

# Calage du sous-échantillon annuel de l'Enquête emploi en continu sur les résultats trimestriels

*Insee*

*DMCSI*

*Département des méthodes statistiques*

*Frédéric Ouradou*

---



# L'enquête emploi en continu (EEC)

L'EEC : enquête trimestrielle, reposant sur un panel rotatif de logements, renouvelé par 1/6 chaque trimestre.

L'échantillon d'un trimestre est donc constitué de six vagues :

- une vague « entrante » en 1re interrogation
- la vague entrante du trimestre précédent, en 2e interrogation...
- la vague « sortante », en 6e interrogation

		Trimestre d'entrée des vagues								
		Année N				Année N-1				Année N-2
		T4	T3	T2	T1	T4	T3	T2	T1	T4
Échantillon de l'année N	T1				V1	V2	V3	V4	V5	V6
	T2			V1	V2	V3	V4	V5	V6	
	T3		V1	V2	V3	V4	V5	V6		
	T4	V1	V2	V3	V4	V5	V6			

Pour une année donnée :

- 24 vagues d'enquête
- réparties sur 9 sous-échantillons distincts de logements

# Les résultats de l'EEC

---

La diffusion des résultats de l'EEC :

1. des résultats trimestriels issus de la compilation des six vagues d'enquête du trimestre
2. des résultats publiés en moyenne annuelle, reposant – pour certaines variables d'intérêt – sur l'utilisation de certaines vagues seulement.

Selon un règlement européen, pour compiler les résultats annuels, « *si plus d'une vague est utilisée, le sous-échantillon total utilisé doit être constitué d'observations indépendantes.* »

En France, le sous-échantillon annuel peut être constitué des seules vagues 1 trimestrielles, ou bien sur des vagues 1 et 6.

Dans ce cas, « **la cohérence entre les totaux annuels des sous-échantillons et les moyennes annuelles des échantillons complets doit être assurée pour l'emploi, le chômage et l'inactivité par sexe et pour les tranches d'âge : 15-24 ans, 25-34 ans, 35-44 ans, 45-54 ans et 55 ans et plus** ».

Cette présentation expose la méthode utilisée pour le « calage annuel » de l'année 2013, permettant de garantir cette cohérence.

# Les marges de calage

---

Les marges utilisées lors du calage de l'enquête trimestrielle :

- 130 marges issues du recensement : effectifs de la population par sexe, tranche d'âge et région
- 25 marges issues de la base de sondage, le fichier de la taxe d'habitation (TH) : décile de revenu du ménage, nombre de pièces du logement, statut d'occupation du logement, type de logement...

Ces marges sont conservées pour le calage annuel.

Le règlement impose le calage du sous-échantillon annuel sur les résultats obtenus en moyenne annuelle à partir de l'échantillon complet pour les croisements âge x sexe x statut d'activité : 30 marges supplémentaires.

On peut souhaiter ajouter d'autres marges de calage, notamment pour des variables issues de l'enquête trimestrielle.

Le calage est effectué séparément sur chacune des 6 vagues :

- les marges issues du recensement sont identiques pour toutes les vagues
- les marges issues du fichier TH sont fonction du millésime du fichier.

# Les options de calage

---

Plusieurs options de calage :

1. trois méthodes de constitution et de calage du sous-échantillon annuel (« sous-ech »)

*meth=1* : le sous-ech est composé des seules vagues 1 de chaque trimestre.

*meth=2* : le sous-ech est composé des vagues 1 et 6 et le calage a lieu séparément sur chacune des vagues.

*meth=3* : le sous-ech est composé des vagues 1 et 6 et le calage se fait globalement sur ces deux vagues.

2. deux options de calcul des marges des variables de calage

*optionTH=1* : les marges de calage issues du fichier TH sont calculées à partir des seules vagues du sous-ech

*optionTH=2* : les marges de calage issues du fichier TH sont calculées à partir de toutes les vagues

3. pour chaque variable de calage, choix entre « cohérence annuelle » et « cohérence trimestrielle »

# Généralités sur le calage

---

Les calages sont effectués à partir de la macro Calmar.

Plusieurs méthodes sont proposées :

- méthode du *Raking ratio*
- méthode *Logit*, où les rapports de poids sont contraints dans une fenêtre [*Low, Up*].

Les critères de qualité d'un calage :

- les statistiques des rapports de poids (écart-type, min et max)
- le nombre d'itérations de Calmar
- l'effet sur des variables « de contrôle »

# Premiers scénarios de calage

Scénario	S1	S2	S3	S4	S5	S6
Type de calage	<i>Raking ratio</i>	<i>Raking ratio</i>	<i>Raking ratio</i>	<i>Raking ratio</i>	<i>Raking ratio</i>	<i>Raking ratio</i>
Option TH	1	2	1	1	1	1
Méthode	1	1	2	3	1	1
Modalité de DIP11 non calée	Sans diplôme	Sans diplôme	Sans diplôme	Sans diplôme	Aucune	Sans diplôme
Modalité de CSTOTR non calée	Ouvriers	Ouvriers	Ouvriers	Ouvriers	Ouvriers	Aucune
Nombre d'itérations de Calmar	4	7	4	6	4	23
Rapport de poids minimal	0,21	0,20	0,21	0,10	0,54	0,05
Rapport de poids maximal	3,15	4,73	3,15	3,25	1,78	21,6
Ecart-type des rapports de poids	0,11	0,15	0,11	0,10	0,05	0,17

Référence

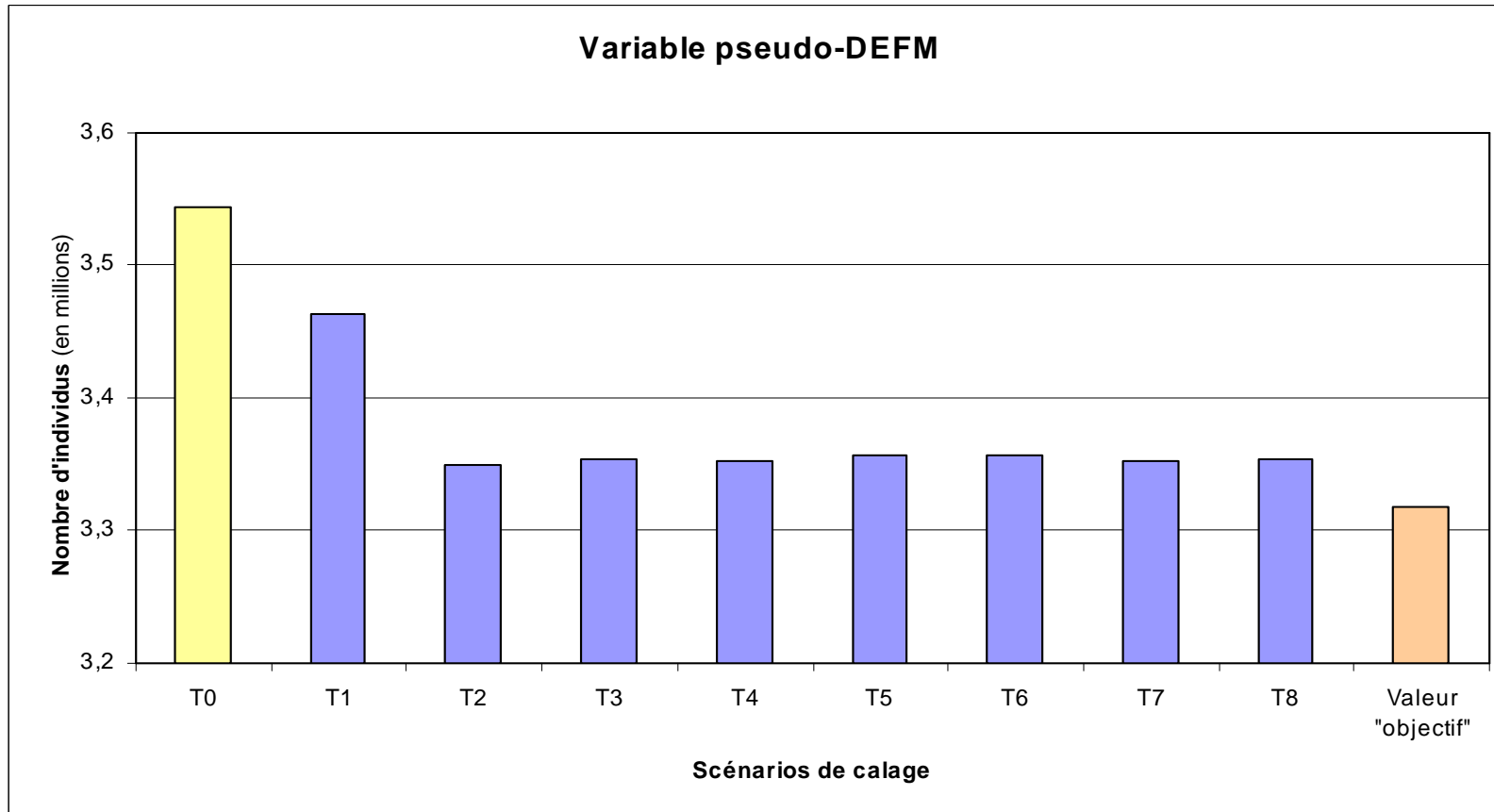
Paramètre différent du scénario 1

# Effet du nombre de contraintes sur les statistiques de calage

Scénario	Cohérence annuelle	Cohérence trimestrielle	Option TH	Nb contraintes	Nb itérations	Écart-type	Rapports de poids			
							Min	D1	D9	Max
T0	Aucun	Variables de calage trimestriel	1	620	1	0	1	1	1	1
T1	+ Statut d'activité x âge x sexe		1	650	4	0,03	0,63	0,98	1,02	1,70
T2	+ Halo		1	651	4	0,07	0,47	0,96	1,05	1,79
T3	+ Halo x âge x sexe		1	660	4	0,07	0,38	0,96	1,04	1,90
T4	+ Diplôme, CSP		1	677	4	0,10	0,41	0,89	1,12	1,85
T5 / S1		+ Statut d'activité x âge x sexe	1	767	4	0,11	0,21	0,88	1,13	3,15
T6		+ Halo x âge x sexe	1	797	5	0,12	0,19	0,87	1,13	3,20
T7		+ Diplôme, CSP	1	848	5	0,14	0,24	0,85	1,15	3,16
T8			2	848	7	0,17	0,23	0,83	1,18	4,82



# Effet du calage sur les variables « de contrôle »

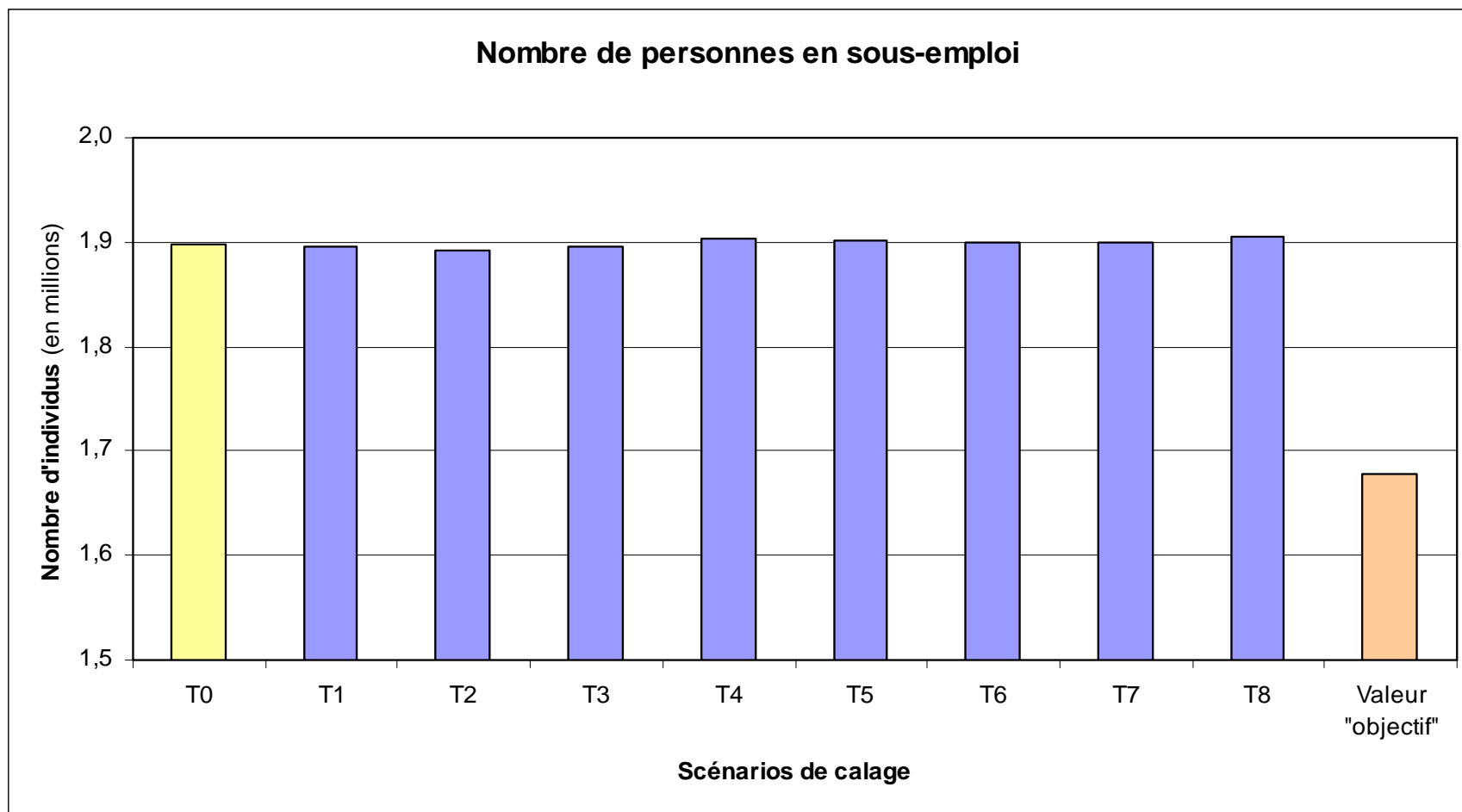


**Pseudo-DEFM** = Personne n'étant pas en emploi au sens du BIT et déclarant – dans l'enquête Emploi – être inscrite auprès de Pôle Emploi, un opérateur de placement ou une association d'insertion

**Valeur objectif** = Résultat en moyenne annuelle sur l'échantillon complet

**T0, T1, ..., T8** = Résultat obtenu sur le sous-échantillon d'intérêt, en fonction du scénario de calage annuel retenu

## Effet du calage sur les variables « de contrôle » (2)



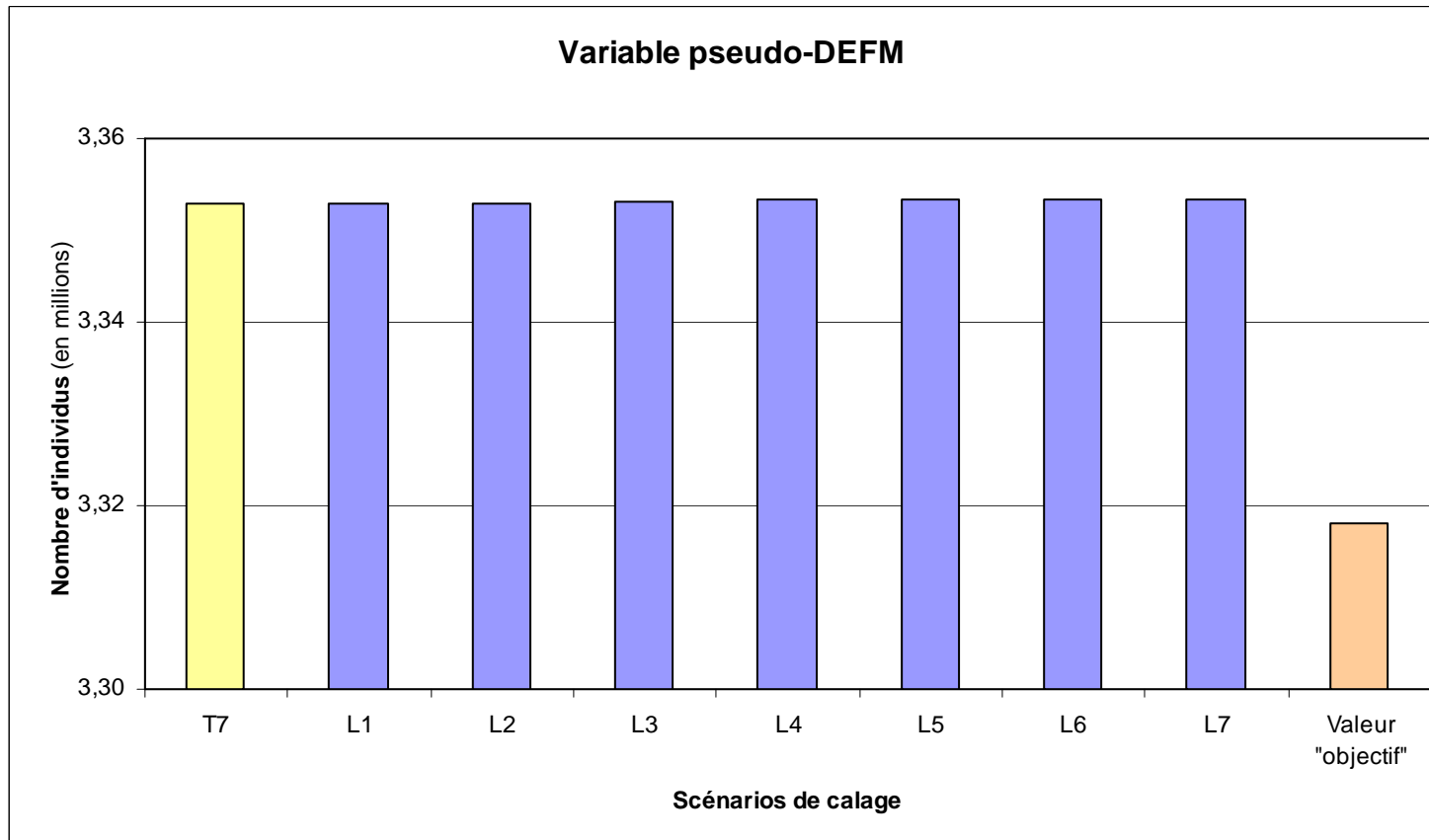
Valeur objectif = Résultat en moyenne annuelle sur l'échantillon complet

T0, T1, ..., T8 = Résultat obtenu sur le sous-échantillon d'intérêt, en fonction du scénario de calage annuel retenu

# Effet de la fenêtre des rapports de poids dans la méthode Logit

Scénario	Méthode	Fenêtre des rapports de poids		Nb itérations	Rapports de poids		
		<i>Low</i>	<i>Up</i>		Écart-type	Min	Max
T7	<i>Raking ratio</i>	sans objet		5	0,1403	0,244	3,158
L1	<i>Logit</i>	0,100	10,0	5	0,1403	0,262	3,122
L2	<i>Logit</i>	0,167	6,0	5	0,1403	0,278	3,085
L3	<i>Logit</i>	0,250	4,0	5	0,1403	0,308	3,012
L4	<i>Logit</i>	0,333	3,0	6	0,1402	0,351	2,883
L5	<i>Logit</i>	0,357	2,8	8	0,1402	0,357	2,800
L6	<i>Logit</i>	0,385	2,6	7	0,1402	0,385	2,600
L7	<i>Logit</i>	0,400	2,5	7	0,1402	0,400	2,500
L8	<i>Logit</i>	0,417	2,4	Ne converge pas			

# Effet de la méthode Logit sur les variables « de contrôle »



**Pseudo-DEFM** = Personne n'étant pas en emploi au sens du BIT et déclarant – dans l'enquête Emploi – être inscrite auprès de Pôle Emploi, un opérateur de placement ou une association d'insertion

**Valeur objectif** = Résultat en moyenne annuelle sur l'échantillon complet

**T7, L1, ..., L7** = Résultat obtenu sur le sous-échantillon d'intérêt, en fonction du scénario de calage annuel retenu