

## 9<sup>th</sup> Workshop on LFS methodology, Rome, 15-16 May 2014

### How is the uniform distribution of reference weeks provided in the French LFS?

Marianne Juillard, Sylvie Le Minez

Insee - May 9, 2014

#### An European obligation

The uniform distribution of reference weeks is included in the settlement of the LFS (see Article 1 of Council Regulation (EC) No 577/98 of 9 March 1998 on the organisation of a labour force sample survey in the Community ).

“In the case of a continuous survey:

- the reference weeks are spread uniformly throughout the whole year;
- the interview normally takes place during the week immediately following the reference week. The reference week and the date of the interview may not be more than five weeks apart, except in the third quarter;
- the reference quarters and years are respectively groups of 13 or 52 consecutive weeks. A list of the weeks making up a given quarter or year is drawn up according to the procedure laid down in Article 8”

This constraint of a uniform distribution of the reference weeks (RW) is obviously important :

- Persons surveyed are interviewed as a reference week: they must answer a series of questions to determine their particular situation on the labor market according to the ILO during this week;
- Uniform distribution of reference weeks is essential to estimate quarterly average unemployment rate, activity, employment etc.;
- It also ensures a correct estimation of annual means: this is for example the case of the number of actual hours performed in the year

#### Short collection periods, including face to face, led to the choice of an areolar sample

For the implementation of a continuous<sup>1</sup> LFS continuously, Insee immediately placed this constraint of a uniform distribution of reference weeks in the sample survey, the uniform distribution of reference weeks being provided in each region. With the new sample drawn from tax bases, the uniform distribution of reference weeks is provided in each region and in each subsample<sup>2</sup>.

The people interviewed in the survey are those living in ordinary homes. Respondents are interviewed at their main residence. The LFS sample is made up of dwellings.

The collection is conducted partly face to face. To meet the deadlines for collection (set at 2 weeks and 2 days maximum after the reference week ), giving several dwellings to survey in the same place and at the same week for each investigator was necessary. **Sampling**

---

<sup>1</sup> In 2003 in France, with a gradual introduction of sub-samples from Q3 2001.

<sup>2</sup> The presentation of the new sample of the Labour Force Survey is based on a lot of the presentation Loonis Vincent, who designed and implemented the new design, based on tax bases (see bibliography ).

**relates on groups of nearby dwellings, called "clusters"<sup>3</sup> : the sample is areolar.** This geographical concentration of, interviews allows savings in terms of both the time required for and the cost of data-gathering and may be also a positive knock-on effect on the response rate. The challenge is to calibrate the cluster size in number of main residences: the size must be large enough to ensure a sufficient workload per investigator, but mustn't be too consistent in order not to lessen too much the precision of estimates (via a clustering effect). The size of these clusters must also be the same throughout the country for reasons of maximum precision : it was set at about 21 main residences.

### **The choice of a rotation pattern (6-) led to sample sectors containing 6 geographically close clusters**

Finally, the LFS main issues is to provide both an estimate of quarterly changes in the unemployment rate and an estimate of the average annual unemployment rate ; this led to a tradeoff between conserving the sample from one quarter to the other in the one hand, and renewing completely the sample from one quarter to another in the other hand.

Insee choose the following rotation pattern, knowing that It wanted in addition follow individuals over a relatively long period ( a little beyond year) :

- Dwellings are interviewed six consecutive quarters
- 1/6 of the dwellings is renewed each quarter

Each quarter , the sample is thus made up of six sub-samples : a sample of clusters of housing interviewed for the first time face to face, 4 samples respectively interviewed over the phone for the 2nd , 3rd, 4th and 5th time, and a sample interviewed face-to-face for the sixth and last time and leaving the sample in the following quarter. Each quarter, a sub-sample should be replaced.

Due to constraints collection (face to face, time limited due to the collection of an interrogation under reference week), **Insee choose to replace outgoing sample after 6 quarters with a sample of clusters geographically close**, with similar characteristics, which limits the losses in accuracy of quarterly changes. It also allows to rely on the same network investigators.

Consequently, it was necessary to build sectors, i.e packages of clusters of dwellings geographicallyly close.

To satisfy this last constraint , the solution was to draw the survey sample at once to cover a relatively long period of 9 years<sup>4</sup>, which implies sectors including 6 clusters of 21 main residences are constructed.

---

<sup>3</sup> These groups have accommodation close called " areas" as the sample was drawn in the population census, then " clusters" where the sample was taken in the new frame, made files tax of the property tax.

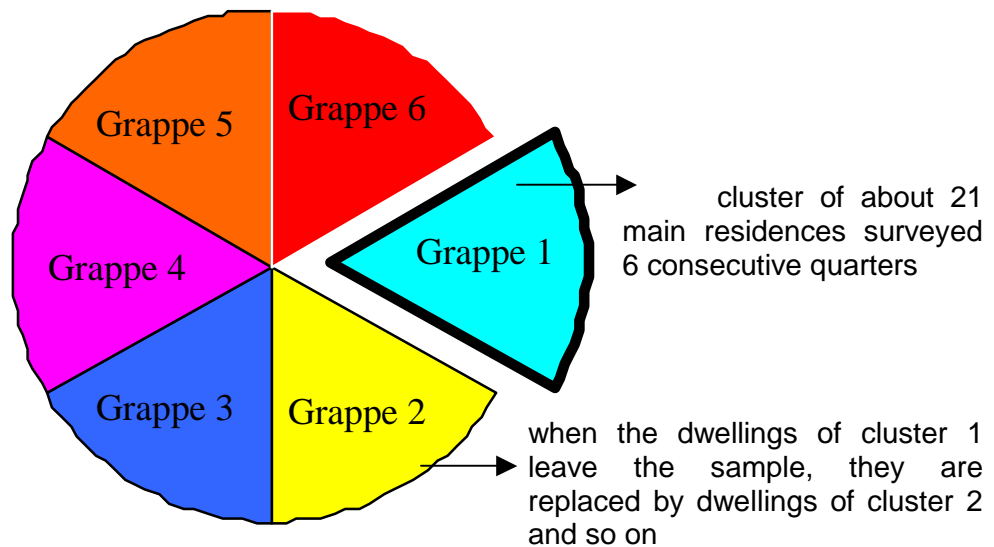
In the first case ( sample drawn from the censu) areas consisted of 20 ordinary housing identified in the 1999 census ; investigators should " rake " the whole area , so as to identify all housing units, new housing created since 1999 and only investigate primary residences.

In the second case ( sample drawn from the tax register , the units are close but not contiguous : a cluster consists of 21 main homes within the meaning of the property tax, but also includes second homes and vacant housing as defined in the Tax register; clusters were formed from 2006 tax register, and data updated every year through matching with the latest release of tax register (new dwellings , change of status residences, change the name of occupants ); investigators therefore are assigned specific dwellings lists to investigate.

In the following text, only the term "cluster " is used ; the following text deals primarily with the new sample drawn from the tax register.

<sup>4</sup> This is the time between two population censuses before the census is continuous from 2006.

**A geographical sector including 6 clusters**  
**Sector life span : 6 x 6 clusters quarters query = 9 years**



### **The sampling design**

Finally, the surveying plan can be summarised in the following ample design is determined as follows :

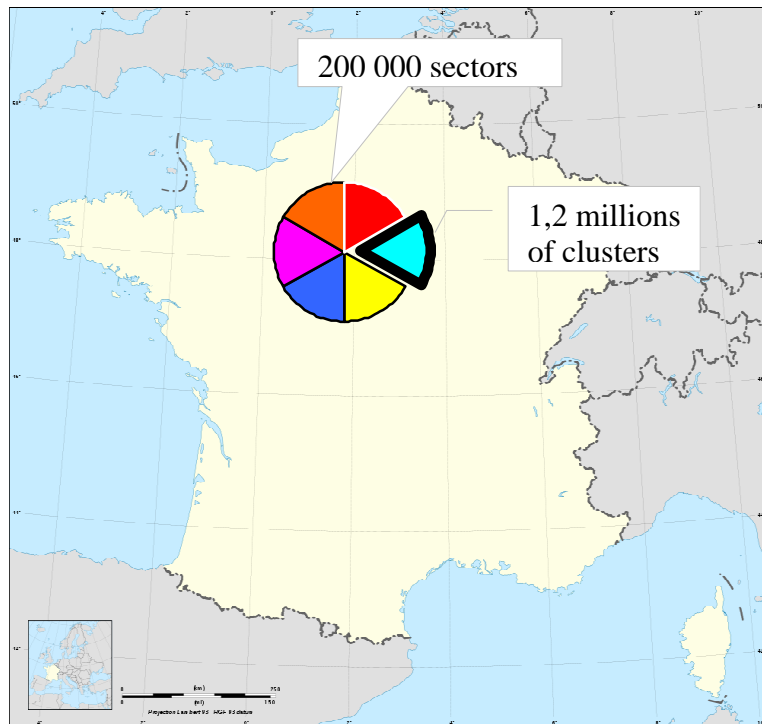
1. the number of main homes to investigate each quarter is determined ( 67,350 per quarter under the new sample drawn from the tax register and mature from Q3 -2011)
2. the number of sectors is deduced, knowing that a cluster of 21 main homes is interviewed each quarter (sectors approximately are  $3210 = 67\ 350/21$  )
3. sectors are drawn and then assigned an order of entry into the sample
4. each of the 6 clusters constituting each sector is assigned a cluster number
5. each sector is assigned a reference week in the quarter : cluster 1 will be surveyed for the first time in quarter determined in step 3; subsequent interrogations are spaced 13 weeks, and so are the interrogations of the remaining clusters in the sector.

### **Partition of metropolitan France in clusters and sectors**

This procedure implies that french territory has been squared so as to form clusters of 21<sup>5</sup> dwellings occupied as a principal residence and geographically close, then sectors composed of 6 clusters geographically close. With geolocation data contained in tax files, this grid could be done in record time and by computer (which was not previously the case with the sample drawn from the census which required a very extensive field work in order to achieve this division ) .

---

<sup>5</sup> Clusters also contain non-main homes, sampled if their number exceeds 10 On the ground, investigators must verify the status of the housing field. There may be a second home has become a main residence and vice-versa. In addition, each year, the sample is refreshed with the last delivery of tax bases: aggregating each cluster new homes by sampling so as not to grow too cluster size. Further away from the date of initial sample selection, the larger the number of clusters varies main homes (some clusters incorporating new homes are larger, some clusters have a size decreases because homes are destroyed, homes become second homes ...).



Once the partition into sectors has been made, the sample design has to be defined. The following decision was made :

- The sectors sample is stratified by region, according to a proportional allocation. (probability proportional to the size in terms of number of main residences)  
The issue of allocation is very important : it determines regional and national precision, and regional workload.
- The selection of sectors<sup>6</sup> within each region is proportional to the number of main homes according to a balanced<sup>7</sup> socio-demographic criteria survey.

### **Determining the rank of introduction of sectors in the sample**

Within each region, sectors of sub-sample 1 were drawn with a probability of 1/6th according to a balanced survey ( same variables as above or less in small regions). Then, the second sample was selected from the remaining 5/6th of sectors with a probability equal to 1/5th according to a balanced survey and so on until the sixth and final subsample.

Using the same methodology, a cluster number from 1 to 6 was assigned to each of six clusters comprising each sector.

### **Uniform distribution of reference weeks ( RW )**

Each sector is assigned as noted above a reference week. Once the reference week is attributed, the survey for later quarters are spaced exactly 13 weeks apart.

<sup>6</sup> To avoid the selection of isolated sectors in sparsely populated areas, sectors have been aggregated by two in municipalities where the estimated probability that at least three geographic areas are selected by the procedure is less than 0.95.

<sup>7</sup> See J-C. Deville and Y. Tillé (2004); practical implementation with SAS macro FastCube.

The uniform distribution of reference weeks had to meet the following requirements<sup>8</sup> :

- A uniform distribution of the RW by region
- A uniform distribution of the RW for each of the 6 sub-samples<sup>9</sup>
- The size of the sample by region and sub-sample is fixed
- In urban areas having at least 5 sampled , the difference between the reference weeks is higher than or equal to 2

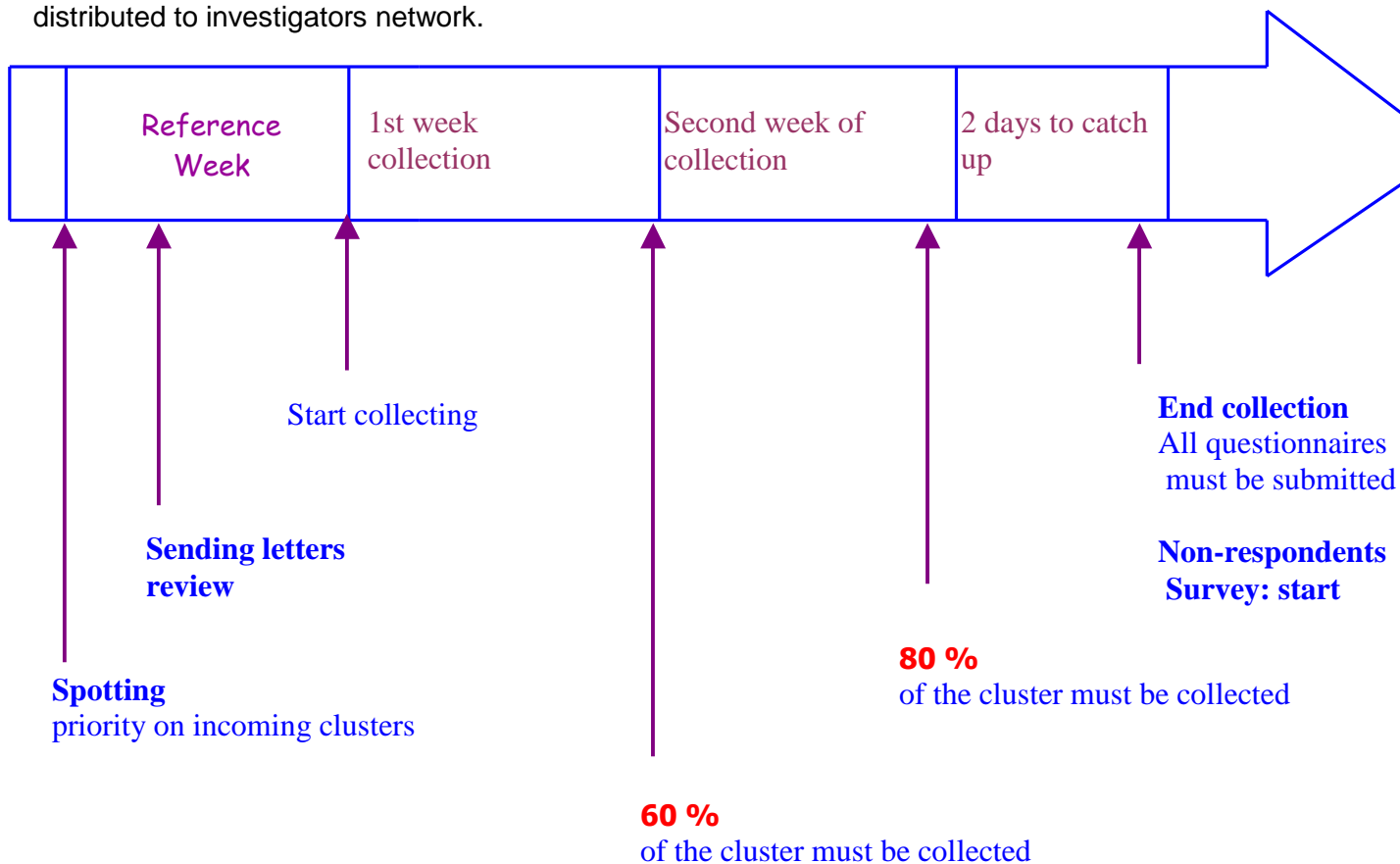
The resolution of this problem is similar to a game SUDOKU.

### Organization of field collection that ensures the uniform distribution of reference weeks (RW)

Investigators have two weeks and two days after the end of the reference week to collect a cluster associated with the reference week (except extension collection : holiday periods for example).

The last two days belonging to the 3rd week of the RF are awarded for "catching up". They must serve the last reminders, confirm the "unreachable" and refusals. To take holiday and celebrations periods into without compromising the smooth collection, extensions of collection are granted on the occasion of bridges and public holidays, summer holidays and year-end holidays.

Each year, a calendar is based on these events. It is addressed to regional offices and then distributed to investigators network.



<sup>8</sup> Other requirement is needed because of adjustments made to ensure a sufficient workload in sparsely populated residential areas (compare note 7) : in a two-sector , the difference between the reference weeks of the two sectors is greater than or equal to 3 (noting that the sub-sample is the same in both sectors ).

<sup>9</sup> This second condition was not imposed during the first sample drawn from the census; or it proves important: some structural variables are addressed q'aux one sub-sample units, as well as the ad hoc module; correction procedures of non-response and improved accuracy are on each subsample.

Given the brevity of the collection period, it is essential to begin the interviews from the first day of the period of investigation, which involves sending or filing the notice letters from the reference week, and trying to make an appointment this week then for the next two. This is why Insee set collection targets: investigators must make every effort to at least 60% of their cluster is collected during the first week, and at least 80% at the end of the second. Trying to make the most of investigations at the beginning of the first week following the reference facilitates interviews, but it is also the best way to get a high response rate in the presence of an extended period of collection short. Indeed, we observe that most of the non-responses do not correspond to refusal, but impossible to reach or long-term absentees. They are also partially recovered through a survey of non-respondents (PAPI and CAWI collection) that runs on weeks 3 and 4 after RW: this survey only covers the ILO module and enriches the margins calibration (number of employed and unemployed).

Making contact attempts spread over three weeks (ie after the reference week) therefore gives more chances to get to join the households concerned. This is even more important than these non-respondents have very special characteristics on the labor market point of view (these are often frames or professions very busy, as shown by surveys specific postcards made from them): noninvestigating them influences strongly the raw results of the survey.

In addition, investigators must transfer the data collected every day and especially at the end of the first week. Finally, as soon as possible and no later than the beginning of the second week, investigators should ask regional offices to send letters to non-respondents.

Quarterly monitoring enables regional offices to know each of their investigators and steer closer collection:

- The rate of progress of the collection on main residences (situation after 1 week, 2 weeks and 2 weeks and 2 days): this indicator is the assessment of the actual progress of investigators (number of main residences surveyed at a time given the total number of main residences to investigate);
- The failure rate is the share of refusal, absent long-term, impossible to reach and impossible to achieve among all units to investigate;
- The surveys accepted rate is the ratio between the number of address files giving rise to an accepted survey and the total number of units of the field. It is available by querying the cluster rank;
- The rate of investigations conducted by proxy: it gives the share of individual questionnaires conducted by proxy.

The LFS application, which was renovated in 2013, now allows multiple players to have all these indicators permanently from their computer workstation.

### **The weighting scheme does not take into account the RW**

Correction of non-response and improved accuracy are in one step with a calibration wave by wave (CALMAR procedure in SAS): variables from the tax register and external data on demographic and housing are used<sup>10</sup>.

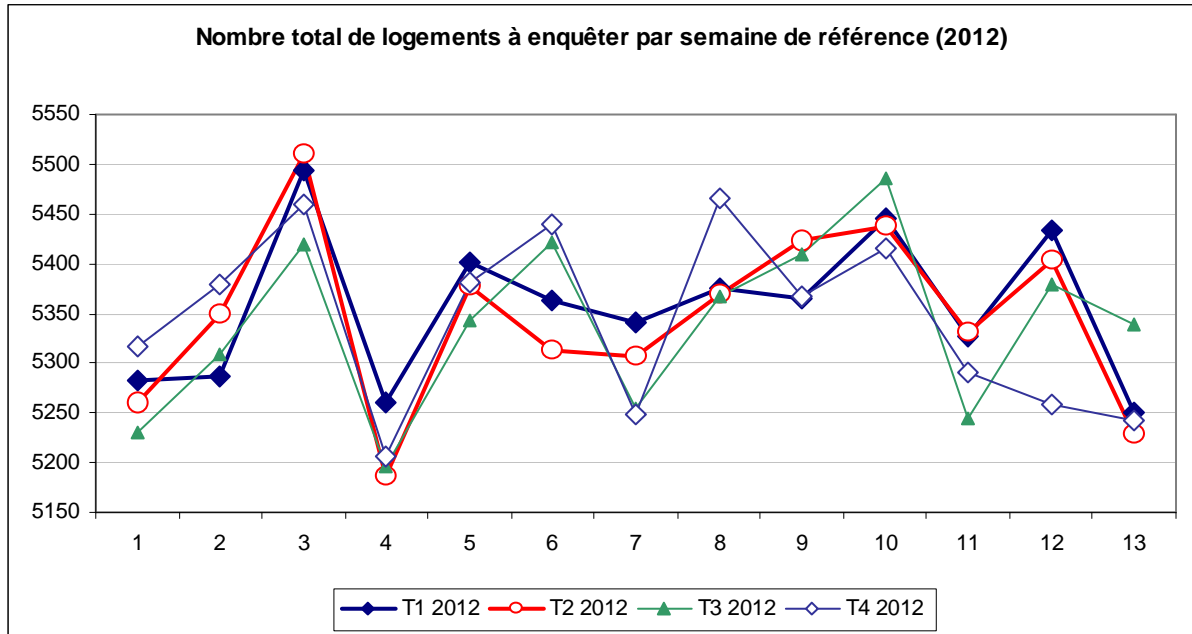
The only data available per week would be those of the frame: This exercise was conducted in Q1 2012 due to a social movement of investigators, which led to decrease the response rate of nearly 10 points some weeks in the first quarter; estimates of the main variables proved identical to those estimated with the usual calibration.

---

<sup>10</sup> See the presentation of Fabien Guggemos (Insee) during the workshop for more details.

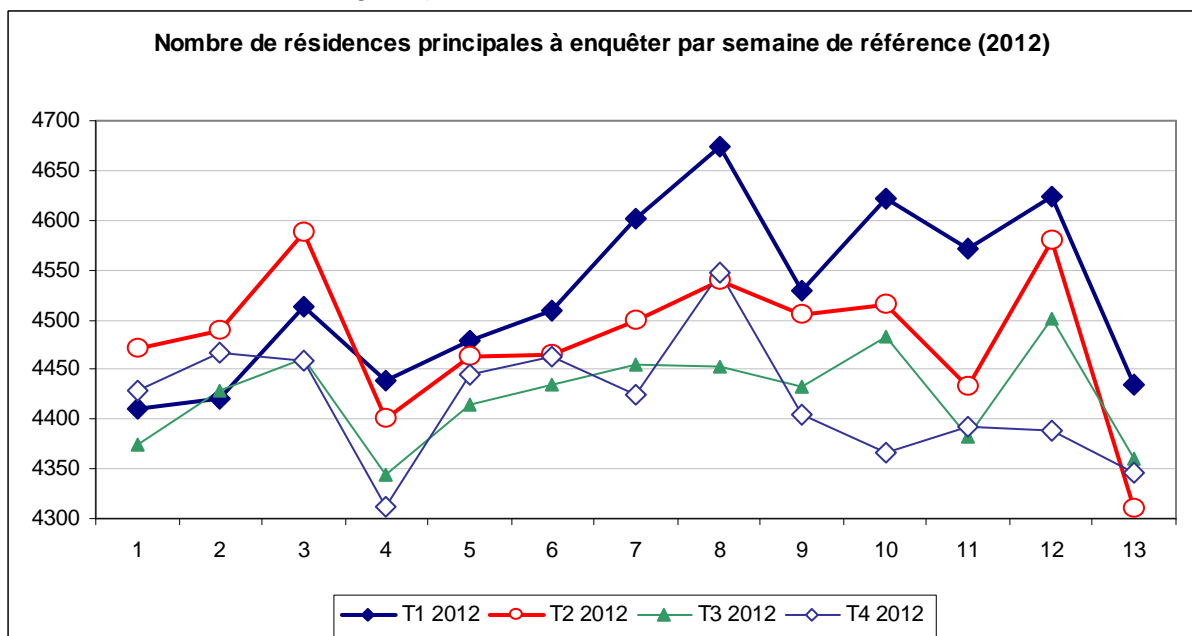
## Some stylized facts

### Number of dwellings to survey each quarter

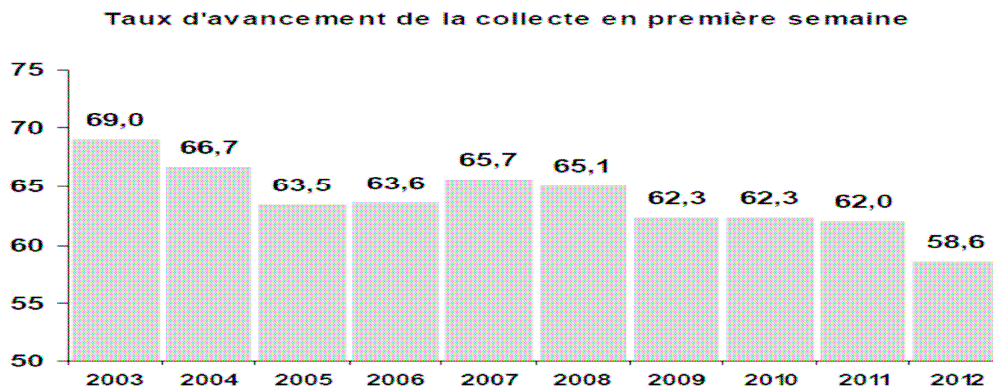


### Number of main residences to survey each quarter

(here, the status of "main residence" is observed by the investigator; It is sometimes different from the status in the tax register)

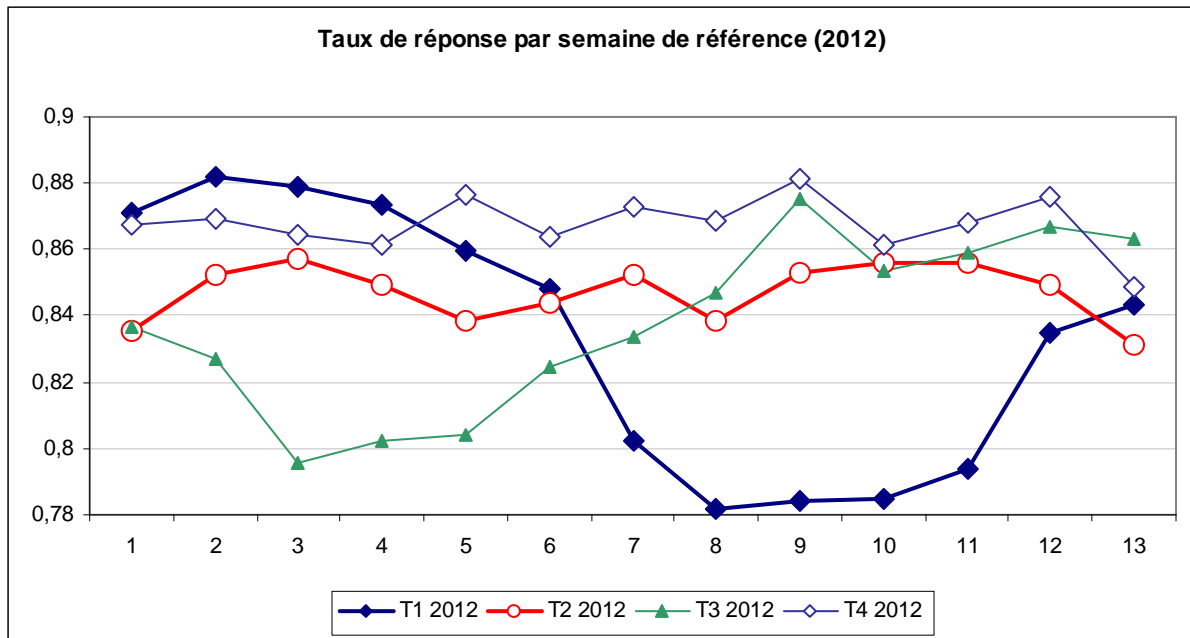


### The rate of progress of the collection after one week



Source des graphiques : INSEE - Enquêtes Emploi 2003 à 2012

### The response rate per week



### Conclusion

The sampling and monitoring of the timeliness of field collection ensures a fairly good distribution of reference weeks per quarter.

INSEE is organized to implement an Internet collection in the LFS. Experiments will take place from 2015. Uniform distribution of RW will obviously be respected. The challenge will be to respect a short collection period with several modes of collection: in case of no answer on the Internet, the collection will be an investigator.



## Bibliography

V.Loonis (2009), « La construction du nouvel échantillon de l'enquête Emploi en continu à partir des fichiers de la taxe d'habitation », actes des Journées de méthodologie statistique, Insee

[http://jms.insee.fr/files/documents/2009/52\\_4-JMS2009\\_S02-5\\_LOONIS-ACTE.PDF](http://jms.insee.fr/files/documents/2009/52_4-JMS2009_S02-5_LOONIS-ACTE.PDF)

M. Christine (2002), « La construction de l'échantillon de la future l'Enquête Emploi en Continu à partir du Recensement 1999 », in Actes des Journées de Méthodologie Statistique, Insee Méthodes, n° 100, pp. 175-229.

J-C. Deville et Y. Tillé (2004), Efficient balanced sampling : the cubemethod, *Biometrika*, 91, 893-912.

G. Chauvet (2006), De nouvelles macros SAS d'échantillonnage équilibré. Technical report, ENSAI, Rennes