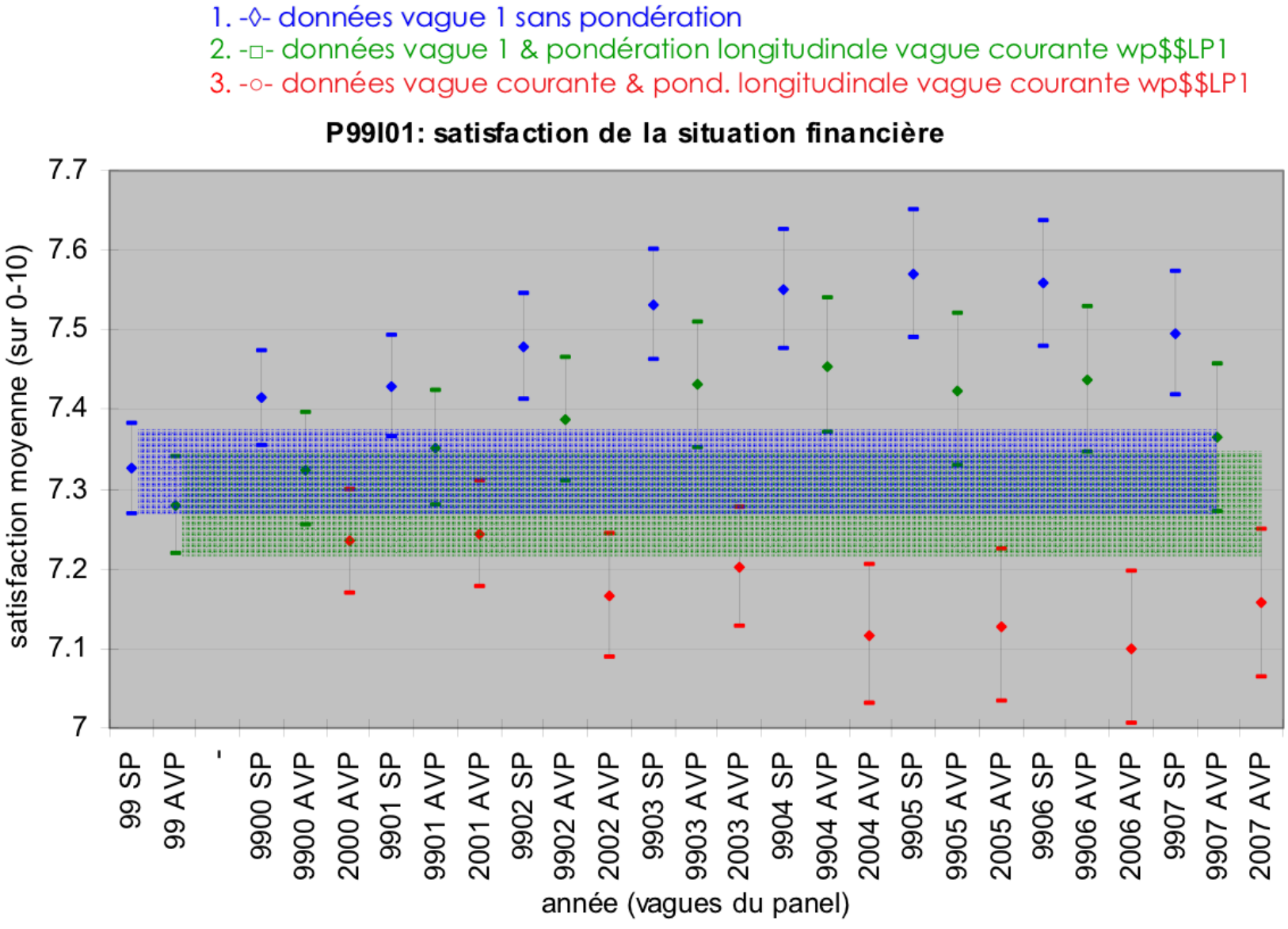

Figure: *Diagnostic method used for longitudinal weights, (Graf 2010)*

Outline

Outline

Key points

- Descriptive statistics: weighting is mandatory
- Estimating effects: a diagnostic step must be performed
- But in most cases weights have to be used
- Check for the right method in the statistical software

Key points

- Descriptive statistics: weighting is mandatory
- Estimating effects: a diagnostic step must be performed
- But in most cases weights have to be used
- Check for the right method in the statistical software

Key points

- Descriptive statistics: weighting is mandatory
- Estimating effects: a diagnostic step must be performed
- But in most cases weights have to be used
- Check for the right method in the statistical software

Key points

- Descriptive statistics: weighting is mandatory
- Estimating effects: a diagnostic step must be performed
- But in most cases weights have to be used
- Check for the right method in the statistical software

Key points

- Descriptive statistics: weighting is mandatory
- Estimating effects: a diagnostic step must be performed
- But in most cases weights have to be used
- Check for the right method in the statistical software

Outline

On the theoretical point of view

- Better understand side effects of weighting
- Summarizing diagnostic tests to perform before running into analysis
- Providing a practical guide for helping user to handle weights
- May look for someone with a statistical background for writing this working paper

In another time, try to summarize litterature on partial non-response handling

On the theoretical point of view

- Better understand side effects of weighting
- Summarizing diagnostic tests to perform before running into analysis
- Providing a practical guide for helping user to handle weights
- May look for someone with a statistical background for writing this working paper

In another time, try to summarize litterature on partial non-response handling

On the theoretical point of view

- Better understand side effects of weighting
- Summarizing diagnostic tests to perform before running into analysis
- Providing a practical guide for helping user to handle weights
- May look for someone with a statistical background for writing this working paper

In another time, try to summarize litterature on partial non-response handling

On the theoretical point of view

- Better understand side effects of weighting
- Summarizing diagnostic tests to perform before running into analysis
- Providing a practical guide for helping user to handle weights
- May look for someone with a statistical background for writing this working paper

In another time, try to summarize litterature on partial non-response handling

On the theoretical point of view

- Better understand side effects of weighting
- Summarizing diagnostic tests to perform before running into analysis
- Providing a practical guide for helping user to handle weights
- May look for someone with a statistical background for writing this working paper

In another time, try to summarize litterature on partial non-response handling

On the theoretical point of view

- Better understand side effects of weighting
- Summarizing diagnostic tests to perform before running into analysis
- Providing a practical guide for helping user to handle weights
- May look for someone with a statistical background for writing this working paper

In another time, try to summarize litterature on partial non-response handling

On the theoretical point of view

- Better understand side effects of weighting
- Summarizing diagnostic tests to perform before running into analysis
- Providing a practical guide for helping user to handle weights
- May look for someone with a statistical background for writing this working paper

In another time, try to summarize litterature on partial non-response handling

For the *Dataset* R package

- Run statistical analysis through methods provided by the survey package instead of native R methods
- Allow to better describe the survey design
- For objects of class "WeightingVariable", allow to specify
  - Variables involved in the computation of the weights
  - Individuals concerned (Myot, must be one wave, ...)
  - Make the sample representative to which population (year 1999)
- Better alert/inform the user of misusage of weighting

For the *Dataset* R package

- Run statistical analysis through methods provided by the `survey` package instead of native R methods
- Allow to better describe the survey design
- For objects of class "WeightingVariable", allow to specify
  - Variables involved in the computation of the weights
  - Individuals concerned (14yo+, miss no one wave, ...)
  - Make the sample representative to which population (year 1999)
- Better alert/inform the user of misusage of weighting

For the *Dataset* R package

- Run statistical analysis through methods provided by the `survey` package instead of native R methods
- Allow to better describe the survey design
- For objects of class "WeightingVariable", allow to specify
  - Variables involved in the computation of the weights
  - Individuals concerned (14yo+, miss no one wave, ...)
  - Make the sample representative to which population (year 1999)
- Better alert/inform the user of misusage of weighting

For the *Dataset* R package

- Run statistical analysis through methods provided by the `survey` package instead of native R methods
- Allow to better describe the survey design
- For objects of class "WeightingVariable", allow to specify
  - Variables involved in the computation of the weights
  - Individuals concerned (14yo+, miss no one wave, ...)
  - Make the sample representative to which population (year 1999)
- Better alert/inform the user of misusage of weighting

For the *Dataset* R package

- Run statistical analysis through methods provided by the `survey` package instead of native R methods
- Allow to better describe the survey design
- For objects of class "WeightingVariable", allow to specify
  - Variables involved in the computation of the weights
  - Individuals concerned (14yo+, miss no one wave, . . .)
  - Make the sample representative to which population (year 1999)
- Better alert/inform the user of misusage of weighting

For the *Dataset* R package

- Run statistical analysis through methods provided by the `survey` package instead of native R methods
- Allow to better describe the survey design
- For objects of class "WeightingVariable", allow to specify
  - Variables involved in the computation of the weights
  - Individuals concerned (14yo+, miss no one wave, . . .)
  - Make the sample representative to which population (year 1999)
- Better alert/inform the user of misusage of weighting

For the *Dataset* R package

- Run statistical analysis through methods provided by the `survey` package instead of native R methods
- Allow to better describe the survey design
- For objects of class "WeightingVariable", allow to specify
  - Variables involved in the computation of the weights
  - Individuals concerned (14yo+, miss no one wave, ...)
  - Make the sample representative to which population (year 1999)
- Better alert/inform the user of misusage of weighting

For the *Dataset* R package

- Run statistical analysis through methods provided by the `survey` package instead of native R methods
- Allow to better describe the survey design
- For objects of class "WeightingVariable", allow to specify
  - Variables involved in the computation of the weights
  - Individuals concerned (14yo+, miss no one wave, ...)
  - Make the sample representative to which population (year 1999)
- Better alert/inform the user of misusage of weighting

Bibliography

Bibliography I

[Graf 2008] Éric Graf. Pondérations du PSM *Office Fédéral de la Statistique*. Rapport de méthodes, 2008.

[Graf 2010] Éric Graf. Étude empirique de l'attrition du Panel suisse de ménages, Vers une caractérisation du profil du non-répondant *Office Fédéral de la Statistique*. Rapport de méthodes, 2010.

[Solon et al.] Solon, G., Haider, S.J., and Wooldridge, J.M. What Are We Weighting For?. *Preliminary Draft, not for citation*

[Voorpostel et al.] Voorpostel, M., Tillmann, R, Lebert, F., Weaver, B., Kuhn, U., Lipps, O., Ryser, V.-A., Schmid, F., Rothenbühler, M., and Wernli, B. *Swiss Household Panel Userguide (1999-2010), Wave 12*. Lausanne: FORS.(October 2011).

[Winship and Radbill] Christopher Winship and Larry Radbill. Sampling Weights and Regression Analysis. Sociological Methods & Research, Vol. 23(2), November 1994.

LIVES Swiss National Centre of Competence in Research

Thank you for your attention

Any question?