

EU-SILC workshop

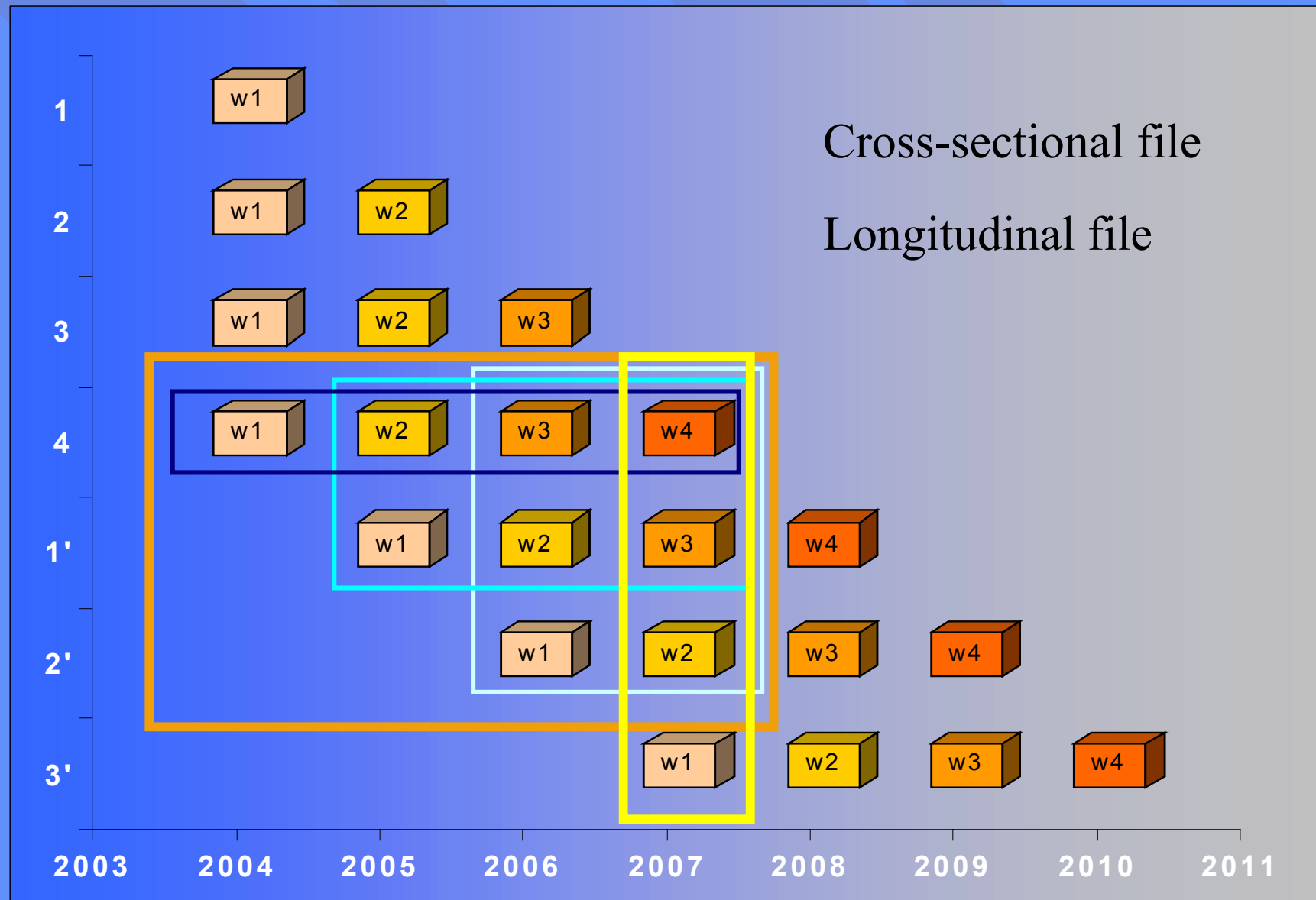
Weighting for EU-SILC

JM. Museux (Eurostat)

EU-SILC – Weighting issues

- Objective
 - A unified and practical structure for the weighting of EU-SILC 4 year rotational design
 - Cross sectional estimation
 - Longitudinal estimation
- Work done by Vijay Verma (University of Sienna) – contract Advanced Estimation Methods with Eurostat
- First Eurostat paper on this issue by P Seoane (Eurostat – INE Spain) and G. Osier (Eurostat)
- Seminal theoretical work by P. Ardilly and P. Lavallée (Symposium Stat Canada 2003)

Eurostat 4 years rotational design



EU-SILC various aspects

- Rotational design is made 4 independent sub samples (4 year panels) each representative of the full population at the moment of selection
- Cross sectional
 - Cross sectional files: X (annual release)
 - Cumulates information from 4 panels
 - Inference on cross sectional population at time t
- Longitudinal
 - Longitudinal files: L (annual release with all existing trajectories ending at time t)
 - Cumulates information from 3 panels
 - Inference on $t-1; t$ trajectories (3 panels); on $t-2, t$ trajectories (2 panels); on $t-3, t$ trajectories (1 panel)

EU-SILC design weight variables

- Theoretical and data checking interest
- For X and L files, first wave unit only
- DB080 household design weights
 - For all households (respondents and non respondents) in the D file (household register)
- PB070: personal design weight for selected respondent (register countries only)
 - For all selected respondents in P file (individual file containing selected household members 16+) eligible for personal interview

EU-SILC cross-sectional weight variables

- For estimation purpose
- In X and L files all waves
 - DB090 household cross sectional weight
 - For all respondent households in D file
 - Cross sectional target population of resident private households
 - RB050 personal cross sectional weight
 - For all individuals in R file (individual register containing all members / former household members of respondent households)
 - Cross sectional target population of all individual living in private households

EU-SILC cross-sectional weight variables

- PB040 Personal cross sectional weight
 - For all individuals in P file (personal files with info from register and interview) containing all members 16+ of respondent households
 - For estimation purpose (cross sectional target population of all individuals 16+ living in private households)
- PB060: personal cross sectional weight for selected respondent (register countries only)
 - For all interviewed selected respondents in P file
 - Cross sectional target population of all individuals 16+ living in private households for personal info collected through interview with the selected respondent only

EU-SILC longitudinal weight variables

- To be used as a basis for longitudinal estimations
- In L files; all waves

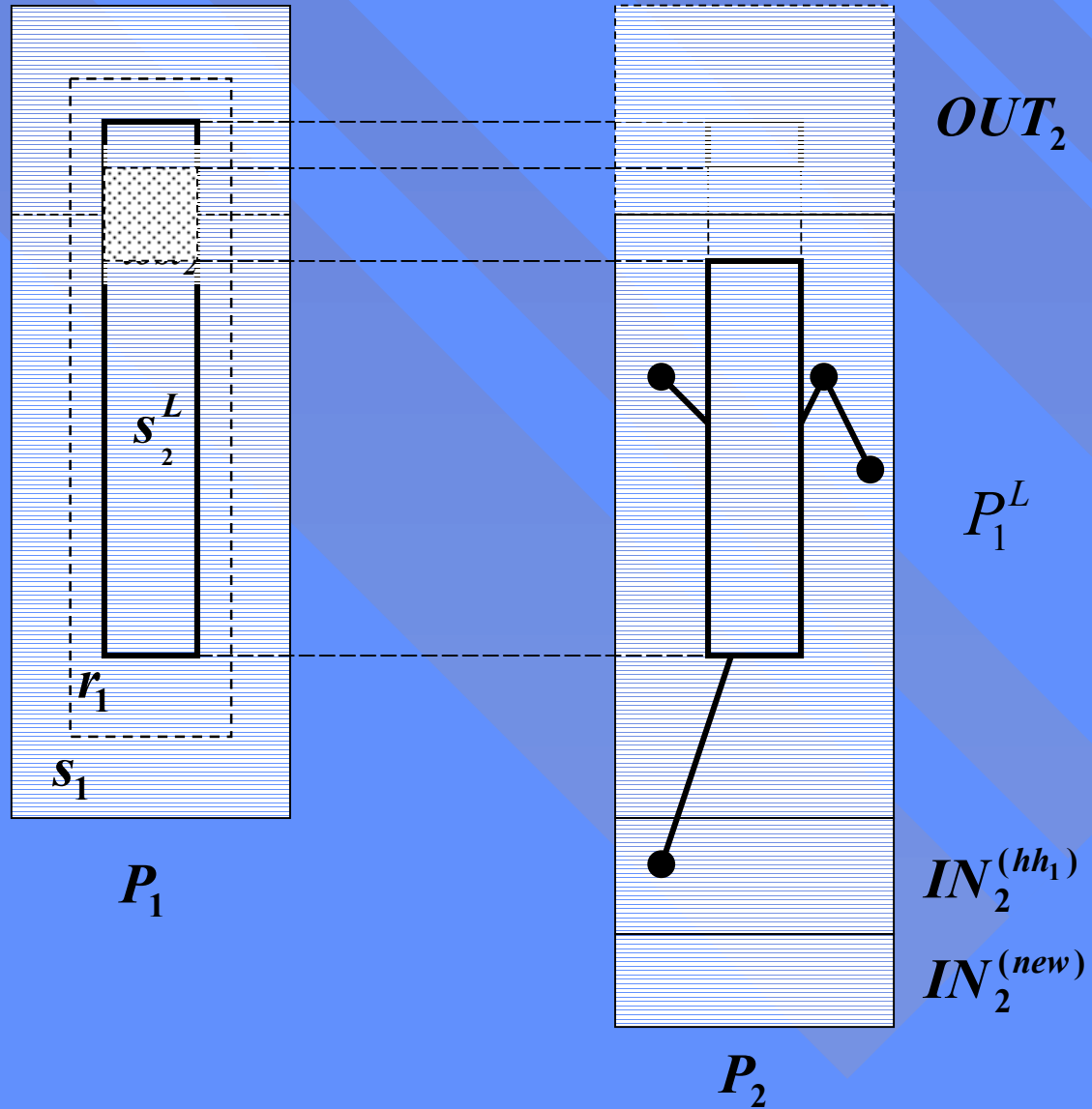
- RB060 personal base weight
 - For all individuals in R files in scope from wave_1 to wave_t
 - estimations for longitudinal population (all individuals)

- PB050 Personal base weight
 - For all individuals in P file in scope from wave_1 to wave_t
 - estimations for longitudinal 16+ population

EU-SILC longitudinal weight variables

- PB080: personal base weight for selected respondent (register countries only)
 - For all selected respondents in P file in scope from wave 1 to wave t
 - estimations for longitudinal 16+ population – personal variables

Population/sample dynamic



Algorithm step by step

(for each panel independently)

- Cross sectional weight (first wave)
 - Design weight
 - Inverse of selection probability
 - Adjustment for initial non response
 - Adjusted cell or logistic regression on household characteristics
 - Non contact / refusal
 - Trimming
 - Avoid spreading of weight
 - Trade off variance/bias
 - Calibration
 - Integrative household and individual characteristics
 - $\omega_1^{(RB)}$ = base weight for wave 1

Algorithm step by step

(for each panel independently)

- Base weight (panel persons)

- $\omega_j^{(RB)}$ = Base weight wave j

- Adjustment for non response (wave j - j+1)
 - Need to impute status of non respondent
- Trimming

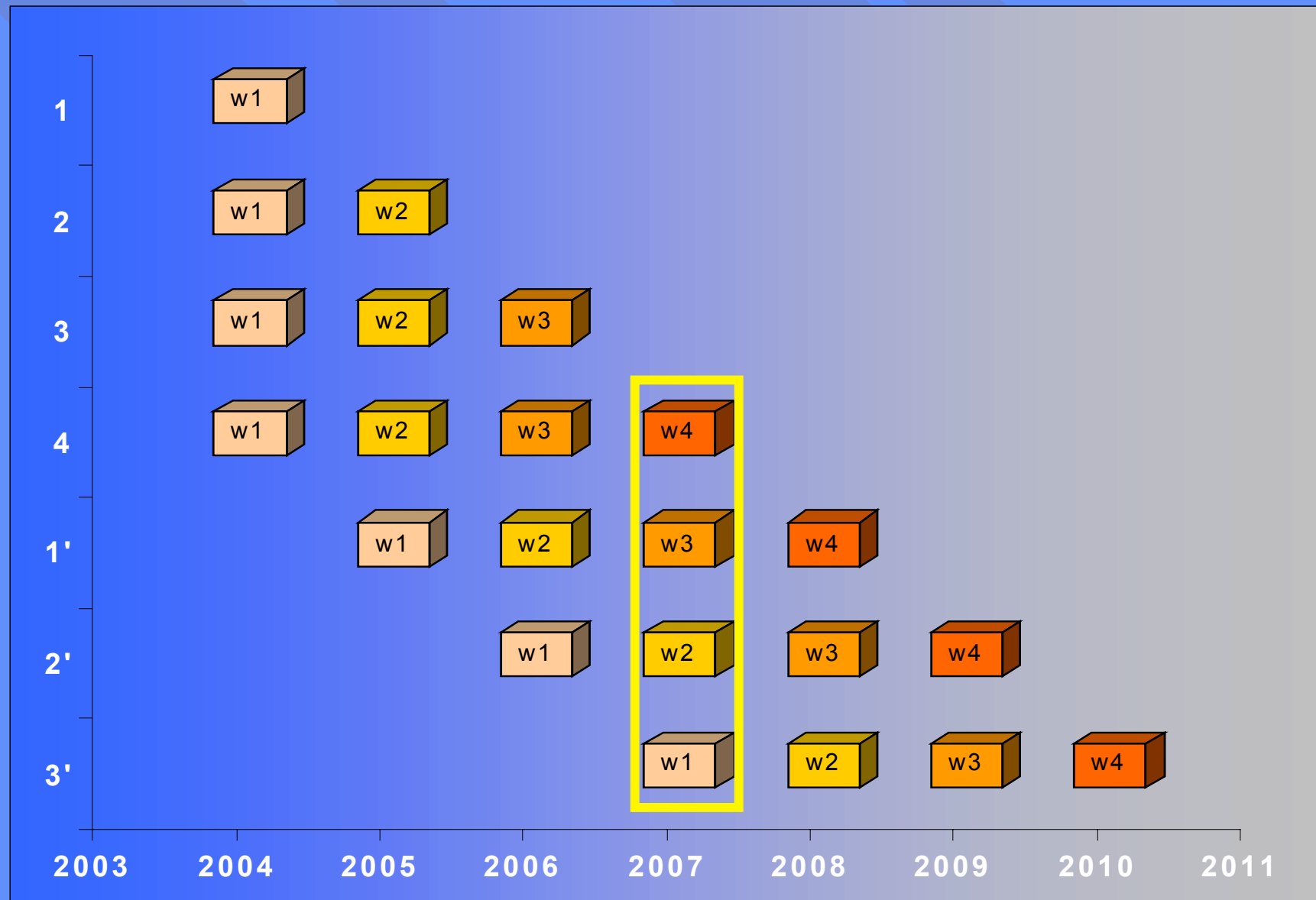
- $\omega_{j+1}^{(RB)}$ = Base weight wave j

Algorithm step by step

(for each panel independently)

- Cross sectional weight (wave $j > 1$)
 - $\omega_j^{(RB)}$ Base weight wave j sample persons
 - $\omega_j^{(RB)}$ Correction for re-entries (wave $j > 2$)
 - Extension of weights to other persons
 - New born
 - Co-resident: Weight sharing method
 - Calibration (possibly)
 - $\omega_j^{(RC)}$

Cross sectional weight variables at time t



Cross sectional weight variables at time t

4 panels will contribute with different waves

| | Current wave | Weight | Population |
|-------------------------|--------------|-------------------|---|
| Panel selected at t | wave 1 | $\omega_1^{(RC)}$ | P_t |
| Panel selected at $t-1$ | wave 2 | $\omega_2^{(RC)}$ | $P_t - IN_{t-1}^{(new)}$ |
| Panel selected at $t-2$ | wave 3 | $\omega_3^{(RC)}$ | $P_t - IN_{t-1}^{(n)} - IN_{t-2}^{(n)}$ |
| Panel selected at $t-3$ | wave 4 | $\omega_4^{(RC)}$ | $P_t - IN_{t-1}^{(n)} - IN_{t-2}^{(n)} - IN_{t-3}^{(3n)}$ |

Cross sectional weight variables at time t

- RB050 personal cross sectional weight

- For $j \in IN_t^{(new)}$ **RB050** = $\omega_1^{(RC)}$

- For $j \in IN_{t-1}^{(new)}$ **RB050** = $\omega_2^{(RC)} / 2$

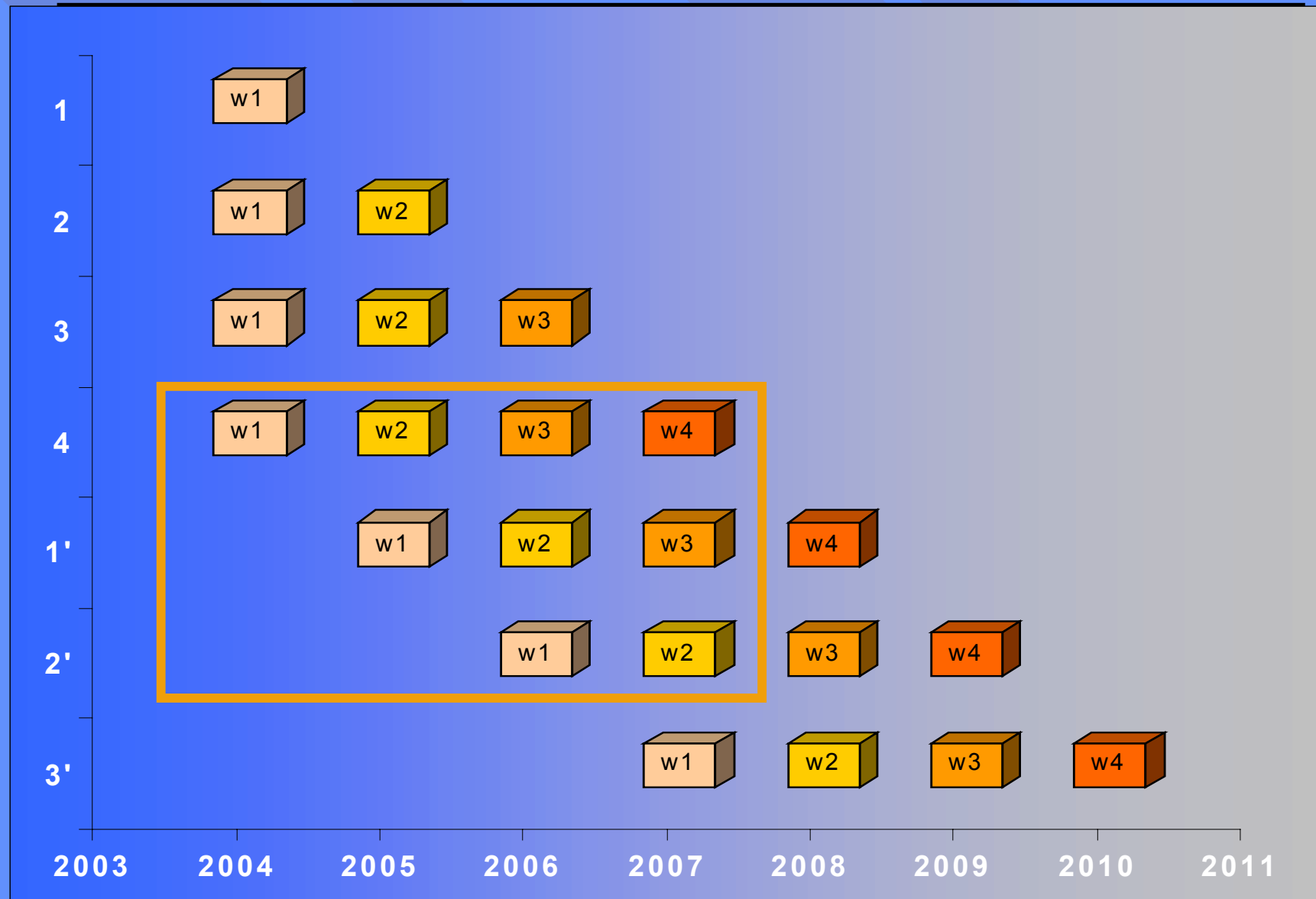
- For $j \in IN_{t-2}^{(new)}$ **RB050** = $\omega_3^{(RC)} / 3$

- *Else* **RB050** = $\omega_4^{(RC)} / 4$

Cross sectional weight variables at time t

- DB090 are by construction equal to RB050 (weight sharing)
- PB040 personal cross sectional weight
 - Based on RB050
 - Possibly modified for individual non response by post calibration on R variables
- PB060 personal cross sectional weight selected respondent
 - Are scaled on RB050 for convenience and consistency

Eurostat 4 years rotational design

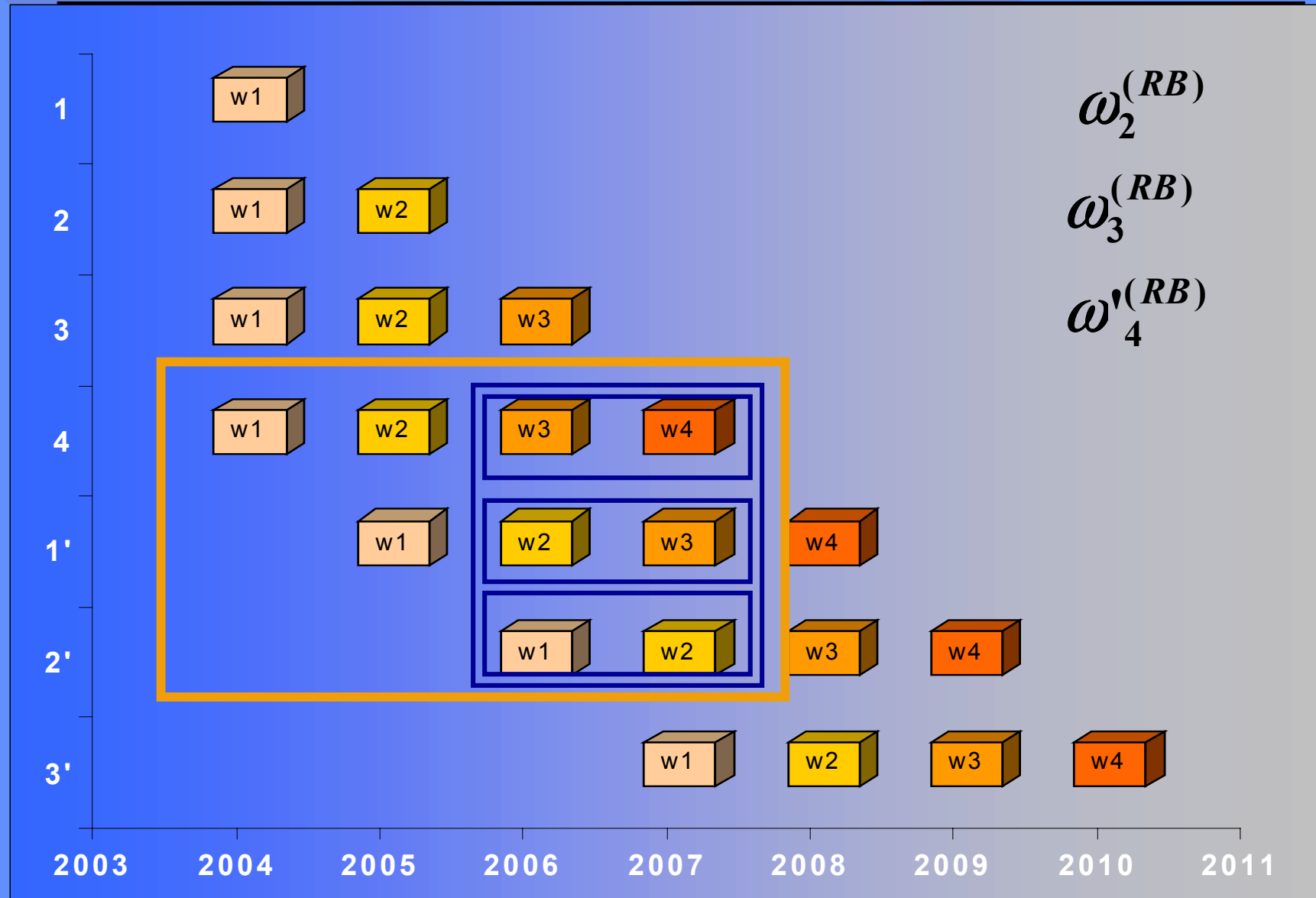


Target longitudinal weight variables in the file release at time t

- For the sake of flexibility, RB060 collect base weight as derived independently for each panel

| RB060 | w1 | w2 | w3 | w4 |
|-----------------------|-------------------|-------------------|-------------------|-------------------|
| Panel selected at t-1 | $\omega_1^{(RB)}$ | $\omega_2^{(RB)}$ | | |
| Panel selected at t-2 | $\omega_1^{(RB)}$ | $\omega_2^{(RB)}$ | $\omega_3^{(RB)}$ | |
| Panel selected at t-3 | $\omega_1^{(RB)}$ | $\omega_2^{(RB)}$ | $\omega_3^{(RB)}$ | $\omega_4^{(RB)}$ |

Longitudinal weight t-1 -> t



Longitudinal weight (wave j -> wave t)

- $\omega_t^{(RB)}$ Base weight at wave t
- $\omega_t'^{(RB)}$ Correction for re-entries (wave j > 2)
- Extension of weights to other persons to be considered in longitudinal analysis
 - New born at wave j
 - Longitudinal Co-residents wave j – wave t :
Weight sharing method

Longitudinal weight for $t-1 \rightarrow t$

3 panels will contribute with different waves

| | Current waves | weight | Longitudinal population |
|-----------------------|---------------|--------------------|---|
| Panel selected at t-1 | 1-2 | $\omega_2^{(RB)}$ | P_{t-1}^L |
| Panel selected at t-2 | 2-3 | $\omega_3^{(RB)}$ | $(P_{t-1} - IN_{t-1}^{(new)})^L$ |
| Panel selected at t-3 | 3-4 | $\omega_4'^{(RB)}$ | $(P_{t-1} - IN_{t-1}^{(new)} - IN_{t-2}^{(new)})^L$ |

Longitudinal weight for t-1 -> t study

- New weight variable to be created

- **For** $j \in IN_{t-1}^{(new)}$ $\omega_2^{(RB)}$
- **For** $j \in IN_{t-2}^{(new)}$ $\omega_{.}^{(RC)} / 2$
- **Else** $= \omega_{.}^{(RC)} / 3$

Longitudinal weight for t-2 -> t study

- New weight variable to be created

- **For** $j \in IN_{t-2}^{(new)}$ $\omega_3^{(RB)}$

- **Else** $= \omega_{.}^{(RC)} / 2$

Conclusions

- 4 year rotational design allows for variety of analysis
- The weighting system is complex but algorithm can be programmed. New variables need to be defined
- The possibility of cumulating sample over time needs to be investigated