Main issues of the weighting procedure in the French LFS

9th Workshop on Labour Force Survey Methodology

Rome, 15th & 16th of May 2014





Fabien Guggemos
France, INSEE
Employment and Earned Income Department

Presentation based also on the work of Vincent Loonis & Marianne Juillard (INSEE)

INSEE
INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

Some important dates for the French LFS Methology

From 2003 : Quarterly survey with a 6- rotation pattern, reference weeks uniformly distributed over the whole year

(until 2002, annual survey in March with 1/3 of the sample renewed each year)

2007 : Introduction of the Non-Respondent Survey (NRS)

(2007 : postal survey, 2009 : postal or internet survey)

2009 : Progressive increase of the sample size (50%) & progressive introduction of a new sampling frame (Tax registers instead of the pop. census)

(Q2 2010 : Increase of the sample size completed

Q4 2011 : sample drawn only within tax registers)

2013 : redesign of the IT tools for the LFS, Questionnaire entirely rewritten

Outline

I. Short general overview on the main charasteristics of the French LFS

II. Main issues of the weighting procedure

Methods used for the French LFS (how, why ?),
Lessons learned, Questions arising

Initial sampling & Annual updates of the sampling frame

Production of the quarterly corrected weights and integration of the NRS results

Revisions

Part I

I.

Short general overview on the main charasteristics of the French LFS

Part I : Short general overview on the main charasteristics of the French LFS

- Survey of ordinary dwellings (≠ households)
- Every individual interviewed in their main residence (only main individual characteristics asked for children below 15)
- Rotation pattern : 6 – (6 waves)

  Each dwelling surveyed 6 consecutive quarters

  → Q-to-Q overlap = 83 %

  → Y-to-Y quarterly overlap = 33 %
- Collection Mode : Face to Face in 1st & 6th (last) interviews

  Phone, otherwise

Part I : Short general overview on the main charasteristics of the French LFS

Wave approach :

- For some structural variables, questions asked only in the first wave (e.g. : situation one year before survey) or only in first and last waves (e.g. : salaries)

→ Annual LFS weights (COEFFY) based on the first wave only (vs quarterly ones COEFFQ, based on the whole sample)

→ Three different weights for national needs : COEFFQ, COEFFY + weights based on 1st-6th waves

- Ad-hoc modules : over the whole year (not only Q2), on a subsample of the 6th wave

Part I : Short general overview on the main charasteristics of the French LFS

Some illustrative figures, for one quarter :

- ~ 80,000 ordinary dwellings
- ~ 67,000 main residences
- ~ 57,000 respondent households
- A bit more than 100,000 people aged 15 or more
- Response Rate ≈ 85 % in 2008-2012 (≈ 80% in 2013)

Part I : Short general overview on the main charasteristics of the French LFS

Collection timing & NRS for one reference week :

- Interviews during the 16 days (2 weeks and 2 days) after the reference week
- Then, 5 days for editing, gathering the data from all regional offices, defining the set of non-respondents
- Then, 2 weeks of collection for the non-respondent survey(NRS)

Part II

II.

Main issues of the weighting procedure

Part II.1 : Initial sampling & Annual updates of the sampling frame

French LFS Sample for 9 years (2009 - 2018) :

- Selected within a new sampling frame : (Housing-) Tax Registers
- One-shot random selection (carried out in 2008) from the 2006 Tax Reg.
- Annual updates of the sample : modifications in the dwelling status (main residence or not,…), new dwellings…

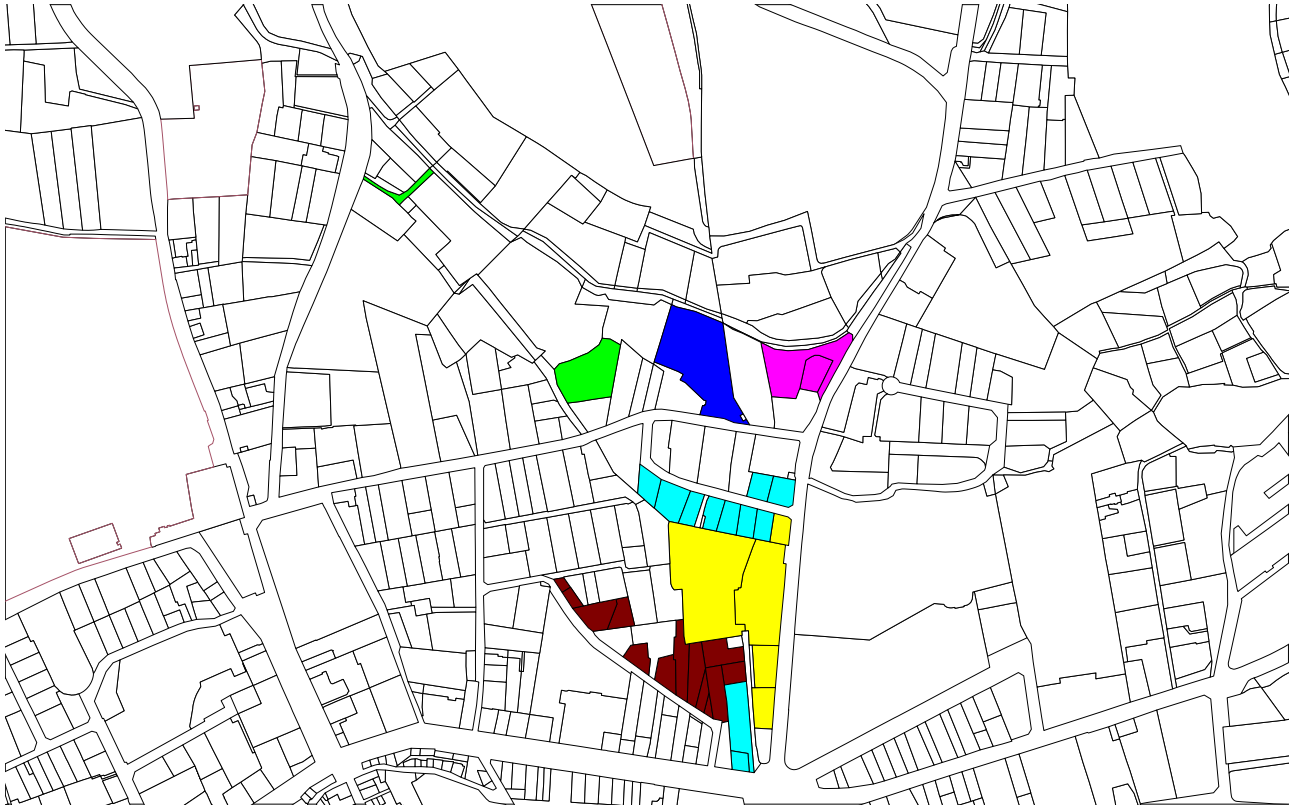
  In April N, tax registers for year N-1 available and used for updating the sample of Q4 N and Q1, Q2, Q3 N+1.

- Descending approach to select the sample :

  1. Selection of sectors (3200), divided into 6 clusters = nearby dwellings (of which ~ 20 main residences).

     Only one cluster of a given sector surveyed at a time; after its 6th wave, replaced by another cluster of the same sector

Part II.1 : Initial sampling & Annual updates of the sampling frame



The six clusters of one given sector.

One given cluster surveyed 6 consecutive quarters

→ Sector life span = 6*6 = 36 quarters = 9 years (2009 - 2018)

Part II.1 : Initial sampling & Annual updates of the sampling frame

Selection of sectors :

- Inclusion probabilities π_{sector} proportional to the number of main residences
- Stratification by region (NUTS 2, i.e. 22 regions)
- Balanced sampling (Fast-Cube Method, see Deville & Tillé, 2004, Chauvet & Tillé, 2006) on age criteria, type of urban/rural areas, incomes, resident status (owner/tenant), type of dwelling…

2.Uniform random attribution of ranks (1 to 6) to the clusters inside the sectors (proba 1/6), such that - for each quarter - the sample of clusters really surveyed (and not only the sample of sectors) is well balanced and stratified by region.

3.Uniform random attribution of reference weeks inside quarters (1 to 13) to the clusters, carrried out by NUTS 2.

Part II.1 : Initial sampling & Annual updates of the sampling frame

Advantages and drawbacks :

- The new sampling frame, tax registers, useful for…
  - Exhaustiveness, in terms of dwellings
  - Good identification of the dwellings over time
  - Recentness with annual updates
  - Information on the closeness of dwellings -> construction of clusters, dwelling by dwelling
  - Important auxiliary information available (charasteristics of the dwellings, of the households), interesting frame to correct for non-response

Part II.1 : Initial sampling & Annual updates of the sampling frame

- But…
  - Ordinary dwellings other than main residences (second homes, vacant dwellings,…) must be surveyed too to check their status.
  - New dwellings and dwellings other than main residences sampled and attached to clusters.

    Complex procedure consisting in sampling some *groups* of such dwellings, in order to ensure that :

    1. The sample is still representative
    2. The size of clusters does not increase dramatically

    Inclusion probabilities in these cases derived by Monte-Carlo simulations

  → Quite homogeneous *a priori* distribution of the dwelling sampling weights $\left(6 \cdot \pi_{\text{sector}}^{-1}\right)$, but distortion and possibly few extreme values after updating the frame (10:1 and sometimes even more)

Part II.1 : Initial sampling & Annual updates of the sampling frame

- Sampling design subject to numerous constraints :
  - Declination by NUTS2 (for both statistical & organisational purposes), by wave (wave approach, rotation bias). Only then, balancing.
  - Practical constraints linked to the geographical breakdown of the network of interviewers
    - Clustering → Non-negligible design effect (~ 1.75 in 2012 for Unemp. Ratio 15-74). Global sample size not excessive
    - Additional Constraints : selection of neighbouring sectors in some little surveyed areas, reference weeks not too close for such sectors

  → Contribution to the distortion of the weights

Questions arising : Trimming, winsorisation of extreme values appeared recently ? New sample to be selected as from 2018, Reallocation by NUTS2?, taking into account the new constraints of the interviewer's network….

Part II.2 : Production of the quarterly corrected weights & integration of the NRS results

General principle to compute the quarterly individual weights :

- Determination of the individuals considered as respondents : for those aged 15 or more, must have answer at least some key questions (regarding the status on the labour market)
- Calibration of dwelling weights, wave by wave, *simultaneously* to correct for non-response bias and to reduce sampling variance (no preliminary separate step to deal with non-response). - CALMAR (Sautory, 1993)
  - Calibr. Variables from the sampling frame, near to those for the balanced selection of the sample, correlated to non-response
  - Calibr. on external margins (age-groups*sex*NUTS2 derived from the census, statistics on dwellings)

Part II.2 : Production of the quarterly corrected weights & integration of the NRS results

- Derivation of individual weights (individual multiplicative factor to account for partial non-response in dwellings)

Advantages & drawbacks :

- No dramatic distortion of the distrib. of the weights at this step ; extreme values arise from previous steps (updates of the frame)
- Sampling frame quite recent and with important auxiliary information related to non-response (see Part II.1)
- But… with only one calibration,
  - Difficulty to separate the effects of correction of non-response vs reduction of variance.
  - Limitation of the number of margins
- Dwellings other than main residence (with possible extreme values) kept in calibration (status may be ≠ between frame & observed)

Part II.2 : Production of the quarterly corrected weights & integration of the NRS results

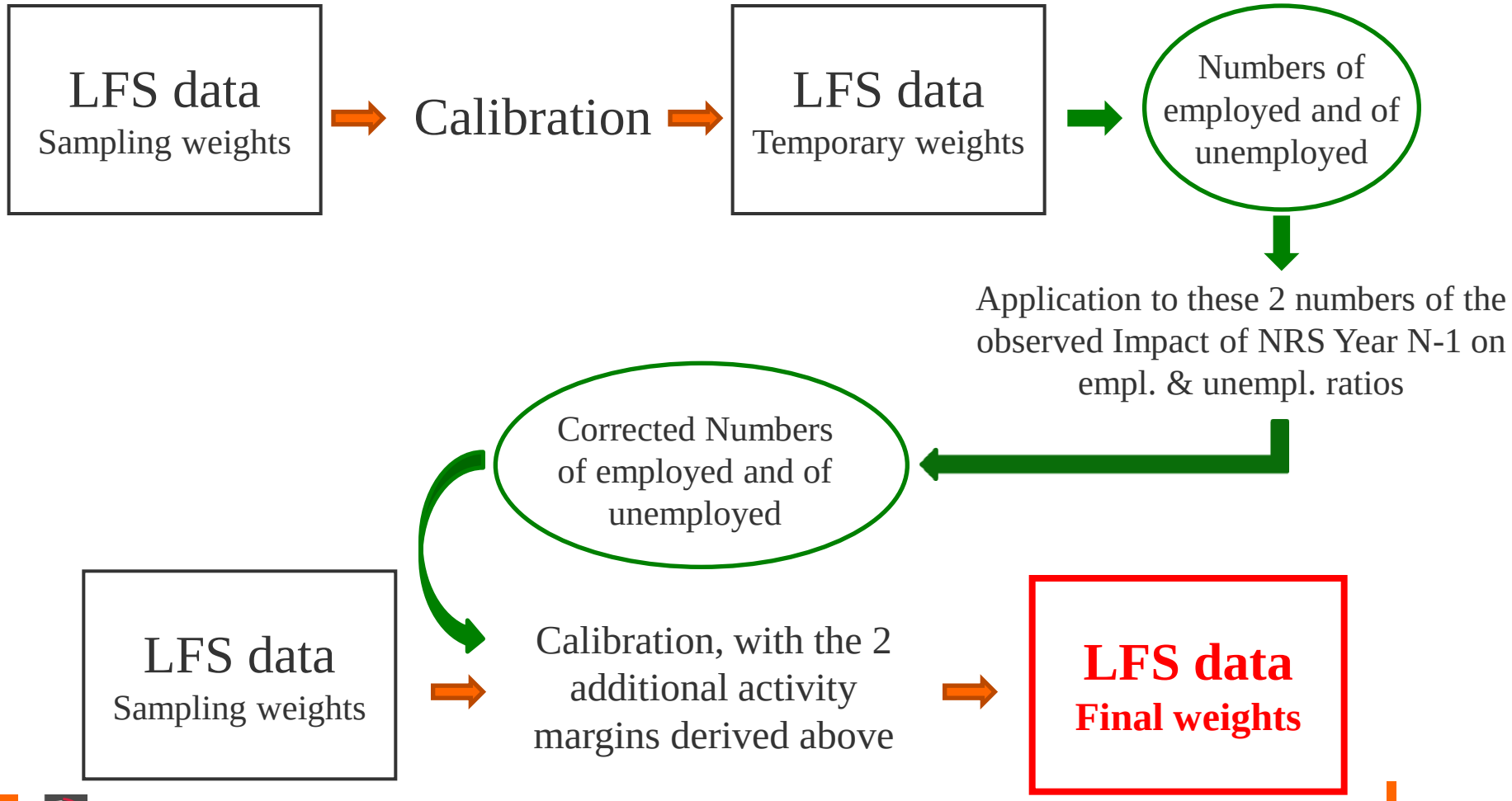
Non-respondent Survey (NRS) :

- Response rate ~ 20% (essentially by postal mail)
- Very short survey (only main questions asked, to derive ILOSTAT)
  - → Respondents to the NRS not incorporated as individual lines in the French LFS microdata file.
  - → Actually, results of the NRS taken into account in calibration
- NRS of quarter Q fully available only in Q+1.
  - → Anticipated effect of NRS in Q based on last year, revision in Q+1.
- Impact

	Unempl. Rate (15+)			Activity Rate (15-64)		
Quarter	Without NRS	With NRS	Diff	Without NRS	With NRS	Diff
T111	9,5	9,5	-0,03	69,8	69,9	0,04
T112	8,7	8,7	-0,05	70,0	70,0	0,07
T113	9,1	9,0	-0,07	70,5	70,6	0,05
T114	9,6	9,6	-0,05	70,3	70,4	0,04

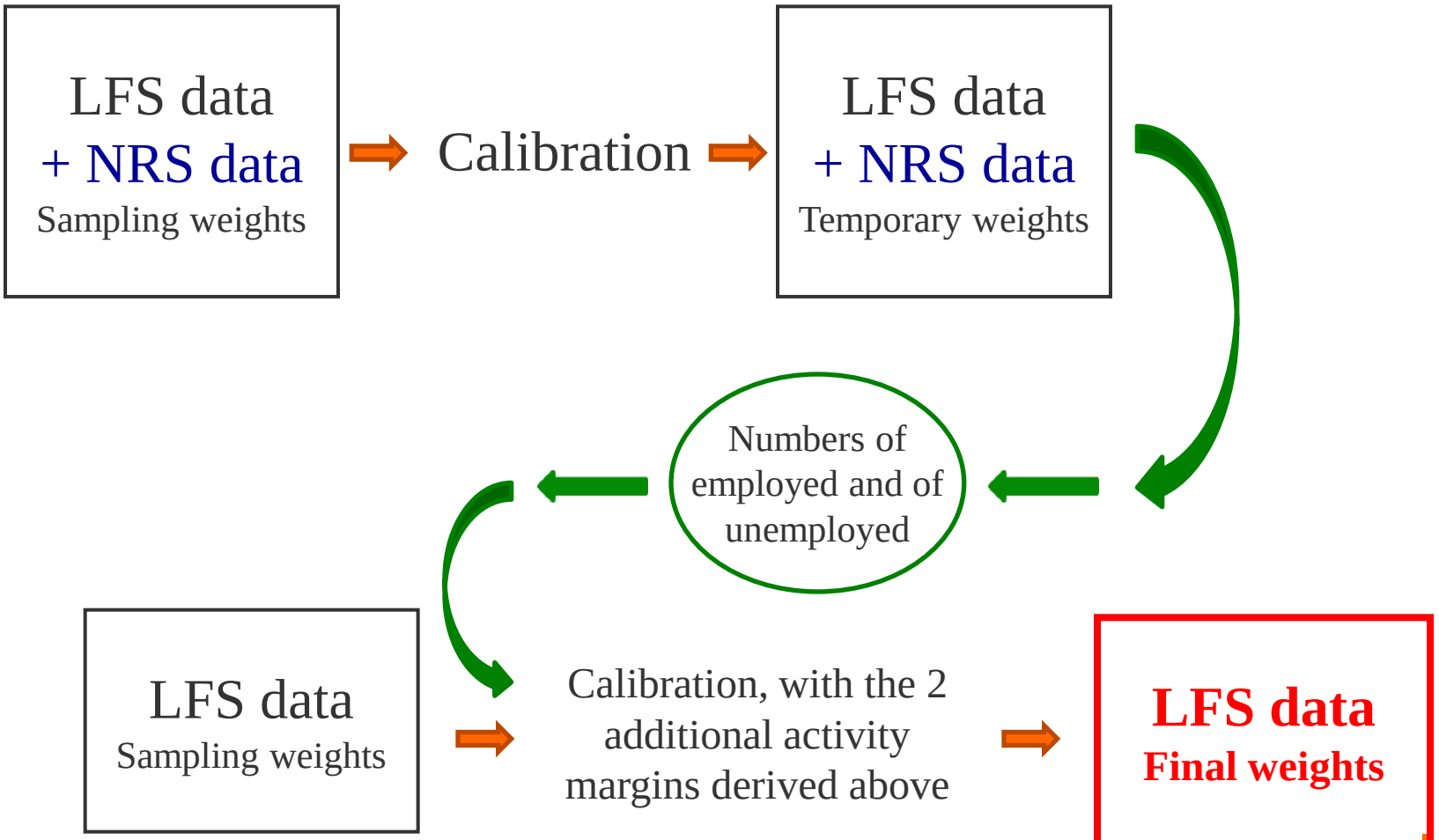
Part II.2 : Production of the quarterly corrected weights & integration of the NRS results

First production of Quarter Q :



Part II.2 : Production of the quarterly corrected weights & integration of the NRS results

First Revision of Quarter Q (when producing Q+1 for the first time) :



Part II.2 : Production of the quarterly corrected weights & integration of the NRS results

Questions arising (/ongoing) :

- Replacing the global calibration by two steps : first, (model-based) correction for non-response, second calibration ?
- Make a distinction between « pure » non-response and « non-survey » (dwellings not surveyed because of an inadequate number of interviewers in some regions) ?
- The use of NRS in the previous cases ?
- Testing a global calibration with instrumental variables to correct non-response

…

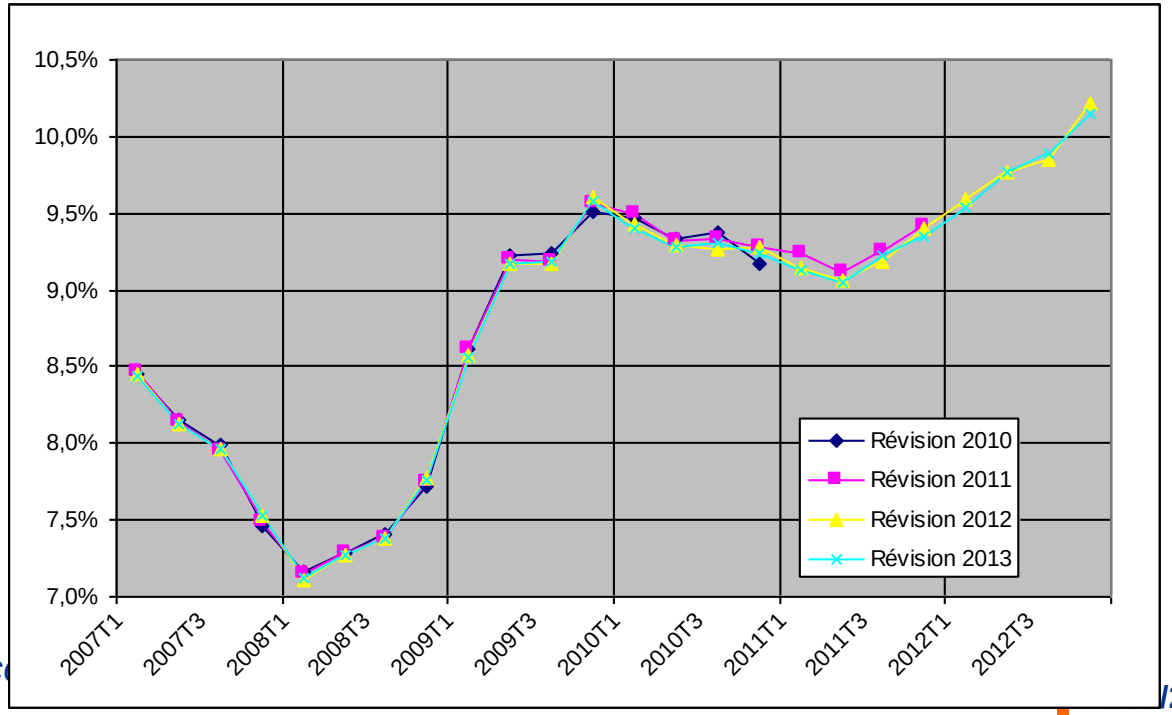
Part II.3 : Revisions

Revisions of the weights of quarter Q, year N:

- The following quarter, due to the results of NRS Q,N (see previous slides)
- In May N+1, N+2, N+3, N+4, due to annual revisions of population margins derived from the census
- Limited impact at aggregated levels

Unemployment rate, 15+

(Calculations performed before the general backcalculation carried out in 2014)



Thank you for your attention

Contact :

M. Fabien Guggemos
Phone : 00 33 1 41 17 54 55
E-Mail : fabien.guggemos@insee.fr

Insee

18 bd Adolphe-Pinard
75675 Paris Cedex 14

www.insee.fr

Informations statistiques :
www.insee.fr / Contacter l'Insee
09 72 72 4000
(coût d'un appel local)
du lundi au vendredi de 9h00 à 17h00