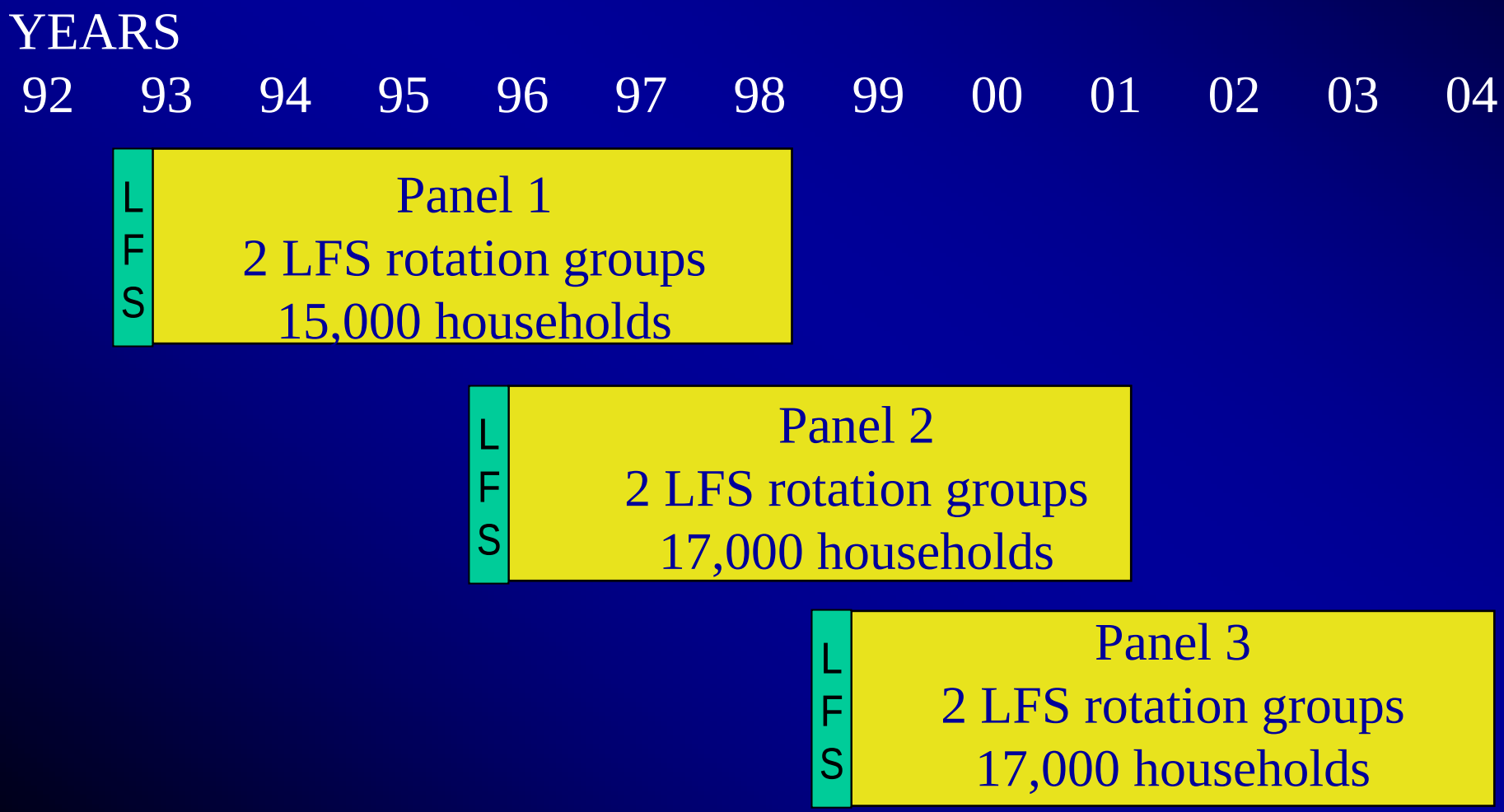
The Survey of Labour and Income Dynamics

- → Methodology
- → Weighting strategy
  - longitudinal
  - cross-sectional
- → Overview of the Imputation Strategy
- → Conclusion

SLID Design and Weighting Strategy



Methodology

→ 2 CATI collection periods per year

- labour interview in January
  - » personal characteristics
    - – origin, education, demographics, etc.
  - » household composition
  - » monthly labour activity

- income interview in May
  - » previous calendar year
  - » survey or tax route, 25-75 split
    - – some differences between the two sources

Methodology

→ Longitudinal person:

- Person in a sampled household at the time of the selection of the panel

→ Cohabitant:

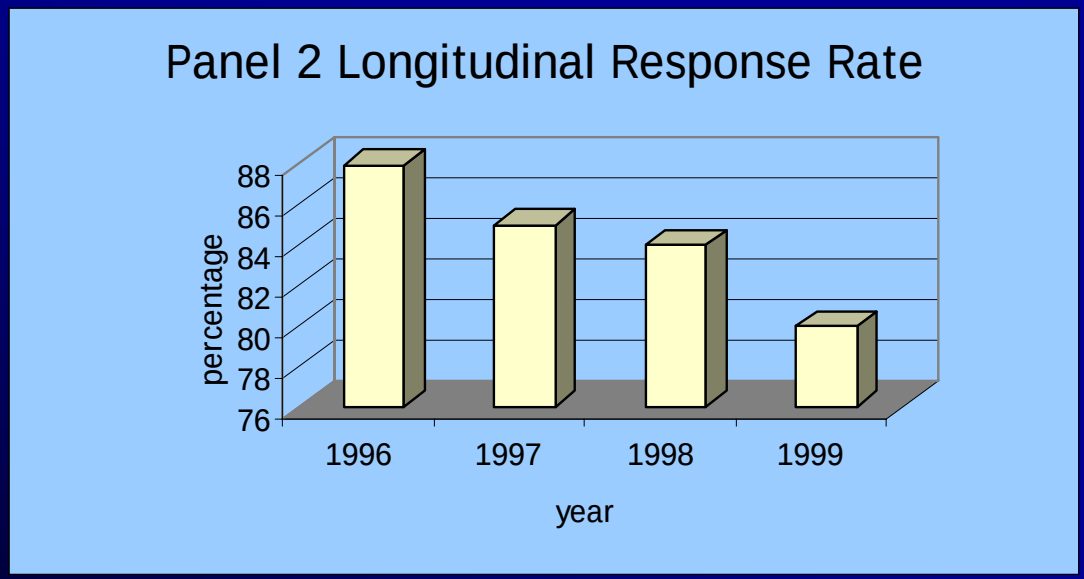
- Person living with a longitudinal individual on December 31 of the reference year

→ Definition of a respondent

- consider only the first and the year being processed
- household level
- a household is respondent if and only if one can determine its composition for the current reference year
- implies that at least one person participated in the January interview

Methodology

- Definition of a respondent
  - nonresponding persons within a responding household are imputed
    - » cross-sectional and historical methods of imputation are used
    - » so far, only the current year is imputed



Weighting strategy

Current weighting strategy

Longitudinal
Subweight
Nonresponse adjustment
Analytical adjustment:
outliers
Post-stratification

Weighting strategy

Current weighting strategy

Longitudinal	Cross-sectional
Subweight	Subweight
Nonresponse adjustment	Nonresponse adjustment
	Panel combination
	Addition of cohabitants (weight share)
Analytical adjustment:	Analytical adjustments:
outliers	inter-provincial migration and outliers
Post-stratification	Post-stratification

Both weights are computed each year

Longitudinal Weighting strategy

- Longitudinal weights are intended to represent the population of Canada's ten provinces
  - on December 31, 1992, for Panel 1
  - on December 31, 1995, for Panel 2
  - on December 31, 1998, for Panel 3
- Longitudinal weight is done independently for each panel

Longitudinal Weighting strategy

→ Nonresponse adjustment

- performed at the person level
- based on observed response rate within homogeneous response groups
- modelling is done every year
- use of segmentation analysis rather than logistic regression
  - » built a tree where the built nodes are studied independently
  - » studies suggest that segmentation analysis is better
- origin of the explanatory variables:

Longitudinal Weighting strategy

Main Variables used in the Making of the HRG

Panel 1	Panel 2	Panel 3
Ethnicity		
Country of origin		
Education level	Education level	Education level
Marital status	Marital status	Marital status
Size of area of residence		Size of area of residence
Labour force status	Labour force status	
	Class of worker (self empl., public, etc.)	Class of worker
Household income		Household income
Owner	Owner	Owner
	Household size	Household size
		Economic family size
	Family type (lone parent, couple, etc)	Family type
Province	Province	Province
Age group	Age group	Age group
Sex	Sex	Sex

Longitudinal Weighting strategy

→ Adjustment for Influential Values

- Same as cross-sectional weighting

→ Post-stratification

- The control totals are based on population projections generated by the Demography Division.
- provinceXageXsex groups (220)

Cross-sectional Weighting Strategy

→ Create a set of weights representative of Canada's ten provinces on December 31 of a given reference year.

- All cross-sectionally in-scope longitudinal individuals and their cohabitants are included in the production of this set of weights.

→ 2 types of weights

- Individual weight : Each member of a household may have a different weight (pp only)
- Integrated weight : Every member of that household has the same weight. (HH and PP)

Cross-sectional Weighting Strategy

→ Application of Panel Allocation Factors

- In theory
  => The two panels could be used independently of one another to produce cross-sectional estimates
- In practice
  => Better to combine the two cross-sectional samples to produce the cross-sectional estimates
  - » Doubles sample size
  - » Reduces variability of the estimates
  - » Permits the use of more control totals in the post-stratification

Cross-sectional Weighting Strategy

→ How to combine the two panels:

- Use of a combined estimator which assigns a relative importance to each panel based on a certain factor p
- Factor p :
  - » depends on the sample sizes of each panel
  - » on the design effect
  - » calculated to minimize the variance of a point estimate
  - » This factor is not applied to individuals in panel 2 that were not in the target population at the time when panel 1 was selected

Cross-sectional Weighting Strategy

→ Weight Share

- The weight share step is essential and unique to cross-sectional weighting.
- It is necessitated by the longitudinal nature of SLID. In a longitudinal survey, the sample evolves over time.
- Cohabitants received an average weight computed at the household level

$$\overline{w} = \frac{\sum_{long.\in level}^{L} w_l}{\#long.+\#IP}$$

Cross-sectional Weighting Strategy

→ Adjustment for Interprovincial Migration

- Required because, some people in the sample move from one province to another.
- Probability of selection vary widely from one region to another
- Substantial difference at the national level between the largest weights and the smallest ones.
- The method adopted by SLID uses the 95th percentile of the distribution of weights

Cross-sectional Weighting Strategy

→ Adjustment for Influential Values

- The first step is to identify the influential items, that is, observations that contribute too heavily to the weighted estimate of total personal income for the province.
- Based on comparisons with external sources
- Adjustment factors to the weights of influential observations

Cross-sectional Weighting Strategy

→ Post-stratification

- Same process as for the longitudinal weighting process
- Adjustment based on the control totals of the Canadian population at December 31st of the reference year
- ProvinceXageXsex groups (220)

SLID Weighting Strategy

→ Future

- Combined panel longitudinal weights
- New calibration scheme that will incorporate family size counts, and household size counts
- Development of a between wave imputation strategy

Overview of the SLID Imputation Strategy

Step 1

→ Income

Step 2

→ Demographic

→ Education Attainment

→ Labour Market Activities

→ Job Characteristics

→ Educational Activity

Overview of the SLID Imputation Strategy

→ Certain amount of knowledge of respondent is known:

- Age, sex, household size, province of residence

→ Donors must meet certain criteria

- (For example the top 0.2% of all income groups are excluded from being income donors)

Overview of the SLID Imputation Strategy

→ Pre-edit and coherence check
→ Historical imputation
→ Cross-sectional imputation
→ Deterministic imputation
→ Modeling
→ Validity check

- Imputation can be partial or complete, and may involve any combination of the 4 types of imputation.

Overview of the SLID Imputation Strategy

→ Income imputation

- Imputation method
  - » 14 variables are done historically then Nearest Neighbor
  - » Federal and Provincial Tax modeled
  - » Government assistance variables imputed deterministically

Overview of the SLID Imputation Strategy

→ Nearest neighbor Imputation

- Matched on 20 qualitative variables, 18 quantitative variables, best qualitative match of 25 best quantitative matches selected
- 16 categories (2 panels x 2 sex x 4 age)
  - » If no donor is found in first category then try other panel, still if no donor try all other categories without crossing age 70 barrier.

Overview of the SLID Imputation Strategy

→ Other sections

- Nearest Neighbor Imputation
  - » 10 categories used (5 Regions x 2 Sex)
  - » between 12 and 24 matching variables used depending on the section
  - » Each section done individually (with the exception of Labour Market Activities and Job Characteristics, which are done at the same time)

Conclusion

→ SLID base weights are not simple

- They are further modified on several occasions
- Several weights are available
- One should make sure to use the good one

- Some missing data are on the file and the currents weights do not account for them
  - » CAUSION

- The WHOLE survey process can affect your analyses

Please give me some feedback

S.V.P. donnez-moi de la rétro-action