

L'impact des changements intervenus dans l'enquête Emploi en 2003 sur la qualité de ses résultats

*Dominique GOUX**

INSEE, division Emploi

Cette étude analyse la qualité des informations recueillies dans l'enquête Emploi depuis le 4^{ème} trimestre 2002 et jusqu'au 4^{ème} trimestre 2004. Elle s'intéresse en particulier au mode de collecte, par téléphone ou par visite, et au rang d'échantillon, c'est-à-dire si l'enquête est la première, deuxième, ... ou sixième enquête effectuée dans le même logement. Dans la première partie, on présente la nouvelle enquête Emploi, mise en place au second trimestre 2001, au départ sous forme expérimentale, ensuite et à partir du premier trimestre 2003 sous forme définitive. La seconde partie décrit les variables utilisées pour l'analyse de la qualité des résultats obtenus par l'enquête. Les deux parties suivantes détaillent les résultats fournis par l'enquête sur deux thèmes : l'accès à la formation continue (partie 3) et le chômage (partie 4).

1. La nouvelle enquête Emploi

Le 1^{er} janvier 2003 l'enquête Emploi est devenue trimestrielle avec une collecte « en continu », c'est-à-dire répartie uniformément sur toutes les semaines de chaque trimestre (Givord, 2003). Cette nouvelle enquête se substitue à l'enquête Emploi annuelle, en place depuis 1950. Au cours de cette longue période de 53 ans, l'enquête a eu lieu en général une fois par an, au printemps, et de temps en temps deux fois par an, au printemps et à l'automne (Goux, 2003).

L'enquête Emploi interroge désormais chaque trimestre environ 54 000 logements. Cet échantillon est constitué de 6 sous-échantillons indépendants et de même taille (environ 9 000 logements chacun). Chaque sous-échantillon est interrogé pendant 6 trimestres consécutifs. Un logement échantillonné reste ainsi pendant 6 trimestres dans l'enquête Emploi. Chaque trimestre, l'échantillon de logements interrogés pour la sixième et dernière fois est remplacé par un « nouvel » échantillon de logements, interrogés pour la première fois (pour plus de détails sur l'échantillon, voir Christine, 2002).

* Dominique Goux, Insee, division Emploi, 18, boulevard Adolphe Pinard, 75 675 Paris Cedex 14. e-mail : dominique.goux@insee.fr

2. Description de l'échantillon et des principales variables étudiées

Dans cette étude, on analyse les résultats des enquêtes Emploi réalisées du quatrième trimestre 2002 au quatrième trimestre 2004 inclus. Ce n'est en effet qu'à partir du 4^{ème} trimestre 2002 que tout l'échantillon de l'enquête provient de la nouvelle base de sondage créée à partir du recensement de 1999 d'une part, et que chaque sous-échantillon correspond à des logements dont l'ancienneté de présence dans l'échantillon de l'enquête est la même d'autre part. Ce nouvel échantillon a été introduit, sixième par sixième, à partir du 3^{ème} trimestre 2001. Pour ce trimestre-là, seul l'échantillon « sixième entrant » était constitué de logements tirés dans le recensement de 1999 ; au 4^{ème} trimestre 2002, 2 sixièmes de l'échantillon étaient dans ce cas ; au 1^{er} trimestre 2002, 3 sixièmes, etc. jusqu'à ce que les 6 sixièmes soient issus du recensement au 4^{ème} trimestre 2002.

On distingue les 6 groupes de répondants correspondant aux 6 sous-échantillons indépendants qui constituent l'échantillon de l'enquête. Ces groupes sont particulièrement intéressants à comparer, car chacun est censé être représentatif de la population française, ce qui n'est pas le cas des groupes constitués de personnes ayant déjà été interrogées le même nombre de fois. Le rang d'interrogation individuel n'est égal au rang d'interrogation de l'aire que si la personne habite dans le logement dès la première enquête et accepte de répondre à chaque trimestre (tableau 1).

Tableau 1 : Effectifs enquêtés selon les rangs d'interrogation du logement et de la personne

Rang détaillé d'interrogation de la personne :	Rang d'interrogation du sous-échantillon						Total
	1	2	3	4	5	6	
Première interrogation	103 858	9 884	3 815	3 035	2 315	6 091	128 998
Seconde interrogation	-	81 202	9 381	3 714	2 742	2 273	99 312
Troisième interrogation	-	-	76 022	10 286	3 738	2 895	92 941
Quatrième interrogation	-	-	-	71 148	10 286	4 360	86 035
Cinquième interrogation	-	-	-	-	71 148	11 724	78 819
Sixième interrogation	-	-	-	-	-	78 071	78 071
Ensemble	103 858	91 086	89 218	88 183	86 417	105 414	564 176

Source : enquêtes Emploi, trimestre 4 de 2002, 2003, 2004, Insee.

Champ : : 15 ans et plus (hors personnes non réinterrogées en vagues intermédiaires pour cause de logement comportant uniquement des inactifs de plus de 65 ans). Le fichier complet comporte 636 258 observations, dont 72 082 sont imputées par recopie des réponses du trimestre précédent.

Note : Le rang détaillé d'interrogation de la personne (variable RGI) comptabilise le nombre de réponses à l'enquête pour une même personne. Le rang d'interrogation du sous-échantillon (variable RGA) indique le nombre de vagues d'enquêtes ayant déjà eu lieu dans le même logement.

On distingue aussi 2 groupes de répondants à l'enquête selon le mode de collecte :

- les personnes pour lesquelles l'enquête est réalisée au téléphone ;
- les personnes pour lesquelles l'enquête est réalisée par visite de l'enquêteur.

Les principes de collecte sont les suivants. Toutes les enquêtes ayant lieu dans un logement interrogé pour la première fois (« sixième entrant ») doivent se faire par visite de l'enquêteur. De fait, c'est presque toujours ce qui se passe : seules 0,5% des enquêtes réalisées dans ces logements sont téléphoniques (tableau 2). De même, toutes les enquêtes ayant lieu dans des logements du « sixième entrant », c'est-à-dire dans des logements présents pour la sixième et dernière fois,

doivent se faire par visite. Cette règle semble un peu moins bien respectée que pour les logements du « sixième entrant » : 2,6% des enquêtes de ce type se font de fait par téléphone.

Les enquêtes dans les logements dits « de rang intermédiaire », c'est-à-dire dans des logements interrogés pour la seconde, la troisième, la quatrième ou la cinquième fois, doivent se faire en général par téléphone. Il y a quelques exceptions à cette règle générale : les logements dont le résident ne possède pas le téléphone, ne souhaite pas le communiquer à l'enquêteur, les logements dans lesquels il y a de nombreux habitants et les logements dans lesquels tous les habitants sont nouveaux par rapport au trimestre précédent sont des logements dans lesquels ces enquêtes intermédiaires peuvent se dérouler par visite. En pratique, 90% des enquêtes de rang intermédiaire ont été réalisées par téléphone entre le 4^{ème} trimestre 2002 et la fin 2004.

Tableau 2 : Nombre et taux d'enquêtes réalisées par téléphone

Rang détaillé d'interrogation de la personne :	Rang d'interrogation du sous-échantillon						Total
	1	2	3	4	5	6	
Première interrogation	466 0,5%	620 6,3%	321 8,4%	264 8,7%	231 10,0%	48 0,8%	1 950 1,5
Seconde interrogation	-	76 251 94,4%	8 475 90,8%	3 277 88,5%	2 432 88,9%	69 3,0%	90 504 91,6%
Troisième interrogation	-	-	72 178 95,4%	9 436 92,2%	3 421 91,8%	129 4,5%	85 164 92,1%
Quatrième interrogation	-	-	-	67 543 95,4%	9 764 93,1%	182 4,2%	77 489 90,5%
Cinquième interrogation	-	-	-	-	63 842 95,6%	466 4,0%	64 308 82,0%
Sixième interrogation	-	-	-	-	-	1 874 2,4%	1 874 2,4%
Ensemble	466 0,5%	76 871 84,8%	80 974 91,2%	80 520 91,7%	79 690 92,6%	2 768 2,6%	321 289 57,2%

Source : enquêtes Emploi, trimestre 4 de 2002, 2003, 2004, Insee.

Champ : cf. tableau 1.

Note : cf. tableau 1.

In fine, 57% des enquêtes sont téléphoniques, 43% se font en vis-à-vis. L'effet du téléphone sur les résultats sera appréhendé de deux façons : d'une part en comparant les résultats des enquêtes par visite et ceux des enquêtes par téléphone, d'autre part en comparant les résultats des enquêtes des logements « entrants » et « sortants » à ceux des autres logements. La première méthode est de fait moins bonne, car les échantillons des personnes répondant par téléphone et des personnes répondants par visite ne sont pas, chacun pris séparément, des échantillons représentatifs, et il est difficile de corriger leur non représentativité, celle-ci étant potentiellement liée aux résultats de l'enquête.

On ne s'intéressera pas à distinguer les réponses obtenues en interrogeant la personne concernée par l'enquête de celles obtenues en interrogeant une personne du ménage qui n'est pas celle directement concernée par l'enquête (obtenues via un « proxy »), n'ayant pas assez d'éléments pour comprendre les différences et interpréter les différences observées.

3. L'accès à la formation continue

La proportion d'actifs ayant suivi une formation au cours des 3 mois précédant la date d'enquête est très variable selon le rang de l'enquête et certaines différences sont significatives au seuil de 5% (tableau 3). Les différences se retrouvent quel que soit le trimestre. Toutes les différences ne sont pas significatives : ce sont essentiellement les différences entre le résultat d'une vague intermédiaire, quelle qu'elle soit, et le résultat d'une vague extrême, quelle qu'elle soit, qui divergent.

Tableau 3 : Taux d'accès à la formation continue selon le rang d'interrogation

Rang d'interrogation du logement	Trim 1 2003	Trim 2 2003	Trim 3 2003	Trim 4 2003	Trim 1 2004	Trim 2 2004	Trim 3 2004	Trim 4 2004
Première interrogation	12,03 (0,36)	12,51 (0,39)	9,68 (0,34)	12,46 (0,36)	13,06 (0,36)	11,86 (0,36)	9,06 (0,33)	10,53 (0,33)
Interrogation intermédiaire	9,85 (0,19)	9,53 (0,19)	6,34 (0,16)	9,75 (0,19)	9,89 (0,19)	10,30 (0,19)	7,06 (0,16)	9,88 (0,19)
<i>dont : 2^{ème} interrogation</i>	<i>10,48 (0,35)</i>	<i>9,56 (0,34)</i>	<i>6,51 (0,31)</i>	<i>10,11 (0,36)</i>	<i>10,55 (0,35)</i>	<i>10,59 (0,35)</i>	<i>7,39 (0,31)</i>	<i>10,08 (0,36)</i>
<i>3^{ème} interrogation</i>	<i>10,58 (0,37)</i>	<i>9,96 (0,36)</i>	<i>6,31 (0,30)</i>	<i>9,73 (0,37)</i>	<i>10,27 (0,37)</i>	<i>10,80 (0,37)</i>	<i>7,15 (0,31)</i>	<i>10,20 (0,37)</i>
<i>4^{ème} interrogation</i>	<i>9,77 (0,38)</i>	<i>9,86 (0,38)</i>	<i>6,56 (0,31)</i>	<i>9,83 (0,37)</i>	<i>9,61 (0,38)</i>	<i>10,50 (0,39)</i>	<i>6,95 (0,32)</i>	<i>10,20 (0,38)</i>
<i>5^{ème} interrogation</i>	<i>9,16 (0,39)</i>	<i>9,35 (0,39)</i>	<i>6,96 (0,34)</i>	<i>10,23 (0,40)</i>	<i>9,56 (0,39)</i>	<i>9,72 (0,41)</i>	<i>7,59 (0,36)</i>	<i>9,45 (0,39)</i>
Dernière interrogation	14,32 (0,46)	13,30 (0,44)	9,89 (0,39)	14,44 (0,45)	11,84 (0,42)	11,77 (0,42)	9,03 (0,38)	11,75 (0,42)
Ensemble	10,97 (0,16)	10,69 (0,16)	7,57 (0,14)	11,08 (0,16)	10,92 (0,16)	10,84 (0,16)	7,77 (0,14)	10,31 (0,15)
Taille de l'échantillon	38 455	37 793	36 758	38 307	38 882	38 579	37 724	38 815

Champ : actifs au sens du BIT.

Note : données brutes non pondérées. Au total en 2003, 40 personnes n'ont pas répondu à la question parmi les 151 353 interrogées. Les écarts-types sont indiqués entre parenthèses (calculés comme pour un sondage aléatoire simple).

Source : enquêtes Emploi, 2003, 2004, Insee.

Ces différences peuvent s'expliquer de plusieurs façons : d'une part le mode de collecte est différent selon les rangs d'interrogation, téléphone pour les enquêtes intermédiaires et visite pour les première et dernière enquêtes. D'autre part, le fait d'avoir suivi une formation (ou pas) le trimestre précédent peut influencer deux comportements : le comportement d'acceptation de l'enquête (plus ou moins de refus, en rapport avec le fait d'avoir suivi une formation) et, en cas de réponse à l'enquête, la réponse à la question sur la formation. Il est possible de tester ces hypothèses. Si l'on regarde la proportion de personnes que l'on ne retrouve pas dans l'échantillon d'un trimestre sur l'autre, toutes raisons confondues, on n'observe pas de différence significative entre celles qui ont suivi une formation et celles qui n'en ont pas suivi (tableau 4). L'hypothèse que le comportement de réponse à l'enquête dépendrait du fait d'avoir suivi une formation est à écarter. Il est plus difficile de tester l'hypothèse complémentaire sur le lien entre la réponse à la question sur la formation à deux dates consécutives. Cependant, si seule cette hypothèse expliquait les différences de résultat selon les vagues, on ne devrait pas observer le même lien entre les enquêtes de rang 1 et 2 qu'entre les enquêtes de rang 3 et 4 ou les enquêtes de rang 5 et 6. Tel n'est pas le cas (tableau 3 à nouveau).

Tableau 4 : Proportion de personnes absentes ou non retrouvées d'un trimestre sur l'autre

Situation initiale	Taux de non retrouvés au ...						
	2 ^{ème} trim 2003	3 ^{ème} trim 2003	4 ^{ème} trim 2003	1 ^{er} trim 2004	2 ^{ème} trim 2004	3 ^{ème} trim 2004	4 ^{ème} trim 2004
Suivi d'une formation							
Inconnu	5 sur 7	7 sur 10	4 sur 11	3 sur 7	7 sur 8	6 sur 8	1 sur 4
Oui	7,2	10,0	7,0	6,5	7,1	10,6	7,4
Non	7,2	10,2	7,1	6,4	7,0	10,7	7,9
Taille de l'échantillon	32 291	31 602	30 692	31 688	32 450	32 258	31 647

Champ : personnes actives de 15 ans ou plus, hors aires du sixième sortant des 4 trimestres 2003.

Note : données brutes non pondérées. Parmi les personnes actives occupées au premier trimestre 2003, 6,8% ne sont pas répondantes au 2^{ème} trimestre 2003, contre 10,9% des personnes au chômage au premier trimestre 2003.

Source : enquêtes Emploi, 2003, 2004, Insee.

Le tableau 5 met en évidence les différences de taux de formation selon qu'il est estimé sur la population répondant par téléphone ou par visite. Ces deux sous-échantillons ne sont cependant pas tout à fait représentatifs de la population métropolitaine puisque le mode de collecte dépend de la taille du ménage et du fait qu'il possède ou non le téléphone ou souhaite le communiquer ou pas (cf. infra).

Tableau 5 : Formation continue selon le mode de collecte

Mode de collecte	4 ^{ème}	1 ^{er}	2 ^{ème}	3 ^{ème}	4 ^{ème}	1 ^{er}	2 ^{ème}	3 ^{ème}	4 ^{ème}
	trim 2002	trim 2003	trim 2003	trim 2003	trim 2003	trim 2004	trim 2004	trim 2004	trim 2004
Interrogation téléphonique	8,27 (0,18)	9,99 (0,19)	9,63 (0,19)	6,43 (0,16)	9,90 (0,20)	10,06 (0,20)	10,39 (0,20)	7,10 (0,17)	10,01 (0,20)
Interrogation en vis-à-vis	9,58 (0,24)	12,56 (0,27)	12,51 (0,28)	9,43 (0,25)	12,85 (0,27)	12,21 (0,26)	11,57 (0,28)	8,89 (0,24)	10,80 (0,25)
<i>Ecart vis-à-vis / téléphone</i>	<i>+1,31 (0,30)</i>	<i>+2,57 (0,34)</i>	<i>+2,88 (0,34)</i>	<i>+3,00 (0,30)</i>	<i>+2,95 (0,33)</i>	<i>+2,15 (0,33)</i>	<i>+1,18 (0,33)</i>	<i>+1,79 (0,29)</i>	<i>+0,79 (0,32)</i>
Taille de l'échantillon total	37 798	38 453	37 792	36 758	38 306	38 881	38 572	37 718	38 808

Champ : actifs au sens du BIT.

Note : données brutes non pondérées. Au total en 2003, 40 personnes n'ont pas répondu à la question parmi les 151 353 interrogées et 4 ont un mode de collecte inconnu. Les écarts-types sont indiqués entre parenthèses (calculés comme pour un sondage aléatoire simple).

Source : enquêtes Emploi, trimestre 4 de 2002, 2003, 2004, Insee.

Les différences entre enquête téléphonique et enquête par visite sont particulièrement élevées en 2003. Elles ont augmenté brutalement entre fin 2002 et début 2003, baissé de façon importante entre fin 2003 et début 2004 et semblent continuer de diminuer tout au long de 2004. Il faut y voir un lien avec l'existence, en 2003, d'un module complémentaire sur la formation continue. Ce module faisait partie intégrante de l'enquête Emploi, mais venait à la suite des questions classiques sur la formation. Aucune des questions de l'enquête Emploi n'a été modifiée, le questionnaire a simplement été allongé, et les enquêteurs ont suivi une formation pour ce complément. Malgré toutes les précautions prises pour ne pas modifier les résultats de l'enquête Emploi, l'existence du module ad hoc a conduit à estimer un taux de formation beaucoup plus élevé qu'à l'habitude. On peut raisonnablement penser que cela est dû à ce que les enquêteurs ont été sensibilisés, en 2003, aux questions concernant la formation tout au long de la vie. Cela souligne l'importance de la formation des enquêteurs, mais aussi la sensibilité des résultats d'une enquête à cette formation.

Au total, on peut conclure à l'existence d'un effet « téléphone » sur la mesure de l'accès à la formation, les réponses étant significativement différentes suivi que ce mode de collecte est utilisé ou pas. L'analyse du cas de la formation en continue souligne par ailleurs l'importance de la formation des enquêteurs.

4. Le « biais de rotation »

Si l'on s'intéresse maintenant au nombre des chômeurs ou au taux de chômage, la différence majeure par rapport à l'analyse sur la formation continue tient à ce que le taux de non-réponse à l'enquête, toutes causes confondues, est lié à la situation de chômage ou non chômage le trimestre précédent.

Les personnes repérées comme au chômage au sens du BIT un trimestre donné ont une probabilité significativement plus élevée de ne pas être enquêtées le trimestre suivant (tableau 6). C'est le cas d'en moyenne 11% d'entre elles, contre seulement 8% des personnes ayant un emploi et 6% des personnes inactives. Ce phénomène s'observait déjà dans les enquêtes Emploi des années 1976-1981 (voir tableau 6, note n° 2671/432). Dans les enquêtes Emploi annuelles récentes, c'était également le cas. Par exemple, parmi les personnes actives occupées en mars 2001 et appartenant aux sous échantillons réinterrogés en mars 2002, 18,1% ne sont pas retrouvées en mars 2002, toutes causes confondues. C'est le cas de 27,5% des chômeurs de mars 2001 et de 13,8% des inactifs de mars 2001 (voir tableau A1 de l'annexe).

Tableau 6 : Proportion de personnes absentes ou non retrouvées d'un trimestre sur l'autre selon l'activité précédente

Situation initiale	Taux de non retrouvés au ...						
	2 ^{ème} trimestre 2003	3 ^{ème} trimestre 2003	4 ^{ème} trimestre 2003	1 ^{er} trimestre 2004	2 ^{ème} trimestre 2004	3 ^{ème} trimestre 2004	4 ^{ème} trimestre 2004
Type d'activité BIT							
Actif occupé	6,8	9,8	6,7	6,1			
Chômeur	10,9	13,7	10,4	9,7			
Inactif	5,1	7,3	4,5	4,1			
Taille de l'échantillon	60 080	59 056	56 874	58 616			

Champ : personnes de 15 ans ou plus, hors aires du sixième sortant.

Note : données brutes non pondérées. Parmi les personnes actives occupées au premier trimestre 2003, 6,8% ne sont pas répondantes au 2^{ème} trimestre 2003, contre 10,9% des personnes au chômage au premier trimestre 2003.

Source : enquêtes Emploi, 2003, 2004, Insee.

Si l'on analyse les causes de la non-réponse à l'enquête suivante, on constate que plusieurs raisons expliquent les différences de taux de réponse selon le type d'activité. D'une part, le taux de déménagement d'un trimestre sur l'autre est plus de 50% plus élevé pour les chômeurs que pour les autres actifs (tableau 7), ce qui empêche leur réinterrogation. D'autre part, le taux de refus-absence de longue durée est aussi de 50% plus élevé parmi les personnes enregistrées au chômage le trimestre précédent que parmi les autres actifs. La première cause de non réinterrogation n'est pas source de biais pour l'enquête. Les personnes qui déménagent sont remplacées par d'autres, les chômeurs qui partent sont remplacés par des chômeurs qui arrivent. En revanche, le surcroît de refus et absences de longue durée observé parmi les chômeurs est potentiellement source de biais. Ces personnes ne sont pas remplacées par d'autres (qui leur ressembleraient en moyenne) dans l'échantillon de l'enquête.

Tableau 7 : Causes de non-réponse selon le statut d'activité

	Actif occupé	Chômeur	Inactif	Ensemble
Déménagement	2,81	4,46	1,92	2,48
Absence	3,64	5,13	2,37	3,13
Refus	0,49	1,03	0,53	0,54
<i>Total</i>	7,03	10,75	4,96	6,27

Note : données non pondérées.

Lecture : 2,8% des actifs occupés avait déménagé d'un trimestre à l'autre.

Source : enquêtes Emploi, 4^{ème} trimestre 2002 au 1^{er} trimestre 2004, Insee.

Comme pour la formation continue, comparons les résultats obtenus sur les 6 sous-échantillons indépendants constituant l'échantillon total de l'enquête. En général, le taux de chômage est plus élevé sur le sous-échantillon interrogé pour la sixième et dernière fois que pour l'échantillon interrogé pour la cinquième fois (tableau 8).

En général, le taux de chômage a tendance à diminuer avec le rang d'interrogation à l'exception de la dernière vague. Le fait que le taux de chômage soit plus élevé en 6^{ème} vague qu'en 5^{ème}, alors que de façon assez systématique le taux de chômage est plus faible sur le $n^{\text{ème}}$ vague que sur la $n-1^{\text{ème}}$, peut s'interpréter comme un « effet téléphone » (tableau 9). Cet effet téléphone, sur le taux de chômage, serait ainsi estimé à -0,23 point. Les sous-échantillons enquêtés par téléphone représentant 4/6^{ème} de l'échantillon, l'effet moyen du téléphone sur une enquête serait de -0,16 point.

L'enquête « Protocole », réalisée auprès de 2 5000 personnes en juin 1994, concluait cependant à l'absence d'effet « téléphone » sur la mesure de l'activité au sens du BIT. Des différences étaient bien sûr observées selon le mode de collecte, mais ces différences n'étaient pas significatives au seuil de 5% (Lagarenne et Schuhl, 1995), probablement à cause de la petite taille de l'échantillon. De fait, l'essentiel des divergences entre les deux modes de collecte se concentrait sur les questions permettant de repérer les chômeurs (partie Q du questionnaire de l'époque).

Tableau 8 : Taux de chômage selon le rang d'interrogation du sous-échantillon

Rang du sous-échantillon	T4 2002	T1 2003	T2 2003	T3 2003	T4 2003	T1 2004	T2 2004	T3 2004	T4 2004
Premier sixième	10,31 (0,38)	10,32 (0,38)	10,52 (0,39)	10,20 (0,39)	11,18 (0,40)	10,37 (0,37)	9,90 (0,37)	9,53 (0,37)	11,80 (0,40)
Second sixième	9,53 (0,36)	10,20 (0,38)	9,62 (0,37)	10,98 (0,40)	10,39 (0,38)	10,64 (0,38)	9,81 (0,36)	9,96 (0,37)	10,19 (0,37)
Troisième sixième	9,69 (0,37)	9,62 (0,36)	9,07 (0,36)	9,28 (0,37)	10,83 (0,39)	10,61 (0,38)	9,62 (0,37)	9,81 (0,37)	10,01 (0,37)
Quatrième sixième	9,44 (0,37)	9,80 (0,37)	8,75 (0,35)	9,18 (0,37)	9,82 (0,37)	11,36 (0,40)	9,21 (0,36)	10,03 (0,38)	9,14 (0,36)
Cinquième sixième	9,10 (0,37)	9,29 (0,37)	9,18 (0,37)	9,14 (0,36)	9,38 (0,37)	9,57 (0,37)	10,08 (0,38)	10,15 (0,38)	10,38 (0,39)
Sixième sixième	8,65 (0,34)	10,01 (0,38)	9,40 (0,37)	9,79 (0,38)	10,55 (0,38)	10,16 (0,38)	8,77 (0,36)	10,42 (0,39)	10,52 (0,38)
<i>Diff. dernier / premier</i>	<i>-1,66</i> <i>(0,51)</i>	<i>-0,31</i> <i>(0,54)</i>	<i>-1,12</i> <i>(0,54)</i>	<i>-0,41</i> <i>(0,55)</i>	<i>-0,63</i> <i>(0,55)</i>	<i>-0,21</i> <i>(0,53)</i>	<i>-1,13</i> <i>(0,51)</i>	<i>+0,89</i> <i>(0,54)</i>	<i>-1,28</i> <i>(0,55)</i>
<i>Diff. 5^{ème} / 2^{ème}</i>	<i>-0,43</i> <i>(0,51)</i>	<i>-0,91</i> <i>(0,53)</i>	<i>-0,44</i> <i>(0,52)</i>	<i>-1,84</i> <i>(0,54)</i>	<i>-1,01</i> <i>(0,53)</i>	<i>-1,07</i> <i>(0,53)</i>	<i>+0,27</i> <i>(0,53)</i>	<i>+0,19</i> <i>(0,53)</i>	<i>+0,19</i> <i>(0,54)</i>
<i>Diff. 6^{ème} / 5^{ème}</i>	<i>-0,45</i> <i>(0,50)</i>	<i>0,72</i> <i>(0,53)</i>	<i>0,22</i> <i>(0,52)</i>	<i>0,65</i> <i>(0,53)</i>	<i>1,17</i> <i>(0,53)</i>	<i>0,58</i> <i>(0,53)</i>	<i>-1,31</i> <i>(0,52)</i>	<i>0,27</i> <i>(0,55)</i>	<i>0,14</i> <i>(0,54)</i>

Champ : actifs au sens du BIT.

Note : données non pondérées.

Lecture : au quatrième trimestre 2002, parmi les actifs interrogés dans les aires entrantes, 10,31% sont au chômage, contre 8,65 % parmi les personnes interrogées dans les aires sortantes.

Source : enquêtes Emploi, trimestre 4 de 2002, 2003, 2004, Insee.

Tableau 9 : Effets moyens du nombre passé d'enquêtes dans le même logement

	Rang d'interrogation du sous-échantillon					
	1 ^{ère}	2 ^{ème}	3 ^{ème}	4 ^{ème}	5 ^{ème}	6 ^{ème}
Sur le taux de chômage :						
Enquête en continu	Ref	-0,41	-0,79	-0,92	-1,14	-0,91
Enquête en continu		ref	-0,38	-0,51	-0,73	
Enquête annuelle	Ref	-0,28	-0,38	-	-	-

Source : enquêtes Emploi en continu, trimestre 4 de 2002 à trimestre 4 de 2004, Insee, et enquêtes Emploi annuelles, 1990 à 2002, Insee.

Les différences selon le rang d'interrogation ne sont clairement pas simplement dues au téléphone. Si tel était le cas, il n'y aurait pas de différence entre les estimations issues des 4 vagues intermédiaires d'une part, et on n'observerait aucune différence entre sous-échantillons dans l'enquête Emploi annuelle, celle-ci se déroulant exclusivement par visite. Ce mode de collecte n'explique donc pas tout. Le fait que les résultats soient significativement différents entre les sous-échantillons est repéré dans la littérature académique sous le nom de « biais de rotation » ou de « d'effet du temps passé dans l'échantillon » (*rotation bias, time-in-sample effect*). Bailar a construit un indice permettant de quantifier le biais de rotation (Bailar, 1975) :

$$B_{i,t} = 100 \times Y_{i,t} / \left(\frac{1}{N} \sum_{j=1}^N Y_{j,t} \right)$$

$$I_{Y(n)} = 100 \times Y(n) / \left(\frac{1}{N} \sum_{i=1}^N Y(i) \right)$$

où $Y_{i,t}$ est l'estimation de Y à la date t et sur le sous-échantillon i .

En l'absence de biais de rotation, les $B_{i,t}$ sont tous égaux à 100. Si $B_{i,t}$ est supérieur à 100, cela indique qu' $Y_{i,t}$ est plus élevé sur le sous-échantillon i qu'en moyenne sur tout l'échantillon. On a calculé les indices de Bailar, corrigés des biais d'échantillonnage sur le taux de chômage (tableau 10) et sur le nombre de chômeurs (tableau 11) pour les enquêtes Emploi du 4^{ème} trimestre 2002 au 4^{ème} trimestre 2004. Ces calculs confirment l'existence d'un biais de rotation. Ce type de biais a déjà été repéré dans des enquêtes étrangères, équivalentes à l'enquête Emploi. Le biais observé dans l'enquête française est de moindre ampleur que dans le CPS américain actuel (Mansur et Shoemaker, 1999 et voir aussi tableau A5 de l'annexe). Il serait légèrement supérieur à celui observé au Canada, d'après les calculs de Ghangurde (1982). Il existe quelques pays où un tel biais semble être absent : c'est le cas de l'Espagne, tout au moins pour les estimations nationales (Bardaji, 2004 et tableau A6 de l'annexe). En Espagne, l'absence de biais de rotation sur le chômage national semble en effet être la résultante de biais de sens variable selon les régions (Bardaji, 2004).

Tableau 10 : Indice de Bailar « corrigé » sur le taux de chômage dans l'enquête Emploi

	Rang d'interrogation du sous-échantillon					
	1	2	3	4	5	6
Trim 4 2004	107,2	103,8	99,2	99,8	95,8	94,2
Trim 1 2003	107,3	101,0	99,6	96,9	94,9	100,3
Trim 2 2003	112,8	104,2	93,8	95,6	93,2	100,8
Trim 3 2003	104,1	112,2	99,1	92,4	95,4	97,0
Trim 4 2003	110,5	98,8	104,4	97,0	90,2	99,5
Trim 1 2004	98,4	102,6	99,4	111,1	94,0	94,9
Trim 2 2004	103,5	101,4	102,3	95,6	105,7	91,7
Trim 3 2004	97,8	99,1	94,4	101,9	102,3	104,7
Trim 4 2004	115,1	100,1	97,8	85,8	100,1	101,1
moyenne	106,3	102,5	98,9	97,3	96,8	98,3

Source : enquêtes Emploi, trimestre 4 2002 à trimestre 4 2004, Insee.

Note : Données pondérées (corrigées des biais d'échantillonnage).

Tableau 11 : Indice de Bailar « corrigé » sur le nombre de chômeurs dans l'enquête Emploi

	Rang d'interrogation du sous-échantillon					
	1	2	3	4	5	6
Trim 4 2002	107,5	103,8	99,2	100,3	95,1	94,1
Trim 1 2003	106,8	101,6	99,8	96,3	95,3	100,1
Trim 2 2003	110,8	104,7	94,4	96,0	92,8	101,5
Trim 3 2003	104,1	110,9	99,7	93,1	96,0	96,1
Trim 4 2003	109,3	99,6	103,1	97,3	91,1	99,6
Trim 1 2004	99,0	101,0	100,5	109,8	94,0	95,7
Trim 2 2004	102,8	103,0	101,2	96,7	104,6	91,7
Trim 3 2004	98,6	99,0	95,6	101,3	102,9	102,7
Trim 4 2004	115,2	101,2	97,3	86,3	99,3	100,7
moyenne	106,0	102,7	99,0	97,5	96,8	98,1

Source : enquêtes Emploi, trimestre 4 2002 à trimestre 4 2004, Insee.

Note : Données pondérées (corrigées des biais d'échantillonnage).

La comparaison des tableaux 10 et 11 montre que le biais de rotation est plus important sur le taux de chômage que sur le nombre de chômeurs. Cela est probablement dû à ce que le biais du taux de chômage a deux origines : biais dans l'estimation du nombre de chômeurs cumulé avec le biais dans l'estimation du nombre d'actifs occupés.

L'indice de Bailar est critiqué pour sa sensibilité aux effets d'échantillonnage. Tout ou partie des différences observées entre vagues peut en effet être due au fait que l'échantillon d'une vague est petit. Cette critique tombe si l'on considère non pas les indices de Bailar trimestre par trimestre, mais la moyenne de ces indices sur un grand nombre de trimestres. Les tableaux 10 et 11 fournissent aussi cette moyenne, qui confirme l'existence du biais de rotation dans l'enquête française. Pour autant, et afin d'affirmer le diagnostic, on a aussi calculé l'indice proposé par Brisebois et Mantel (1999), indice s'affranchissant des effets d'échantillonnage (il compare les résultats sur le même sous-échantillon, au fil du temps) et des effets de saisonnalité et tendance (il compare les résultats à la même date, sur des sous-échantillons différents). L'indice est construit ainsi :

$$BM_{i,t} = 100 \times Y_{i,t}^* / \left(\frac{1}{N} \sum_{j=1}^N Y_{j,t-i+j}^* \right)$$

$$\text{avec } Y_{i,t}^* = Y_{i,t} - \left(\bar{Y}_t - \frac{1}{N} \sum_{j=1}^N \bar{Y}_{t-i+j} \right)$$

Cet indice nécessite de nombreuses données pour pouvoir être calculé. Il faut disposer d'au moins autant de dates de collecte qu'il y a de sous-échantillons dans l'échantillon total pour calculer un seul indice. Disposant de 9 trimestres d'enquêtes Emploi, il est possible de calculer 4 indices de Brisebois et Mantel par rang d'interrogation (tableau 12).

Tableau 12 : Indice de Brisebois et Mantel sur le taux de chômage dans l'enquête Emploi

	Rang d'interrogation du sous-échantillon					
	1	2	3	4	5	6
Trim 4 2004	110,8					
Trim 1 2003	108,6	104,7				
Trim 2 2003	103,4	105,2	97,6			
Trim 3 2003	103,7	103,3	100,3	95,9		
Trim 4 2003		98,5	96,5	98,0	93,0	
Trim 1 2004			99,2	103,0	94,8	98,0
Trim 2 2004				95,6	97,3	93,1
Trim 3 2004					102,0	96,5
Trim 4 2004						100,9
moyenne	106,6	102,9	98,4	98,4	96,8	97,1

Source : enquêtes Emploi, trimestre 4 2002 à trimestre 4 2004, Insee.

Note : Données pondérées (corrigées des biais d'échantillonnage). Les indices ont des valeurs quasiment identiques à celles ci-dessus quand on ne pondère pas.

Les indices de Brisebois et Mantel confirment l'existence du biais de rotation sur le chômage. Quel que soit l'indice, en comparant les écarts moyens entre les deux derniers sous-échantillons (considérés comme une mesure de l'effet téléphone) et les autres écarts entre sous-échantillon (considérés comme une mesure du « biais de rotation », il semble que l'effet « téléphone » soit du second ordre par rapport au « biais de rotation ».

Bibliographie

- [1] Bailer Barbara, 1975, « The effects of rotation group bias on estimates from panel surveys », *Journal of American Statistical Association*.
- [2] Bardaji José, 2004, « Le biais de rotation dans l'enquête EPA », Rapport de stage ENSAE.
- [3] Brisebois François et Mantel Harold, 1996, « Month-in-sample effects for the canadian labour force survey », *Proceedings of the survey research methods section, SSC Annual meeting*.
- [4] Christine Marc, 2002, « La construction de l'échantillon de la future enquête Emploi en continu à partir du recensement de 1999 », in Actes des Journées de Méthodologie Statistique, Insee Méthodes, n°100, pp 175-229.
- [5] Ghangurde P.D., 1982, « Rotation Group Bias in the LFS estimates », Proceedings of the Survey Research Methods Section, *American Statistical Association*.
- [6] Givord Pauline, 2003, « Une nouvelle enquête Emploi », *Economie et Statistique*, n° 362.
- [7] Goux Dominique, 2003, « Une histoire de l'enquête Emploi », *Economie et Statistique*, n° 362.
- [8] Lagarenne Christine et Schuhl Pierrette, 1995, « Contrôle de la qualité de l'enquête trimestrielle Emploi. Résultats de l'enquête Protocole », note interne n° 32/F232.
- [9] Mansur K. et Shoemaker H., 1999, « The impact of changes in the Current Population Survey on time-in-sample bias and correlations between rotation groups », *Proceedings of the survey research methods section, American Statistical Association*.

Annexe : Eléments sur le biais de rotation dans les enquêtes Emploi 1990-2002

Tableau A1 : Proportion de personnes non retrouvées un an plus tard

a) dans l'enquête Emploi annuelle

Année d'enquête	Situation la première année			Taille de l'échantillon
	actif occupé	chômeur	inactif	
1990	16,5	25,6	11,8	90 080
1991	14,8	22,8	11,5	91 039
1992	15,1	23,5	11,3	95 519
1993	15,3	23,9	12,0	100 996
1994	15,6	23,9	12,3	101 731
1995	15,3	23,5	12,9	100 309
1996	16,4	25,7	12,6	100 069
1997	16,9	26,8	12,5	99 984
1998	16,0	24,5	12,0	99 144
1999	19,2	27,5	14,3	99 082
2000	19,0	27,2	13,7	99 868
2001	18,1	27,5	13,8	97 457

Champ : 15 ans et plus des tiers entrant et médian.

Note : données pondérées. Parmi les personnes au chômage en mars 1991 et appartenant aux tiers entrant ou médian de l'enquête, 22,8% n'ont pas été enquêtées en mars 1992 (toutes causes confondues).

Source : enquêtes annuelles Emploi, 1990 à 2002, Insee.

b) dans l'enquête Emploi trimestrielle

Date d'enquête	Situation la première année			Taille de l'échantillon
	actif occupé	chômeur	inactif	
1 ^{er} trimestre 2003	16,2	23,6	15,3	20 045
dont 1 ^{er} sixième	16,2	25,8	16,1	9 918
dont 2 ^{ème} sixième	16,2	21,5	14,7	10 127

Champ : 15 ans et plus des premier et second sixièmes.

Note : données non pondérées. Parmi les personnes au chômage au premier trimestre 2003 et appartenant aux premier ou second sixièmes de l'enquête, 23,6% n'ont pas été enquêtées au premier trimestre 2004 (toutes causes confondues).

Source : enquêtes Emploi, 1^{er} trimestre, 2003, 2004, Insee.

**Tableau A2 : Taux de chômage selon le rang de l'échantillon
données non pondérées**

Année d'enquête	Sous-échantillon			<i>Ensemble</i>	différence sortant / entrant
	entrant	Médian	sortant		
1990	9,55	9,11	8,86	<i>9,17</i>	-0,69
1991	9,27	9,29	8,61	<i>9,06</i>	-0,66
1992	9,98	9,97	10,20	<i>10,05</i>	+0,22
1993	11,11	11,22	11,07	<i>11,13</i>	-0,04
1994	12,84	11,86	12,47	<i>12,39</i>	-0,37
1995	12,04	11,74	11,04	<i>11,61</i>	-1,00
1996	12,18	12,02	12,12	<i>12,11</i>	-0,06
1997	12,79	12,20	11,97	<i>12,32</i>	-0,82
1998	12,00	11,81	11,72	<i>11,84</i>	-0,28
1999	12,00	11,60	11,72	<i>11,78</i>	-0,28
2000	10,61	9,91	9,62	<i>10,04</i>	-0,99
2001	8,74	8,82	8,77	<i>8,78</i>	-0,04
2002	8,92	8,89	8,91	<i>8,91</i>	-0,01

Champ : actifs au sens du BIT.

Note : données pondérées.

Source : enquêtes annuelles Emploi, 1990 à 2002, Insee.

**Tableau A3 : Taux de chômage selon le rang de l'échantillon
données pondérées**

Année d'enquête	Sous-échantillon			<i>Ensemble</i>	différence sortant / entrant
	entrant	médian	sortant		
1990	9,70 (0,19)	9,13 (0,18)	8,91 (0,18)	<i>9,24</i> (0,11)	-0,79 (0,26)
1991	9,24 (0,19)	9,43 (0,19)	8,62 (0,18)	<i>9,10</i> (0,11)	-0,62 (0,26)
1992	9,93 (0,18)	9,95 (0,19)	10,30 (0,19)	<i>10,06</i> (0,11)	+0,37 (0,27)
1993	10,90 (0,19)	11,13 (0,19)	11,03 (0,20)	<i>11,02</i> (0,11)	+0,13 (0,27)
1994	12,58 (0,20)	11,63 (0,19)	12,36 (0,20)	<i>12,18</i> (0,11)	-0,22 (0,29)
1995	11,83 (0,20)	11,49 (0,19)	10,91 (0,19)	<i>11,41</i> (0,11)	-0,92 (0,27)
1996	12,04 (0,20)	11,78 (0,20)	11,83 (0,20)	<i>11,88</i> (0,11)	-0,05 (0,28)
1997	12,67 (0,21)	12,02 (0,20)	11,75 (0,20)	<i>12,15</i> (0,12)	-0,92 (0,29)
1998	11,94 (0,20)	11,64 (0,20)	11,47 (0,20)	<i>11,68</i> (0,11)	-0,47 (0,28)
1999	11,92 (0,20)	11,50 (0,20)	11,45 (0,20)	<i>11,62</i> (0,11)	-0,47 (0,28)
2000	10,38 (0,19)	9,82 (0,18)	9,58 (0,18)	<i>9,93</i> (0,11)	-0,80 (0,26)
2001	8,58 (0,17)	8,66 (0,17)	8,61 (0,18)	<i>8,61</i> (0,10)	+0,03 (0,25)
2002	8,78 (0,18)	8,69 (0,18)	8,84 (0,18)	<i>8,77</i> (0,10)	+0,06 (0,25)

Champ : actifs au sens du BIT.

Note : données non pondérées. Les écarts-types sont indiqués entre parenthèses (calculés comme pour un sondage aléatoire simple).

Source : enquêtes annuelles Emploi, 1990 à 2002, Insee.

Tableau A4 : Indice de Bailar sur le nombre de chômeurs dans le CPS

	Rang d'interrogation du sous-échantillon							
	1	2	3	4	5	6	7	8
1968-1969	120,0	101,5	96,4	92,8	109,3	96,5	92,6	91,0
1970-1972	109,2	100,3	98,1	101,2	102,3	96,7	94,1	98,2
1987	107,0	100,4	98,8	100,2	100,4	95,6	96,5	101,2
1998	109,5	103,0	101,3	97,8	100,3	94,6	95,7	97,8

Source : Bailar, 1975 et Mansur et Shoemaker, 1999.

Dans le CPS, le nombre de chômeurs est plus élevé en première interrogation, mais aussi en 5^{ème} interrogation, interrogation qui intervient après 8 mois de non interrogation. Le biais est persistant au cours du temps. Il a cependant diminué après les changements de protocole intervenus en 1970 (changement de la position du module « inactivité » dans l'enquête), mais pas du tout après les changements de 1994 (changement de questionnaire, introduction de la saisie portable CAPI et CATI).

Tableau A5 : Indice de Bailar « corrigé » sur le taux de chômeurs dans l'EPA

	Rang d'interrogation du sous-échantillon					
	1	2	3	4	5	6
Trim 1 2003	99,7	104,0	105,9	97,0	99,0	94,2
Trim 2 2003	100,0	99,9	101,9	104,5	96,5	97,1
Trim 3 2003	101,1	101,0	97,8	96,7	104,2	99,5
Trim 4 2003	95,2	99,1	103,1	98,2	100,4	104,2
Trim 1 2004	104,3	95,0	100,2	98,7	99,1	102,6
Trim 2 2004	102,1	103,7	98,1	97,0	99,9	99,0
Moyenne	100,4	100,4	101,2	97,8	99,9	99,4

Source : Bardaji, 2004 (INE, EPA).

L'Encuesta de Poblacion Activa (EPA) espagnole a un schéma identique à celui de l'enquête Emploi : chaque trimestre 1/6^{ème} de l'échantillon est renouvelé. La première enquête a lieu en face à face, les 5 suivantes par téléphone. L'enquête est réalisée auprès de 65 000 logements chaque trimestre. Bardaji utilise plusieurs méthodes pour tester l'existence d'un biais de rotation, qui aboutissent toutes à la même conclusion : au niveau national il n'y a pas de biais de rotation sur le taux de chômage dans l'EPA. En revanche, un tel biais peut être mis en évidence dans 3 communautés autonomes : en Galice et en Andalousie, le taux de chômage est plus faible pour les primo-répondants, en Catalogne, le taux de chômage est plus élevé en première interrogation. Il n'y a pas de biais de rotation dans les autres communautés autonomes. Le tableau A5 fournit les indices de Bailar corrigés, c'est-à-dire corrigés des biais d'échantillonnage, pour le taux de chômage. Sur la période le taux de chômage en Espagne oscille entre 10,9 et 11,7%.

