

# **ESTIMATION DES CESSATIONS D'ENTREPRISES : méthode et résultats**

D. Francoz

## **Introduction**

*La démographie d'entreprises consiste à observer les populations d'entreprises ainsi que les mouvements (créations et cessations) qui affectent ces populations.*

*Le dispositif actuel, qui s'appuie sur l'exploitation statistique du répertoire SIRENE, se compose principalement de séries de créations d'entreprises et de stocks.*

*Outre le fait de permettre une analyse conjoncturelle et d'aider à déterminer les stocks d'entreprises, la mise en place de séries de cessations apporte des éléments complémentaires aux travaux statistiques élaborés au sein du SISA. En effet, on peut imaginer que cette information pourrait être mise à profit au moment de l'élaboration des comptes nationaux et des prévisions conjoncturelles.*

*L'objectif de l'étude est de mesurer le nombre de cessations survenues dans une période de temps déterminée.*

*Pour pouvoir faire une comparaison avec les créations d'entreprises, il est nécessaire de mettre en place deux types de séries :*

- des séries annuelles (comparaison des volumes annuels) ;*
- des séries mensuelles (comparaison des évolutions).*

*La mesure des cessations à partir de l'exploitation des fichiers liés au répertoire SIRENE met en évidence un certain nombre de problèmes qui ne se posent pas dans le cas des créations.*

*Quels sont ces problèmes ? Quelles solutions sont actuellement envisagées ?*

## Problèmes liés à l'utilisation du répertoire

L'exploitation statistique du répertoire SIRENE nécessite de bien connaître et prendre en compte ses spécificités liées à sa vocation administrative.

Le fichier DEMO est un fichier mensuel qui enregistre tous les mouvements de mise à jour du répertoire SIRENE.

Chaque mouvement est caractérisé par deux dates : la date à laquelle est survenu l'événement que l'on prend en compte (date d'événement) et la date à laquelle on a traité cet événement (date de traitement).

La date qui nous intéresse dans la mesure des mouvements d'entreprises est la date d'événement.

### *Estimation des créations*

Le dénombrement des créations d'entreprises repose sur l'hypothèse contrôlée selon laquelle l'écart entre la date d'événement et la date de traitement est très faible.

Cela nous permet d'estimer les créations ayant eu lieu un mois  $m$  donné par les créations traitées ce mois  $m$  et figurant dans le fichier DEMO du mois  $m$ .

On connaît dès le 1<sup>er</sup> du mois suivant le nombre de créations du mois  $m$ .

L'hypothèse que l'on fait est justifiée par la procédure d'enregistrement des créations d'entreprises : la personne qui crée une entreprise a obligation de déclaration auprès d'un Centre de Formalités des Entreprises (CFE), guichet unique des déclarations administratives des entreprises, dans les 15 jours qui suivent le début de l'activité. D'autre part, l'INSEE est tenu de fournir un numéro d'immatriculation à l'entreprise dans les 5 jours suivant le dépôt du dossier au CFE.

La rapidité, à la fois dans la déclaration et dans le traitement des créations d'entreprises valide l'hypothèse de quasi-égalité entre la date d'événement et la date de traitement.

En revanche, les cessations, lorsqu'elles sont déclarées, le sont avec retard. En conséquence, les enregistrements effectués le mois  $m$  ne reproduisent pas les événements réellement intervenus au cours de ce mois.

On ne peut donc pas obtenir les cessations du mois  $m$  dès le début du mois  $m + 1$ .

## ***Circuits d'enregistrement des cessations***

Comme pour l'immatriculation, la cessation doit être déclarée au CFE. Cependant, malgré cette obligation, l'entreprise néglige bien souvent de la faire. Afin de maintenir la qualité du répertoire et de traiter ces non-déclarants, appelés faux-actifs, on procède à des Enquêtes d'Amélioration du Répertoire (EAR). Ces enquêtes, réalisées auprès de populations à risque (exemples : unités n'ayant fait l'objet d'aucune mise à jour depuis plus de 5 ans, personnes physiques de plus de 70 ans, entreprises non répondantes aux Enquêtes Annuelles d'Entreprises ou absentes dans les sources administratives). La stratégie mise en place consiste à repérer à travers ces enquêtes le plus d'unités faussement actives dans le répertoire.

Lors de ces enquêtes, on obtient l'information sur l'existence ou non de ces unités, mais pas toujours sur la date à laquelle s'est produit l'événement. On enregistre alors une date d'événement fictive fixée théoriquement par convention au 25 décembre de l'année  $n-1$ . En pratique, en fonction des pratiques des différentes Directions Régionales, on trouve aussi comme date fictive la date du jour de traitement.

En outre, quand on connaît la date de cessation de l'unité, il s'avère que l'unité peut être morte depuis très longtemps.

On se trouve donc face à deux problèmes qui font obstacle à un dénombrement simple des cessations :

- le délai d'enregistrement des cessations ;
- l'existence de dates d'événement fictives

## **Quelles conséquences pour le dénombrement des cessations ?**

*On ne peut pas estimer les cessations en date d'événement par les cessations en date de traitement.*

Le *tableau 1* fait apparaître des différences significatives entre les cessations en date d'événement et celles en date de traitement. Ces différences n'affectent pas seulement la valeur absolue, mais aussi les évolutions d'une année sur l'autre.

Ainsi, si on utilise la date de traitement, on note une augmentation entre 1987 et 1988 alors qu'en date d'événement on note une baisse du nombre de cessations (origine CFE).

Tableau 1

**Nombre annuel de cessations enregistrées selon la date choisie**

	Date de traitement		Date d'événement	
	origine CFE	origine EAR	origine CFE	origine EAR
1987	215497	72476	215688	88328
1988	218883	96749	213609	73775
1989	218612	69638	227830	47068
1990	232782	49236	230048	50347
1991	225941	50268	220215	53596
1992	222533	76317	219008	58596
1993	219736	67306	205527	58087

Source : DEMO. Données calculées sans tenir compte de l'existence de dates fictives

Ces différences sont encore plus sensibles pour les cessations dues aux enquêtes d'amélioration du répertoire.

## Problème des cessations détectées par enquêtes

Les cessations d'origine EAR représentent, selon les années, entre 17 et 30 % des cessations annuelles. En conséquence, elles ne peuvent pas être ignorées de l'analyse démographique.

Leur prise en compte soulève le problème des dates d'événement fictives. En effet, la proportion de cessations avec une date fictive est d'environ 50 % des cessations d'origine EAR.

*On ne peut pas utiliser directement les cessations en date d'événement.*

Ce problème se pose pour la constitution des séries mensuelles et des séries annuelles.

Le premier obstacle est celui des dates fictives. Si on ne conserve que les cessations pour lesquelles la date d'événement est bonne, on va sous-estimer le nombre annuel de cessations (certaines cessations auxquelles a été attribuée par convention une année de cessation  $n$  se sont effectivement produites cette année  $n$ ).

En calculant les cessations sans tenir compte des problèmes de date, on surestime le nombre de cessations en 1987 et 1988. En effet, l'année 1987 correspond à la mise en place du nouveau système d'exploitation du répertoire SIRENE. Des volumes impor-

tants d'enquêtes d'amélioration ont été réalisés en 1987 et 1998. Ils ont permis de détecter de nombreuses cessations d'entreprises pour lesquelles la date d'événement était inconnue.

Un deuxième obstacle à l'utilisation de la date d'événement est le délai d'enregistrement des cessations : il faut un temps assez long pour avoir la quasi-totalité des cessations survenues à un instant donné. On remarque, dans le tableau n°1, une baisse des cessations d'origine CFE en date de traitement en 1992 et 1993. Cette tendance est en partie due au fait que tous les mouvements survenus ces années-là n'ont pas encore été enregistrés.

Le *tableau 2* analyse le rythme d'enregistrement des cessations sur la période 1987-1991.

Il faut attendre la fin de la deuxième année suivant l'événement pour disposer de 99 % des cessations.

*Tableau 2*

**Répartition des cessations en date d'événement en fonction du délai d'enregistrement des cessations en mois**

0 mois	1 mois	2 mois	3 mois	4 mois	5 mois	6 mois	7 mois	8 mois	9 mois
15,9%	55,1%	73,0%	81,4%	85,9%	88,8%	90,9%	92,4%	93,6%	94,4%

10 mois	11 mois	12 mois	24 mois	36 mois
95,2%	95,8%	96,3%	98,9%	99,6%

*Cessation d'origine CFE*

Si l'on veut fournir une information dans un délai proche de la fin de la période de référence, on ne peut pas se contenter des observations. Il faut donc envisager une modélisation.

Compte tenu des difficultés rencontrées dans l'estimation des cessations d'origine EAR, il ne sera question par la suite que de l'élaboration des séries mensuelles à partir des cessations enregistrées par la voie des CFE.

En ce qui concerne la modélisation des cessations enregistrées par voie d'enquêtes administratives, plusieurs pistes sont à l'étude.

## Méthode d'estimation des séries mensuelles

L'examen des séries de cessations d'origine EAR ne fait pas apparaître de régularités au cours du temps. Il est donc impossible de les prendre en compte dans le calcul des séries mensuelles. En revanche, elles doivent intervenir dans le calcul du volume annuel de cessations, si l'on veut évaluer le solde annuel entre le nombre de créations et le nombre de cessations.

Les travaux réalisés ont principalement porté sur la mise en place de séries mensuelles à partir des cessations d'origine CFE.

L'objectif est de fournir une information conjoncturelle sur l'évolution des cessations. Les données publiées doivent donc être les plus récentes possibles.

L'idée est d'utiliser l'information sur les délais d'enregistrement des cessations pour estimer le nombre de cessations survenues dans les derniers mois.

### *Calcul général*

On utilise les notations suivantes :

$n_{ij}$  : nombre d'unités ayant cessé leur activité le mois  $i$  et enregistrées le mois  $j$ .

$n_i$  : nombre total d'unités cessées le mois  $i$ .

$$n_i = \sum_j n_{ij}$$

$P_k^m = (P_k^m)$  : probabilité d'enregistrer au plus tard  $k$  mois après les cessations survenues le mois  $m$ . Il s'agit d'une proportion cumulée de mouvements enregistrés au titre du mois  $m$  au bout de  $k$  mois.

Des  $P_k^m$  différents sont calculés selon les mois de l'année.

On fait l'hypothèse que l'on a enregistré toutes les cessations survenues avant le 31 décembre 1992. On a donc dans ce cas :

$$\hat{n}_i = \sum_{j=i}^{\text{mar95}} n_{ij} \text{ pour } i = \text{janvier 1987 à décembre 1992}$$

On estime les cessations survenues depuis le 1<sup>er</sup> janvier 1993 à partir des cessations déjà enregistrées et des probabilités d'enregistrement avec retard selon la formule :

$$\hat{n}_i = \frac{\sum_{j=i}^t n_{ij}}{\hat{p}_m^{t-i}} \text{ pour } i \geq \text{janvier 1993}$$

### *Calcul intermédiaire des $P_k^m$*

Dans un premier temps, nous avons calculé les  $P_k^m$  à partir des cessations survenues jusqu'en décembre 1992 suivant la formule :

$$P_k^m = \frac{\sum_{c=0}^k \sum_{a=87}^{92} n_{am,am+c}}{\sum_{a=87}^{92} \hat{n}_{am}}$$

Ce calcul a pour inconvénient de ne pas tenir compte des événements les plus récents, dans le sens où l'on attend d'avoir enregistré toutes les cessations survenues un mois donné pour utiliser l'information dans le calcul.

L'utilisation des  $P_k^m$  ainsi calculés repose sur l'hypothèse d'une stabilité dans le temps des comportements d'enregistrement des cessations d'origine CFE.

Une étude sur le pourcentage cumulé d'enregistrement dans les 12 premiers mois en fonction de l'année et du mois d'événement a montré une évolution de ce pourcentage au cours du temps. En 1987 et 1988, les pourcentages d'enregistrement dans les premiers mois sont plus faibles que dans les années suivantes.

L'année 1989 a été exclue à cause de la grève qui a affecté les centres informatiques de l'INSEE en octobre 1989. Les effets de cette grève se font sentir sur les pourcentages d'enregistrement de presque tous les mois de l'année 1989.

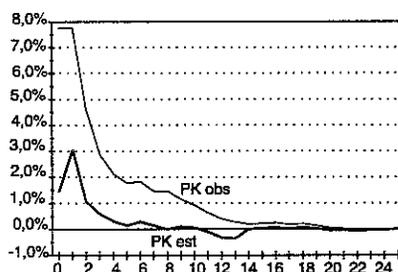
Une méthode a été mise en œuvre pour estimer les  $P_k^m$  en tenant compte des événements les plus récents. On trouvera en annexe le détail des calculs qui font intervenir le nombre de cessations du mois  $m$  enregistré au bout de  $k$  mois et le rapport entre ce nombre et celui obtenu au bout de  $k + 12$  mois.

L'estimation, par cette méthode, des  $P_k^m$  a été testée en estimant le nombre de cessations survenues en 1993 en fonction du nombre de mois d'observations. On a comparé les résultats avec ceux obtenus avec le calcul initial des  $P_k^m$ . On constate qu'en utilisant les événements les plus récents, le nombre estimé de cessations d'un mois donné en fonction du nombre de mois d'observations converge beaucoup plus rapidement vers la vraie valeur des cessations.

C'est ce qu'illustre le *graphique 3* qui donne l'estimation du nombre de cessations du mois de février 1993 en fonction du nombre de mois d'estimation et de la méthode de calcul des  $P_k^m$ .

*Graphique 3*

**Estimation du nombre de cessation du mois de février 1993, en fonction du nombre de mois d'estimation et de la méthode de calcul des  $P_k^m$ .**



Dans ce cas, à partir de trois mois d'observations (délai = 2) l'estimation à partir des  $P_k^m$  estimés se situe dans un intervalle de 1 % autour de la vraie valeur. Il faut attendre 10 mois (11 mois d'observations) pour que les cessations calculées à partir des  $P_k^m$  observés se situent dans le même intervalle.

*Tableau 4*

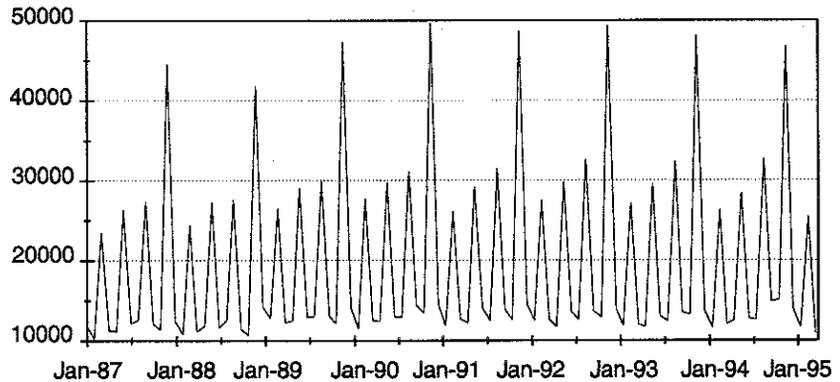
**Résultat de l'étude comparant l'estimation des cessations avec les probabilités d'enregistrement calculées à partir des observations et celles estimées**

délai	1 mois	2 mois	3 mois	4 mois	5 mois	6 mois	7 mois	8 mois et plus
$P_k^m$ observés	1	7	1	4	5	1	1	4
$P_k^m$ estimés	1	10	6	4	1	1	1	0

Lecture : nombre de cas pour lesquels l'estimation des cessations se situe dans un intervalle de 1 % autour de la vraie valeur au bout de K mois. On dispose de 24 observations (soit une observation par mois de janvier 1992 à décembre 1993.) Par exemple, dans 7 cas sur 24, l'estimation réalisée à partir des  $P_k^m$  observés converge vers la vraie valeur au bout de 2 mois. L'estimation réalisée à partir des  $P_k^m$  estimés converge dans 10 cas sur 24 au bout de deux mois.

### Graphique 5

#### **Exemple de séries obtenues Cessations totales brutes**



Une étude plus systématique a été réalisée sur la période janvier 1992-décembre 1993. Pour chaque mois, on observe le délai au bout duquel l'écart entre l'estimation et la vraie valeur des cessations est inférieur à 1 % de la vraie valeur. Le tableau n°4 donne les résultats de cette étude.

### **Résultats et modalités de publication envisagés**

Onze séries de cessations ont été constituées en utilisant la méthode exposée : les cessations totales, les cessations définitives, les cessations par reprise et huit séries de cessations par activité.

Les séries brutes montrent une forte saisonnalité. Le nombre de cessations est systématiquement plus élevé le dernier mois de chaque trimestre et surtout en décembre (cf. graphiques n°5 et 6).

Les séries corrigées des variations saisonnières connaissent une forte évolution entre 1987 et fin 1990. En revanche, depuis 1990, les cessations se caractérisent par la faible amplitude des mouvements observés.

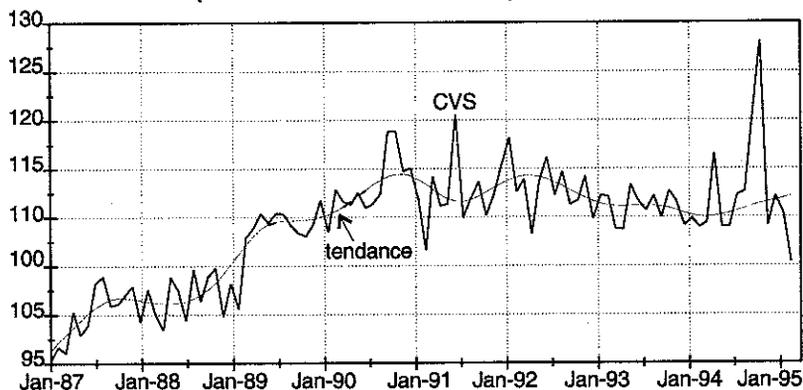
### **Modalités de publication :**

Un problème à résoudre est celui du délai de publication et des modalités de correction des chiffres déjà publiés.

## Graphique 6

### Exemple de séries obtenues

#### Cessations totales (tendances et séries CVS)



Il s'agit de faire l'arbitrage entre deux contraintes :

- publier des chiffres le plus rapidement possible afin que l'information conserve un caractère conjoncturel ;
- les chiffres publiés seront soumis à des corrections minimales qui ne doivent pas remettre en cause le sens de variation de la série CVS.

Le critère utilisé pour déterminer le délai de publication optimal a consisté à comparer le pourcentage d'évolution de la série brute d'un mois sur l'autre et l'écart entre le nombre estimé et la vraie valeur des cessations.

Le plus faible taux d'évolution d'un mois sur l'autre est d'environ 1 %. Il s'observe généralement entre les mois d'avril et mai et entre juillet et août. Dans les autres cas, ce taux est sensiblement plus élevé.

Les résultats de l'étude présentée au § 3.2 nous conduisent à choisir un délai de publication de deux mois.

En effet, dans près de la moitié des cas, ce délai est suffisant pour respecter l'écart de 1 % entre estimation et vraie valeur des cessations.

*Présentation des résultats* : les séries CVS seront publiées en indice (base 100 en 1987) afin de ne pas introduire de comparabilité avec les niveaux de créations mensuelles ni avec les niveaux annuels de cessations lorsque ceux-ci seront disponibles. En effet, l'intérêt de ces séries est bien leur analyse en termes d'évolution.

## Développements futurs

Actuellement, seules les cessations France entière ont fait l'objet d'une estimation.

La même méthode pourrait être utilisée pour établir des séries régionales. Une étude est en cours de réalisation pour examiner les possibilités d'automatisation de cette méthode.

Un autre axe d'analyse consiste à construire un modèle établissant un lien entre créations et cessations par l'intermédiaire du taux de survie des entreprises. Les études réalisées jusqu'à présent sur la survie de générations ont montré une grande stabilité dans le temps du taux de survie des entreprises (les générations 1985, 1987 et 1990 ont été comparées). Ce résultat laisse à penser que l'évolution des créations à un instant donné se répercute mécaniquement sur celle des cessations avec un certain retard. Par exemple, dans le secteur des services aux entreprises qui a connu une explosion démographique à la fin des années 80, l'évolution des cessations sur la période 1987-1994 est atypique par rapport à celle des autres secteurs (croissance des cessations sur l'ensemble de la période).

La mise en évidence d'une telle relation permettrait d'isoler l'effet purement conjoncturel dans l'évolution des cessations.

Un dernier axe consiste à estimer les cessations annuelles. Le calcul du nombre annuel de cessations d'entreprises va permettre de contrôler les évolutions de stocks d'entreprises. En effet, la démographie d'entreprises peut se résumer dans l'écriture de l'équation globale suivante :

$$\text{Stock}_{1/1/n+1} - \text{Stock}_{1/1/n} = \text{Créations}_n - \text{Cessations}_n$$

Le problème qui se pose lors de l'évaluation du stock d'entreprises à un moment donné est celui de l'existence dans le répertoire SIRENE d'entreprises faussement actives. Il s'agit d'entreprises qui ont cessé leur activité mais ne l'ont pas déclaré dans un Centre de Formalités des Entreprises. Certaines vont le faire avec retard, d'autres ne le feront jamais.

L'estimation annuelle des cessations doit prendre en compte à la fois le retard mais aussi l'absence de déclaration. Elle permettra donc de contrôler l'estimation des stocks et leur évolution.

De plus, le calcul annuel des cessations permettra également de quantifier ces mouvements en termes d'emplois perdus. En effet, un des objectifs à moyen terme des travaux de démographie d'entreprises est de réaliser une démographie pondérée en termes d'effectifs. Il sera alors possible de comparer les emplois créés lors des créations d'entreprises à ceux perdus lors des cessations.

L'estimation des cessations annuelles pose actuellement un problème dû à l'absence de régularité de comportement d'enregistrement des cessations d'origine EAR. Il est donc difficile d'estimer les cessations survenues dans la période récente (1994) en appliquant le même schéma d'enregistrement que dans les années antérieures. Un autre problème, lié au précédent est l'estimation des cessations avec date d'événement inconnue.

Les premières estimations réalisées montrent une sous-estimation des cessations d'origine EAR en 1994. Plusieurs pistes sont envisagées pour résoudre ce problème. La première consiste à prendre plus de recul : on remarque qu'avec 3 mois d'informations supplémentaires, le nombre estimé de cessations d'origine EAR augmente de 15 %.

La seconde consiste à introduire dans le modèle une nouvelle variable : la date de création de l'unité cessée.

---

## A N N E X E

---

### Estimation des $P_k^m$

On utilisera les notations suivantes :

$N_k^{mb}$  : nombre cumulé de cessations réelles du mois  $m$  pour les années 1990 à  $b$  enregistrées au bout de  $k$  mois ( $b = 1991$  à  $1995$ ).

$R_k^m$  : rapport entre  $P_k^m$  et  $P_{k+12}^m$ .

On connaît les  $N_k^{mb}$ . L'idée est d'estimer d'abord les  $R_k^m$  puis les  $P_k^m$ . En effet, on peut estimer les  $R_k^m$  par la formule suivante :

$$\hat{R}_k^m = \frac{N_k^{m,b}}{N_{k+12}^{m,b}}$$

On suppose que  $\hat{P}_{48}^m = 1$

On estime les  $R_k^m$  de la façon suivante :

$$\hat{R}_0^m = \frac{N_0^{m,95}}{N_2^{m,95}}$$

$$\hat{R}_2^m = \frac{N_2^{m,94}}{N_{14}^{m,94}}$$

$$\hat{R}_{14}^m = \frac{N_{14}^{m,93}}{N_{26}^{m,93}}$$

$$\hat{R}_{26}^m = \frac{N_{26}^{m,92}}{N_{38}^{m,92}}$$

$$\hat{R}_{38}^m = \frac{N_{38}^{m,91}}{N_{48}^{m,92}}$$

On en déduit les  $P_k^m$  estimés :

$$\hat{P}_{48}^m = 1$$

$$\hat{P}_{48}^m = \hat{P}_{48}^m \times \hat{R}_{38}^m = \hat{R}_{38}^m$$

$$\hat{P}_{26}^m = \hat{P}_{38}^m \times \hat{R}_{26}^m = \hat{R}_{38}^m \times \hat{R}_{26}^m$$

$$\hat{P}_{14}^m = \hat{P}_{26}^m \times \hat{R}_{14}^m = \hat{R}_{38}^m \times \hat{R}_{26}^m \times \hat{R}_{14}^m$$

$$\hat{P}_2^m = \hat{P}_{14}^m \times \hat{R}_2^m = \hat{R}_{38}^m \times \hat{R}_{26}^m \times \hat{R}_{14}^m \times \hat{R}_2^m$$

$$\hat{P}_0^m = \hat{P}_2^m \times \hat{R}_0^m = \hat{R}_{38}^m \times \hat{R}_{26}^m \times \hat{R}_{14}^m \times \hat{R}_2^m \times \hat{R}_0^m$$

On déduit les  $P_k^m$  intermédiaires de la façon suivante :

$$\hat{P}_1^m = \hat{P}_0^m + (\hat{P}_2^m - \hat{P}_0^m) \times \frac{N_1^{m,95} - N_0^{m,95}}{N_2^{m,95} - N_0^{m,95}}$$

$$\hat{P}_1^m = \hat{P}_0^m + (\hat{P}_2^m - \hat{P}_0^m) \times \frac{N_1^{m,95} - N_0^{m,95}}{N_2^{m,95} - N_0^{m,95}}$$

$$\hat{P}_k^m = \hat{P}_2^m + (\hat{P}_{14}^m - \hat{P}_2^m) \times \frac{N_1^{m,94} - N_0^{m,94}}{N_{14}^{m,94} - N_2^{m,94}} \text{ pour } k = 3 \text{ à } 13$$

$$\hat{P}_k^m = \hat{P}_{14}^m + (\hat{P}_{26}^m - \hat{P}_{14}^m) \times \frac{N_k^{m,93} - N_{14}^{m,93}}{N_{26}^{m,93} - N_{14}^{m,93}} \text{ pour } k = 15 \text{ à } 25$$

$$\hat{P}_k^m = \hat{P}_{26}^m + (\hat{P}_{38}^m - \hat{P}_{26}^m) \times \frac{N_k^{m,92} - N_{26}^{m,92}}{N_{26}^{m,92} - N_{14}^{m,92}} \text{ pour } k = 27 \text{ à } 37$$

$$\hat{P}_k^m = \hat{P}_{38}^m + (\hat{P}_{48}^m - \hat{P}_{38}^m) \times \frac{N_k^{m,91} - N_{38}^{m,91}}{N_{48}^{m,91} - N_{38}^{m,91}} \text{ pour } k = 39 \text{ à } 47$$