

## L'enquête emploi : échantillon 1992 et années suivantes

*Nicole ROTH*  
*Insee*

L'Enquête Emploi (EE) a pour but essentiel de décrire la situation de l'emploi à un instant donné, d'en suivre l'évolution au cours du temps. La mesure du chômage en particulier, selon les normes du Bureau International du Travail, constitue l'un de ses objectifs principaux. Par certains aspects, l'EE est une actualisation du recensement de la population (RP), permettant de mettre à jour les structures de la population active entre deux RP. Elle s'appuie sur un questionnaire qui, sur certains points, suit celui du RP tout en le complétant sur d'autres aspects.

De ce fait, la méthodologie adoptée pour l'échantillonnage de l'EE est voisine de celle du RP : le sondage aréolaire consiste, en se basant sur un découpage du territoire en petites zones - les aires - à tirer au sort un échantillon de celles-ci. Chaque zone-échantillon a des frontières très soigneusement définies et l'enquête consiste en un recensement de chaque aire tirée au sort. L'enquêteur ne se limite pas à visiter des logements à partir d'une liste pré-établie, mais identifie tous les logements de son aire.

Ainsi, l'échantillon permet notamment de s'affranchir des oublis éventuels du RP, de saisir toutes les situations nouvelles survenues dans l'aire : logements construits depuis le RP et transformations de locaux en logements à usage d'habitation. C'est pourquoi l'EE est parfois qualifiée de "micro-recensement".

La technique de l'échantillon aréolaire permet également d'aboutir à des taux de réponse nettement plus élevés que les enquêtes comparables issues de l'Echantillon Maître.

	Enquêtes acceptées et complètes	Ménages non contactés	Refus
Enquête Emploi 1987	93,8 %	3,5 %	2,7 %
Conjoncture auprès des ménages 1987	86,2 %	6,4 %	7,5 %

L'EE est sujette à un taux de non réponse de l'ordre de 7 %, soit deux fois moins qu'une enquête Echantillon-Maître standard.

Enfin, la technique aréolaire réduit les coûts de collecte sur le terrain. Cet avantage a cependant une contrepartie évidente qui réside dans l'effet de grappe : à taille d'échantillon constante, cet effet réduit la précision des résultats, compte tenu de l'homogénéité naturelle latente à l'intérieur de chaque aire. L'ampleur de cet effet sera précisé ultérieurement sur quelques grandeurs mesurées à l'enquête.

Ce papier a pour objet de décrire les grandes lignes de l'échantillon qui doit être mis en place à partir de 1992 et d'explicitier les raisons qui ont conduit à certains choix.

La stratégie adoptée pour l'échantillon des années 1992 et suivantes reconduit le principe d'un échantillon d'aires déjà utilisé lors des séries précédentes. Elle repose sur une stratification initiale constituée par le croisement région x tranche d'unité urbaine. Le principe d'une rotation sur 3 ans et d'un renouvellement d'un tiers de l'échantillon chaque année est maintenu.

## I - LA STRATIFICATION

Chaque région est découpée en strates selon la taille des unités urbaines (UU) ou le caractère plus ou moins rural des communes. Ce sont essentiellement les concepts du RP qui président à ce découpage puisque c'est le code TU du RP qui est utilisé. Cependant, dans le rural, une typologie un peu plus fine sera utilisée en 1992, discriminant 4 types de communes et isolant notamment le "rural profond" agricole. Cette typologie issue des résultats du RP 82 a été mise en place par le CREDOC [1] à partir des profils socioprofessionnels des communes.

On a comparé l'efficacité de la nouvelle stratification dans le rural à celle de l'ancienne fondée sur le code ACTR du RP (ACTR : "Appartenance à un Canton Totalement Rural") les résultats présentés comparent les effets de sondage selon les deux stratégies de stratification.

Effets de sondage : strates rurales

	Stratification selon ACTR	Stratification selon la typologie CREDOC (4 groupes)
Chômeurs BIT	1,5	1,5
Actifs BIT	2,7	2,6
Agriculteurs exploitants	7,4	5,8
Artisans, commerçants, chef d'entreprise	2,0	2,0
Cadres et prof. intellectuelles supérieures	2,4	2,1
Professions intermédiaires	2,5	2,2
Employés	2,4	2,2
Ouvriers	2,6	2,4

N.B. : Les effets de sondage sont estimés ici sans tenir compte de la stratification régionale.

Dans les strates urbaines, aucune stratification plus fine n'a été utilisée, si ce n'est celle implicite résultant du tirage systématique sur un fichier ordonné (UU ordonnées par département et taille). La raison en tient à l'absence d'une typologie explicative qui soit opérationnelle. En effet, à ce niveau, une typologie de communes est inefficace car le lien entre les caractéristiques de la commune et celles d'une aire contenue dans la commune est d'autant plus faible que la taille des communes augmente.

Même dans les petites unités urbaines ( $\leq 5\ 000$  habitants), le pouvoir explicatif de la typologie de communes s'est avéré trop faible pour apporter une amélioration significative. Il aurait donc fallu disposer d'une typologie infracommunale ; or il n'existe pas de découpage du territoire qui soit compatible avec le principe de délimitation des aires sur le terrain : en effet, il faut être sûr, qu'autour de chaque point tiré (district du RP), on peut découper une zone appartenant à un type donné et ayant la taille requise. Ceci nécessiterait un travail préalable de regroupement des districts hors de portée dans les échéances actuelles.

## II - LA TAILLE DES AIRES

Lors des séries précédentes, la taille des aires était de 40 logements en moyenne au moment du découpage (avec une fourchette allant de 32 à 48). Le tirage et la délimitation des aires ont lieu dès que les données du RP sont disponibles, et ceci pour l'ensemble des sous-échantillons de la série. La phase "terrain" de ce découpage conduit à fixer des limites d'aires qui doivent être claires et non ambiguës, afin d'éviter des risques d'omission ou de surcouverture au moment de l'enquête. Evidemment, à la date de l'enquête, la fourchette de 32-48 peut varier en raison de la construction neuve ou des modifications du parc de logements.

Ce type d'échantillon conduit à un effet de grappe assez important. Comme le montre l'estimation des effets de sondage sur l'enquête de 1989. (voir annexe)

A partir de 1992, pour réduire cet effet de grappe tout en conservant certains avantages de la technique aréolaire (meilleur taux de réponse, aspect micro-recensement), la taille des aires a été réduite dans les unités urbaines les plus importantes ( $\geq 100\ 000$  habitants) : dans ces strates, les aires sont calibrées autour de 20 logements (avec une fourchette allant de 15 à 25).

La logique du calcul aurait évidemment conduit à aller plus loin dans ce sens, réduire encore plus la taille des aires et étendre cette réduction à l'ensemble des strates. En fait, les choix effectués se justifient par les deux raisons suivantes : premièrement, diminuer sensiblement la taille des aires, notamment dans le rural, conduit à une augmentation importante des coûts de collecte si on veut conserver un faible taux de non-réponse. Ceci mettrait également en péril l'un des objectifs de l'enquête qui est de fournir des résultats rapidement (4 semaines d'enquête sur le terrain, premiers résultats 2 mois plus tard).

Par ailleurs, il nous a paru plus difficile de définir des limites d'aires claires lorsque ces aires sont de petite taille ; la multiplication de situations ambiguës sur le terrain risquerait alors de faire perdre en couverture du territoire. De plus, la variabilité relative de la taille des aires risque d'être d'autant plus importante que les aires sont de faible taille, ce qui nuirait à la précision des résultats.

Ces raisons, plutôt qualitatives, ont donc conduit à adopter une position quelque peu conservatrice ; les limites fixées - calibrage des aires à 20 logements dans les unités urbaines de 100 000 habitants et plus - restent un peu arbitraires. Elles s'expliquent notamment par un souci de prudence.

Les résultats attendus de cette réduction de la taille des aires, à taille d'échantillon constante et à stratification constante, sont les suivants :

	Effectifs en milliers	Ecart type en milliers	deft aire 40	deft aire 20
Chômeurs BIT	2 431	42	2,39	1,96
Actifs BIT	23 933	105	3,18	2,66
Actifs occupés BIT	21 502	107	3,29	2,76
Salariés	18 032	114	3,86	3,26
Non salariés	1 553	56	6,62	6,42
Agriculteurs exploitants	1 312	52	6,80	6,60
Artisans, commerçants chefs d'entreprise	1 739	36	2,48	2,26
Cadres et prof. intellectuelles supérieures	2 267	56	4,54	3,20
Professions intermédiaires	4 327	59	2,82	2,35
Employés	5 812	67	2,88	2,36
Ouvriers	6 045	77	3,59	2,97

N.B. : Les estimations de précision figurant dans ce tableau s'appliquent aux données EE88.

Elles ne prennent pas en compte la poststratification selon le sexe et la tranche d'âge quinquennale. Cette poststratification améliore notablement la précision des estimations pour toutes les grandeurs liées à la structure démographique. Ainsi, avec la poststratification l'écart-type des actifs BIT passe de 105 à 61 milliers. .

### III - LA REGIONALISATION DE L'ECHANTILLON

A partir de 1992, l'échantillon des petites régions sera notablement accru afin de répondre aux demandes européennes émanant d'Eurostat. L'échantillon global est accru de 10 % ce qui le porte à un peu moins de

80'000 ménages (résidences principales), et devrait fournir environ 150 000 individus répondants.

Cet échantillon est alloué entre les différentes régions de façon à assurer un seuil minimum de précision dans chaque région : l'échantillon doit permettre d'aboutir à un coefficient de variation  $CV \leq 8 \%$  pour une caractéristique touchant 5 % des forces de travail et sujette à un effet de sondage analogue à celui du chômage. Ceci nécessite d'avoir environ 5 500 enquêtes réalisées dans chaque région (compte-tenu de la stratégie adoptée). Les petites régions ont donc été fixées à ce seuil ; le reste de l'échantillon a été alloué proportionnellement à la taille des autres régions, ce qui conduit à une diminution de 6 % de la taille de l'échantillon dans les régions les plus importantes par rapport à la situation actuelle. Toutefois, par rapport à la série 1982-1989, la précision des résultats sera meilleure en raison de l'amointrissement de l'effet de grappe.

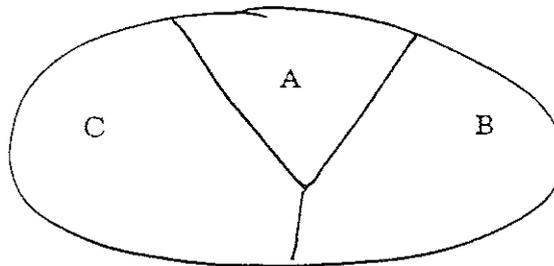
	Allocation	Accroissement p.r. à l'allocation actuelle	Gain en écart type par rapport à la situation actuelle
Ile de France	22.300	- 6 %	0,94
Champagne Ardennes	5 540	+ 76 %	0,68
Picardie	5 540	+ 39 %	0,77
Haute Normandie	5 540	+ 29 %	0,80
Centre	5 540	- 2 %	0,91
Basse Normandie	5 540	+ 44 %	0,76
Bourgogne	5 540	+ 44 %	0,76
Nord-Pas de Calais	9 280	- 6 %	0,94
Lorraine	5 700	- 6 %	0,94
Alsace	5 540	+ 30 %	0,80
Franche-Comté	5 540	+ 116 %	0,62
Pays de Loire	7 100	- 6 %	0,94
Bretagne	6 900	- 6 %	0,94
Poitou Charentes	5 540	+ 45 %	0,75
Aquitaine	6 160	- 6 %	0,94
Midi-Pyrénées	5 770	- 6 %	0,94
Limousin	5 540	+ 200 %	0,52
Rhône -Alpes	11 840	- 6 %	0,94
Auvergne	5 540	+ 69 %	0,70
Languedoc Roussillon	5 540	0 %	0,91
PACAC	10 030	- 6 %	0,94
	<hr/> 151 560	<hr/> + 10 %	<hr/> 0,89

#### IV - LE RENOUVELLEMENT DANS LE TEMPS

A l'issue du RP90, l'ensemble des aires devant servir aux enquêtes Emploi jusqu'au prochain RP sont découpées une fois pour toutes (phase "terrain" en 1991-92). Le plan de rotation de l'échantillon est maintenu : chaque année 1/3 de l'échantillon est renouvelé, de sorte qu'une aire donnée participe 3 années consécutives à l'enquête.

Avec 2/3 d'aires communes pour 2 enquêtes consécutives, la précision d'une variation se trouve sensiblement améliorée par rapport à ce que donneraient 2 échantillons indépendants.

Sur le 1/3 renouvelé, l'aire sortante est remplacée par une aire entrante qui lui est contigüe sur le terrain. En effet, les aires ont été groupées en paquets, à la fois pour faciliter le découpage et pour éviter de trop brusques variations de l'échantillon lors du renouvellement de l'échantillon.



Lorsque l'aire A est sortante, elle est remplacée par B puis par C ; ceci réduit le travail de délimitation des aires : le coût de définition d'une limite commune à 2 aires est divisé par 2. Par ailleurs les aires d'un même groupe présentent certaines similitudes, ce qui réduit la variabilité des estimations lors du renouvellement.

L'amélioration de la précision des estimations de variations grâce au recouvrement partiel des échantillons de deux enquêtes successives repose sur l'existence d'une corrélation positive entre les valeurs observées aux dates  $t$  et  $t - 1$ .

L'importance de cette corrélation dépend de la variable considérée ; elle est évidemment plus élevée sur les aires communes que sur les aires renouvelées.

Coefficient de corrélation entre moyennes par aire aux 2 dates	aires communes	aires renouvelées
chômeurs BIT	0,60	0,25
actifs BIT	0,90	0,40

A partir des totaux calculés sur chaque aire on a estimé la variation sur une aire d'une grandeur  $Y$  entre les dates  $t$  et  $t-1$  (il s'agit d'une vraie variation pour les aires communes aux 2 dates, d'une différence entre aire entrante et aire sortante pour les aires appariées). On a ensuite extrapolé cette variation à l'ensemble de la population. les calculs ont été appliqués soit à  $Y$  directement, soit à un résidu issu d'une linéarisation afin de tenir compte de la post-stratification.

Pour les deux variables, chômage et actifs BIT, les résultats sont les suivants :

	Ecart-type estimé	
	total	variation
Chômeurs BIT	40 000	38 800
Actif BIT	61 000	53 600

L'écart-type de la variation d'une grandeur apparaît donc inférieur à celui de cette même grandeur. Il est toutefois élevé si on le rapporte aux évolutions annuelles.

ANNEXE : LES ERREURS ALÉATOIRES DE L'ENQUÊTE

En milliers ou en % selon les variables

Variable	Valeur de la variable et intervalle de confiance à 95 % (1)			Les deux sexes
	Hommes	Femmes		
<b>1/ CHOMAGE</b>				
Nombre total de chômeurs au sens du BIT.....	982 + ou - 46	1 326 + ou - 48	2 308 + ou - 76	(2,13)
Nombre de chômeurs BIT de 15 à 24 ans.....	249 + ou - 18	343 + ou - 21	592 + ou - 29	(1,24)
Nombre de chômeurs BIT de 25 à 49 ans.....	565 + ou - 34	828 + ou - 36	1 393 + ou - 56	(1,88)
Nombre de chômeurs BIT de 50 ans et plus.....	168 + ou - 15	155 + ou - 14	324 + ou - 21	(1,18)
Taux de chômage BIT.....	7,3 + ou - 0,3	12,6 + ou - 0,5	9,6 + ou - 0,3	(2,35)
Taux de chômage (BIT) des 15-24 ans.....	16,9 + ou - 1,1	24,2 + ou - 1,4	20,5 + ou - 0,9	(1,17)
Taux de chômage (BIT) des 25-49 ans.....	6,1 + ou - 0,4	11,6 + ou - 0,5	8,5 + ou - 0,3	(2,02)
Taux de chômage (BIT) des 50 ans et plus.....	6,1 + ou - 0,5	7,9 + ou - 0,7	6,9 + ou - 0,4	(1,16)
Pourcentage de chômeurs ayant plus d'un an d'ancienneté de chômage.....	40,1 + ou - 2,0	43,4 + ou - 1,7	42,0 + ou - 1,4	(1,22)
Pourcentage de chômeurs de 15 à 24 ans ayant plus d'un an d'ancienneté de chômage.....	18,6 + ou - 2,9	28,0 + ou - 2,8	24,2 + ou - 2,1	(0,97)
Pourcentage de chômeurs de 25 à 49 ans ayant plus d'un an d'ancienneté de chômage.....	41,6 + ou - 2,8	46,4 + ou - 2,2	44,5 + ou - 1,7	(1,20)
Pourcentage de chômeurs de 50 ans et plus, ayant plus d'un an d'ancienneté de chômage.....	64,8 + ou - 4,3	61,4 + ou - 4,6	63,2 + ou - 3,2	(1,04)
<b>2/ POPULATION ACTIVE</b>				
Nombre total d'actifs au sens du BIT.....	13 541 + ou - 59	10 521 + ou - 99	24 062 + ou - 129	(2,35)
Nombre d'actifs BIT de 15 à 24 ans.....	1 473 + ou - 36	1 414 + ou - 35	2 887 + ou - 56	(1,78)
Nombre d'actifs BIT de 25 à 49 ans.....	9 312 + ou - 24	7 150 + ou - 73	16 462 + ou - 79	(1,19)
Nombre d'actifs BIT de 50 ans et plus.....	2 756 + ou - 42	1 957 + ou - 42	4 713 + ou - 68	(1,96)
Taux d'activité (BIT) global.....	64,4 + ou - 0,5	45,8 + ou - 0,6	54,7 + ou - 0,5	(3,24)
Taux d'activité (BIT) des 15-24 ans.....	37,0 + ou - 1,0	35,2 + ou - 1,1	36,1 + ou - 0,9	(1,97)
Taux d'activité (BIT) des 25-49 ans.....	96,3 + ou - 0,2	73,4 + ou - 0,7	84,8 + ou - 0,4	(1,93)
Taux d'activité (BIT) des 50 ans et plus.....	37,4 + ou - 0,8	21,3 + ou - 0,6	28,4 + ou - 0,6	(2,42)

(1) Par exemple, le nombre de chômeurs au sens du BIT mesuré d'après l'enquête est de 2 308 000 ; il a pour intervalle de confiance à 95 % + ou - 76 000 ; l'intervalle 2 232 000 - 2 384 000 a donc 95 chances sur 100 de contenir l'effectif réel des chômeurs BIT dans l'ensemble de la population.

Entre ( .. ) figurent les effets de sondage relatifs aux deux sexes

Variable	Valeur de la variable et intervalle de confiance à 95 %		
	Hommes	Femmes	Les deux sexes
<b>POPULATION ACTIVE (suite)</b>			
Taux d'activité (BIT) des 15-19 ans.....	13,6 + ou - 0,9	9,4 + ou - 0,8	11,5 + ou - 0,6
Taux d'activité (BIT) des 20-24 ans.....	60,1 + ou - 1,5	59,7 + ou - 1,5	59,9 + ou - 1,2
Taux d'activité (BIT) des 50-54 ans.....	90,1 + ou - 1,1	62,2 + ou - 1,5	76,0 + ou - 1,0
Taux d'activité (BIT) des 55-59 ans.....	68,1 + ou - 1,5	44,7 + ou - 1,5	56,1 + ou - 1,2
Taux d'activité (BIT) des 60 ans et plus.....	10,4 + ou - 0,6	5,8 + ou - 0,4	7,7 + ou - 0,4
Taux d'activité (BIT) des femmes de 25 à 29 ans.....	/ /	/ /	/ /
Taux d'activité (BIT) des femmes de 30 à 34 ans.....	/ /	/ /	/ /
Taux d'activité (BIT) des femmes de 35 à 39 ans.....	/ /	/ /	/ /
Taux d'activité (BIT) des femmes de 40 à 44 ans.....	/ /	/ /	/ /
Taux d'activité (BIT) des femmes de 45 à 49 ans.....	/ /	/ /	/ /
<b>3/ POPULATION ACTIVE OCCUPEE (en nombre)</b>			
Actifs occupés au sens du BIT.....	12 559 + ou - 73	9 195 + ou - 107	21 754 + ou - 149
Salariés.....	10 383 + ou - 89	8 015 + ou - 105	18 399 + ou - 157
Non-salariés.....	2 176 + ou - 80	1 179 + ou - 56	3 355 + ou - 126
1. Agriculteurs exploitants.....	795 + ou - 62	466 + ou - 42	1 261 + ou - 100
2. Artisans, commerçants, chefs d'entreprises.....	1 122 + ou - 47	579 + ou - 35	1 701 + ou - 72
3. Cadres et professions intellectuelles supérieures.....	1 611 + ou - 80	640 + ou - 43	2 251 + ou - 111
4. Professions intermédiaires.....	2 533 + ou - 71	1 860 + ou - 59	4 393 + ou - 107
5. Employés.....	1 494 + ou - 71	4 474 + ou - 87	5 968 + ou - 121
6. Ouvriers.....	5 004 + ou - 110	1 176 + ou - 54	6 180 + ou - 141
01. Agriculture.....		1 508 + ou - 107	( 2,74 )
02. Industries agricoles et alimentaires.....		609 + ou - 40	( 2,69 )
03. Energie.....		261 + ou - 30	( 4,22 )
04. Industries des biens intermédiaires.....		1 256 + ou - 71	( 6,54 )
05. Industries des biens d'équipement.....		1 568 + ou - 71	( 2,56 )
06. Industries des biens de consommation.....		1 225 + ou - 62	( 4,76 )
02 à 06. Industrie (en milliers).....		4 920 + ou - 122	( 2,44 )
07. Bâtiment, génie civil et agricole.....		1 636 + ou - 58	( 3,96 )
08. Commerce (en milliers).....		2 564 + ou - 72	( 1,57 )
09. Transports et télécommunications.....		1 318 + ou - 58	( 2,95 )
10. Services marchands.....		4 648 + ou - 101	( 2,89 )
11 à 13. Institutions financières.....		717 + ou - 38	( 2,16 )
14. Services non marchands.....		4 327 + ou - 122	( 1,81 )
08 à 14. Tertiaire.....		13 574 + ou - 184	( 2,39 )

## BIBLIOGRAPHIE

- [1] I. ALDEGHI, T. LAMBERT, N. TABARD : rapport CREDOC octobre 1988 - Ressemblance et diversité sociales des communes françaises
- [2] W.G. COCHRAN : Sampling Techniques Wiley 1963
- [3] N. COEFFIC : Enquête Emploi : Quelques points de méthodologie (note interne n° 676/120 du 1.7.88)
- [4] J. DESABIE : Théorie et pratique des sondages - Dunod 1971
- [5] J.C. DEVILLE, N. ROTH : Economie et Statistique n° 193-194 nov.-déc. 1986 - La précision des enquêtes sur l'emploi
- [6] G. ROY : Courrier des Statistique n° 35 juillet 1985 - Le plan d'échantillonnage des enquêtes sur l'emploi