

Analyse des transitions entre deux vagues dans l'enquête Emploi : le biais de rotation a des causes variables selon les vagues

Version très provisoire

Introduction

Les six interrogations de l'enquête Emploi et le biais de rotation

L'enquête Emploi est la source statistique qui permet de mesurer le chômage au sens du BIT. Depuis 2003, l'enquête a été renouvelée et est devenue trimestrielle (avec une collecte en continu). Il s'agit d'un panel de logements tournant, avec échantillon aréolaire (i.e. obtenu par tirage de zones géographiques) : chaque zone de l'échantillon est interrogée six fois, avant d'être remplacée par une zone géographiquement proche. Ce mode d'échantillonnage permet notamment de limiter la variance des évolutions des grandeurs d'intérêt, au premier rang desquelles le taux de chômage. Un trimestre donné, les données d'enquête exploitées pour fournir le taux de chômage proviennent de six « sous-échantillons » distincts, regroupant chacun des zones géographiques ayant déjà été interrogées un même nombre de fois. Théoriquement, chacun de ces sous-échantillons est représentatif de la population du champ dans son ensemble. Néanmoins, dans les faits, les résultats de ces sous-échantillons montrent (en moyenne) des différences systématiques en fonction du rang d'interrogation, différences qui ont été désignées sous le terme de « biais de rotation ». L'existence d'un biais de rotation pose la question d'un éventuel biais sur les résultats de l'enquête obtenus à partir de l'ensemble des sous-échantillons (et de la stabilité de ce biais dans le temps). Ce type d'argument a été mis en avant lors de la réunion de la formation Emploi-Revenus du CNIS du 8 mars 2007 pour justifier le report de la publication des résultats de l'enquête Emploi alors que ses résultats semblaient présenter des incohérences avec d'autres sources disponibles.

Analyse des transitions : apports par rapport aux analyses longitudinales sur six vagues

A cette époque, l'écart maximal de taux de chômage entre sous-échantillons était de 1,2 point en 2006¹, et même de 1,6 point en 2005². Un tel biais de rotation « en coupe » compare à une même date des sous-échantillons de logements différents. Dans les années précédentes, le biais de rotation avait déjà analysé de cette manière en interne à l'Insee, notamment par D. Goux, et les résultats disponibles en 2005 avaient été présentés aux Journées de Méthodologie Statistique cette année-là³.

¹ En effet, le taux de chômage de 2006 (moyenne annuelle) variait entre 10,5% (estimation à partir des sous-échantillons pondérés de rang 1) et 9,3% (estimation à partir des échantillons pondérés en quatrième ou cinquième interrogation).

² Cf. Lagarde, S. (2007) : « Approches comparées de l'évolution du chômage à travers l'enquête Emploi, els statistiques de l'ANPE et les enquêtes annuelles de recensement », présentation à la formation Emploi-Revenus du CNIS, 8 mars 2007, http://www.cnis.fr/agenda/DPR/DPR_0474.pdf.

³ Goux D. (2005) : « L'impact des changements intervenus dans l'enquête Emploi en 2003 sur la qualité de ses résultats », Journées de méthodologie statistiques, http://jms.insee.fr/site/files/documents/2006/440_1-JMS2005_SESSION15_GOUX_ACTES.PDF.

Plus récemment, dans le cadre du plan qualité de l'Insee sur l'enquête Emploi pour 2007, une nouvelle étape dans l'analyse des causes du biais de rotation de l'enquête Emploi s'est appuyée sur une approche longitudinale (suivi dans le temps des sous-échantillons au cours de leurs six interrogations). Cette stratégie permettait de s'affranchir des effets des fluctuations d'échantillon affectant le biais de rotation en coupe à une date donnée⁴. Cette approche longitudinale, exploratoire, a permis de mettre en évidence deux types d'effets intervenant dans le biais de rotation :

- les entrées et sorties des échantillons répondants (sur lesquelles des analyses détaillées ont été réalisées, selon les différentes causes d'entrées sorties : du logement, du ménage, de la personne...) ;
- le fait que, pour un logement toujours répondant, les réponses fournies peuvent varier en fonction de la vague.

Ces travaux ont montré que les effets des entrées et sorties d'échantillon étaient loin d'être négligeables. Sur données non pondérées (qui seront seules utilisées par la suite⁵), les variations de taux de chômage entre les vagues 1 et 6 passent de 0,8 point sur la population totale (10,6% en première vague contre 9,8% en dernière vague) à 0,4 point pour les seules personnes dont le statut d'activité est renseigné six fois (tableau 1).

Tableau 1 : le taux de chômage en fonction des vagues dans différents champs

Champ	1	2	3	4	5	6	moy enn e	min	max	écar t
Ensemble	10,6	10,4	10,1	9,9	9,7	9,8	10,1	9,7	10,6	0,97
Personnes avec statut d'activité renseigné 6 fois	9,0	9,1	8,8	8,8	8,8	8,6	8,9	8,6	9,1	0,57

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005.

Intéressante sur le plan exploratoire, l'approche longitudinale sur six vagues a d'un point de vue plus analytique l'inconvénient d'être sensible aux effets de « sélection endogène des sous-populations étudiées », et aux effets de cette sélection sur le biais de rotation. Le premier effet s'observe par exemple sur le cas des personnes ayant répondu six fois : se restreindre à cette population conduit à s'intéresser à des personnes dont le taux de chômage est relativement bas (les chômeurs étant comme on le verra, plus mobiles et plus souvent non-répondants pour d'autres raisons). Mais c'est le deuxième effet qui paraît le plus gênant pour des analyses portant sur le biais de rotation : l'approche consistant à comparer des logements ou personnes sur la base de leurs résultats sur six vagues revient dans les faits à adopter des règles de sélection différentes pour chaque vague. Pour une personne ayant répondu six fois, les situations de première et dernière vague ne sont en effet pas exactement comparables : on observe en première vague la probabilité d'être chômeur, sachant que la personne sera répondante aux 5 vagues suivantes, tandis qu'en dernière vague, on observe la probabilité d'être chômeur, sachant que la personne a été répondante aux cinq vagues précédentes. Or ces deux probabilités n'ont pas de raison d'être identiques : tout écart peut contribuer au biais de rotation estimé en longitudinal. Enfin, l'approche sur six vagues a un autre inconvénient, lié : en privilégiant des sélections et comparaisons sur six vagues, elle ne permet pas facilement de

⁴ S'apparentant en cela aux calculs d'indices du type de Brisebois et Mantel mis en œuvre par D. Goux sur données en coupe.

⁵ Dans les faits, la pondération utilisée corrige peu le biais de rotation (les variables pouvant être utilisées pour le calage étant trop peu liées à la situation sur le marché du travail). L'intérêt des analyses n'est donc pas remis en cause par ce choix, qui permet de mieux séparer les différents effets.

mettre en évidence les spécificités de certaines vagues, qui, comme on le verra, sont importantes.

C'est pourquoi une nouvelle approche est développée ici, consistant à étudier les transitions des personnes entre deux vagues successives d'un sous-échantillon. Ces transitions sont étudiées sur l'ensemble des personnes présentes à l'une ou l'autre de deux vagues successives, ce qui permet un mécanisme de sélection identique pour toutes les vagues, les seules différences entre vagues étant celles du protocole de collecte au sens large que l'on est intéressé à étudier. En particulier, les transitions entre les vagues 2 et 3, 3 et 4 et 4 et 5 ne diffèrent que par le nombre d'interrogations précédentes⁶, ce qui permettra d'étudier l'impact de ce nombre d'interrogation. Dans ce travail, les analyses descriptives des transitions sont complétées par une approche plus analytique, décomposant chaque transition en plusieurs étapes (probabilité d'être encore dans le logement, probabilité de répondre, probabilité de changer d'activité sachant que l'on est répondant...). Ces étapes serviront dans une deuxième partie de base à des simulations permettant de mieux évaluer l'impact des différents effets sur le biais de rotation.

Les données utilisées : les 12 sous-échantillons introduits entre le premier trimestre 2003 et le quatrième trimestre 2005

Les données utilisées sont celles des 6 dates d'observations des 12 sous-échantillons introduits entre le premier trimestre 2003 et le quatrième trimestre 2005⁷. 1 031 591 observations ont un statut d'activité au sens du B.I.T. renseigné (correspondant à des personnes de tous âges, y compris les moins de quinze ans). Pour un rang de vague donné (sous échantillon en première interrogation, deuxième interrogation...), les effectifs disponibles sont six fois moindres, de l'ordre de 172 000. Dans la suite toutefois, nous ne fournirons plus les nombres d'observations utilisés (obtenus en agrégeant 12 sous-échantillons pour plus de robustesse), mais les nombres moyens d'observations par sous-échantillon (et vague). En effet, ce qui importe pour l'analyse de l'enquête Emploi, c'est bien de comparer pour un trimestre donné deux sous-échantillons correspondant à deux rangs d'interrogations différents et les ordres de grandeur fournis ici sont donc ceux qui sont pertinents pour cette comparaison.

Un fort impact des entrées et sorties sur le biais de rotation

Les mouvements d'entrées –sorties des répondants concernent plus de personnes que les changements de statut d'activité, et sont d'ampleur plus variables en fonction des vagues.

En moyenne, un peu plus de 14 300⁸ personnes d'un sous-échantillon répondent à un rang d'interrogation donné. Parmi elles, environ 1 000 sortent de l'échantillon de répondants entre

⁶ Dans tout le texte, le rang considéré est le rang d'interrogation du sous-échantillon du logement, et non de la personne.

⁷ Les données utilisées ne tiennent pas compte de la modification du statut d'activité qui a été mise en œuvre à l'automne 2007 et rétroplacée sur l'ensemble des fichiers de l'enquête Emploi en continu.

⁸ Soit 172 000 divisé par 12.

deux vagues, et environ autant y entrent, tandis qu'un peu plus de 600 changent de statut d'activité au sens du B.I.T. entre les deux dates (tableau 2). Ces premiers chiffres confirment que les variations d'échantillons ne sont *a priori* pas négligeables, dans la mesure où elles concernent plus de personnes que les variations de statut d'activité au sens du B. I. T ..

transition	effectifs selon la transition						répartition sur les présents à au moins une des deux dates selon la transition					
	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy.	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy.
tjrs présent – même statut	12561	12989	13010	13007	12179	12749	79%	81%	82%	83%	75%	80%
tjrs présent – autre statut	646	621	596	588	621	615	4%	4%	4%	4%	4%	4%
tjrs présent – total	13208	13610	13607	13595	12800	13364	83%	85%	85%	86%	78%	84%
va sortir de l'échantillon	837	877	867	862	1493	987	5%	5%	5%	5%	9%	6%
va entrer dans l'échantillon	1280	864	850	697	1411	1020	8%	5%	5%	4%	9%	6%
total vague initiale	14044	14488	14474	14457	14292	14351	100%	100%	100%	100%	100%	100%
total vague finale	14488	14474	14457	14292	14211	14384						

Tableau 2 – Changements de situation individuelle entre deux vagues successives

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, personnes présentes à au moins une des deux vagues étudiées.

Toutefois, ce qui importe pour le biais de rotation n'est pas d'étudier la fréquence moyenne des transitions, mais plutôt d'isoler celles qui sont d'importance ou de nature⁹ variable selon les vagues. A cet égard, les nombres d'entrées et de sorties de l'ensemble des répondants culminent tous les deux à plus de 1 400 entre les vagues 5 et 6, et le nombre d'entrées reste élevé (proche de 1 300) entre les vagues 1 et 2. Entrées comme sorties restent inférieures à 900 lors des autres transitions.

Les forts mouvements d'échantillons observés entre les vagues 5 et 6 ne sont que partiellement dus au protocole de collecte théorique.

Le nombre élevé de mouvements en vague 6 s'explique en partie du fait du protocole de collecte « théorique », qui repousse en dernière vague le repérage de nouveaux logements ainsi que la réinterrogation de deux catégories de logements imputés en vagues intermédiaires.

Le premier effet est lié à l'échantillon aréolaire de l'enquête Emploi : les logements à interroger sont les logements « ordinaires » qui appartiennent à une zone géographique définie. Leur liste est établie par l'enquêteur avant le début de la première vague d'enquête lors du « ratissage » : ce dernier consiste pour l'enquêteur à recenser l'ensemble des logements « ordinaires » (i.e. non collectifs) présents sur l'aire géographique qu'il est chargé d'enquêter. Le « ratissage » qui nécessite un déplacement (coûteux) sur le terrain n'est demandé qu'en première et dernière interrogation d'une aire, les mouvements de logements étant jugés trop peu nombreux pour justifier le coût supplémentaire qu'entraînerait leur repérage aux vagues intermédiaires. L'ajout de nouveaux logements dans l'aire n'étant pas

⁹ Se reporter à l'annexe pour un tableau équivalent au tableau 2 détaillé selon le statut d'activité de départ (ou d'arrivée).

possible aux vagues 2 à 5, les habitants des logements concernés¹⁰ ne peuvent être interrogés avant la vague 6. En moyenne, chaque trimestre, 135 personnes habitant de « nouveaux » logements sont effectivement interrogées pour la première fois lors de cette vague-là (tableau 3). Elles ont un taux de chômage un peu plus faible que la moyenne (8,7%).

Tableau 3 – Les entrées en cours d'enquête en fonction du motif d'absence à la vague précédente, du statut d'activité à la vague courante et du rang d'interrogation

vague initiale de la transition	effectifs						taux d'entrée						
	1	2	3	4	5	moy. 1 à 5	1	2	3	4	5	moy. 1 à 5	moy. 1 à 4
total													
actif occupé	516	364	366	302	595	428	9,2%	6,3%	6,4%	5,3%	10,6%	7,5%	6,8%
chômeur	73	63	58	45	104	69	11,0%	9,5%	9,1%	7,2%	17,3%	10,7%	9,2%
inactif	691	437	425	350	712	523	8,9%	5,4%	5,3%	4,3%	8,8%	6,5%	5,9%
total	1280	864	850	697	1411	1020	9,1%	6,0%	5,9%	4,8%	9,9%	7,1%	6,4%
entrées "normales" (correspondant à des entrées dans le champ)													
actif occupé	104	97	107	94	298	140	1,9%	1,7%	1,9%	1,6%	5,3%	2,5%	1,8%
logement en '14' à la vague initiale	0	0	0	0	42	9							
logement pas en '14' à la vague initiale (ni finale)	104	97	107	94	256	131							
chômeur	21	21	23	17	57	28	3,1%	3,2%	3,6%	2,7%	9,4%	4,3%	3,2%
logement en '14' à la vague initiale	0	0	0	0	7	1							
logement pas en '14' à la vague initiale (ni finale)	21	21	23	17	50	26							
inactif	150	140	149	132	374	189	1,9%	1,7%	1,8%	1,6%	4,6%	2,4%	1,8%
logement en '14' à la vague initiale	0	0	0	0	57	12							
logement pas en '14' à la vague initiale (ni finale)	150	140	149	132	316	177							
total	275	258	279	243	728	357	2,0%	1,8%	1,9%	1,7%	5,1%	2,5%	1,8%
logement en '14' à la vague initiale	0	0	0	0	107	21							
logement pas en '14' à la vague initiale (ni finale)	275	258	279	243	621	335							
Dont : nouveau logement													
actif occupé	0	0	0	0	63	13	0,0%	0,0%	0,0%	0,0%	1,1%	0,2%	0,0%
chômeur	0	0	0	0	6	1	0,0%	0,0%	0,0%	0,0%	1,0%	0,2%	0,0%
inactif	0	0	0	0	66	13	0,0%	0,0%	0,0%	0,0%	0,8%	0,2%	0,0%
total	0	0	0	0	135	27	0,0%	0,0%	0,0%	0,0%	0,9%	0,2%	0,0%
entrées suite à non-interrogation de ménages supposés dans le champ													
actif occupé	412	267	260	208	297	289	7,4%	4,6%	4,5%	3,6%	5,3%	5,1%	5,0%

¹⁰ Les nouveaux logements repérés en vague 6 sont aux quatre cinquièmes des logements neufs, mais il y a également de faibles proportions de logements précédemment oubliés, quelques logements issus de la transformation d'un local ou de l'éclatement d'un logement.

chômeur	52	42	35	28	47	41	7,9%	6,3%	5,5%	4,5%	7,9%	6,4%	6,1%
inactif	541	297	276	218	339	334	6,9%	3,7%	3,4%	2,7%	4,2%	4,2%	4,2%
total	1005	606	572	454	683	664	7,2%	4,2%	3,9%	3,1%	4,8%	4,6%	4,6%

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005. Pour les effectifs : personnes dont le statut d'activité est connu à la deuxième des deux vagues étudiées et non à la précédente (dite « vague initiale »).

Lecture : en moyenne un trimestre donné (et donc pour un sous-échantillon de rang 1 donné), 516 actifs occupés répondant en vague 2 n'avaient pas eu leur statut d'activité collecté en vague 1, soit un effectif équivalent à 9,2% des effectifs d'actifs occupés de vague 1 (« taux d'entrée »). Ces 516 entrées se décomposent en 104 entrées repérées comme des entrées dans le champ de l'enquête, tandis que 4121 correspondent à des personnes n'ayant pas pu être interrogées en vague 1 bien qu'a priori déjà dans le champ.

Le deuxième effet est lié au choix d'imputer les réponses des logements d'inactifs de 65 ans aux vagues intermédiaires. Ce protocole se justifie en raison de la faible probabilité de retour en activité des personnes concernées : ne pas les interroger en vague intermédiaire permet d'utiliser le budget alloué à l'enquête de manière plus efficace (en le consacrant aux ménages pour lesquels des changements sont plus probables) et de ne pas laisser les personnes concernées. Mais cette procédure fait également implicitement la double hypothèse que les personnes concernées sont d'une part toujours présentes en vague intermédiaire, et d'autre part, qu'elles auraient répondu si on avait effectivement essayé de les interroger. La réinterrogation de ces logements en vague 6 permet de constater que ce n'est pas toujours le cas puisque pour un sous-échantillon donné, en moyenne 261 de ces personnes (10%, ne représentant que 2% de l'ensemble des inactifs) ne répondent plus en vague 6 (tableau 4). S'agissant par construction de personnes inactives de 65 ans ou plus¹¹, la non-suppression de ces personnes des répondants n'a pas d'influence directe sur le biais de rotation sur le taux de chômage. En revanche, l'observation retardée des personnes les ayant remplacées pourrait en avoir. Ces interrogations retardées peuvent également se produire pour les résidences secondaires qui seraient devenues résidences principales entre la première et cinquième vague. En effet, l'enquêteur n'a pas à vérifier en vague intermédiaire si les logements décrits comme résidences secondaires à une vague antérieure ont conservé ce statut : ils sont automatiquement imputés comme l'ayant conservé et donc comme étant hors champ à ces vagues-là. Ces deux types d'entrées retardées pourraient contribuer au biais de rotation, à condition d'être nombreuses ou très atypiques. Ce n'est heureusement pas le cas : chaque trimestre, en moyenne, seules 107 personnes appartenant à un logement préalablement imputé sont interrogées pour la première fois en vague 6 (tableau 3), dont environ 70 appartenant à un logement précédemment composés d'inactifs, le reste appartenant à une ancienne résidence secondaire. La majorité de ces nouveaux enquêtés sont inactifs ce qui, pour les logements précédemment inactifs, correspond à l'idée que les ménages se succédant au sein d'un même logement restent sociologiquement proches (tableau 3). *In fine*, seule une cinquantaine de nouveaux arrivants sont actifs (avec un taux de chômage est proche de 15 %).

Au total cependant, compte tenu des faibles effectifs concernés, les dispositions décrites ci-dessus ont un impact négligeable sur le biais de rotation¹² comme le confirmeront les simulations effectuées ci-dessous^{13,14}.

¹¹ Parmi les inactifs concernés, il s'avère en effet que 135 ne doivent plus être interrogés (ils n'habitent plus le logement, celui-ci est vacant...) et que 126 autres ne peuvent l'être (refus, impossibles à joindre).

¹² Ou tout du moins, sur le biais de rotation sur données brutes, les éventuelles conséquences sur les données redressées n'étant pas analysées ici.

¹³ Les simulations effectuées plus loin montrent que ces deux dispositions sont à l'origine d'une hausse de 0,02 point du taux de chômage brut en vague 6 (et de variations encore plus négligeables aux autres vagues).

Tableau 4 – Les sorties en cours d'enquête en fonction du motif du statut d'activité à la vague précédente et du rang d'interrogation

	Effectifs						probabilité de sortie						
	1	2	3	4	5	moy. 1 à 5	1	2	3	4	moy. 1 à 5	moy. 1 à 4	
vague initiale													
Total													
actif occupé	377	394	402	398	590	432	6,7%	6,9%	7,0%	7,0%	10,5%	7,6%	6,9%
chômeur	74	70	68	71	80	73	11,2%	10,5%	10,6%	11,3%	13,3%	11,3%	10,9%
inactif	385	413	398	394	822	482	4,9%	5,1%	4,9%	4,8%	10,2%	6,0%	5,0%
<i>inactif en '14' à la vague initiale</i>	0	0	0	0	261	52	0,0%	0,0%	0,0%	0,0%	12,3%	3,2%	0,0%
<i>inactif en '14' à aucune des vagues</i>	385	413	398	394	561	430	6,5%	6,8%	6,6%	6,6%	9,4%	7,2%	6,6%
total	837	877	867	862	1493	987	6,0%	6,1%	6,0%	6,0%	10,4%	6,9%	6,0%
sorties correspondant à des sorties du champ													
actif occupé	135	147	168	148	237	167	2,4%	2,6%	2,9%	2,6%	4,2%	2,9%	2,6%
chômeur	31	32	29	29	37	32	4,7%	4,8%	4,6%	4,7%	6,1%	5,0%	4,7%
inactif	134	150	155	139	362	188	1,7%	1,9%	1,9%	1,7%	4,5%	2,3%	1,8%
<i>inactif en '14' à la vague initiale</i>		0	0	0	135	27	0,0%	0,0%	0,0%	0,0%	6,3%	1,7%	0,0%
<i>inactif en '14' à aucune des vagues</i>	134	150	155	139	227	161	2,2%	2,5%	2,6%	2,3%	3,8%	2,7%	2,4%
total	300	329	352	317	636	387	2,1%	2,3%	2,4%	2,2%	4,4%	2,7%	2,3%
sorties liées à une non-interrogation de personnes supposées dans le champ (prob. cond.)													
actif occupé	242	247	234	249	353	265	4,4%	4,4%	4,2%	4,5%	6,6%	4,8%	4,4%
chômeur	43	38	39	41	43	41	6,8%	6,0%	6,3%	6,9%	7,6%	6,7%	6,5%
inactif	252	264	243	255	461	295	3,3%	3,3%	3,1%	3,2%	6,0%	3,8%	3,2%
<i>inactif en '14' à la vague initiale</i>	0	0	0	0	126	25	0,0%	0,0%	0,0%	0,0%	5,9%	1,6%	0,0%
<i>inactif en '14' à aucune des vagues</i>	252	264	243	255	334	270	4,3%	4,5%	4,1%	4,4%	5,9%	4,6%	4,3%
total	537	548	516	545	857	601	3,9%	3,9%	3,7%	3,9%	6,3%	4,3%	3,8%

¹⁴ Un tel résultat ne doit cependant pas être généralisé, car ces dispositions peuvent avoir un impact important sur d'autres indicateurs. Un exemple extrême est le taux de non-réponse, dont le niveau et le profil par vague est fortement influencé par le traitement adopté pour les logements d'inactifs de 65 ans ou plus. Les complications d'analyse que peut occasionnellement entraîner la non-réinterrogation des inactifs en vague intermédiaire restent toutefois des contreparties légères par rapport aux gains qu'elle permet, notamment en terme d'acceptabilité de la part des ménages de retraités. Ces ménages se sentent en effet souvent peu concernés par l'enquête (bien que sur les dépliants et documents d'information destinés aux ménages elle ait récemment été rebaptisée à dessein « enquête sur l'emploi, le chômage et l'inactivité ») et ils ne sont pas rares à mettre avant le sujet de l'enquête pour justifier leur non-participation, en particulier lors de réinterrogations ou à l'occasion de l'enquête non-répondant. La question des ménages âgés n'est d'ailleurs pas spécifique à la France, et d'autres pays ont fait remonter à Eurostat les difficultés rencontrées pour remonter chaque trimestre des informations les concernant.

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 200, personnes dont le statut d'activité est connu à la première des deux vagues étudiées (« vague initiale ») et (pour les sorties) non à la suivante.

Lecture : en moyenne un trimestre donné (et donc pour un sous-échantillon de rang 1 donné), 377 actifs occupés répondant en vague 1 ne répondent plus en vague 2, soit 6,7% des effectifs d'actifs occupés de vague 1. Ces 377 sorties se décomposent en 135 entrées repérées comme des entrées dans le champ de l'enquête (soit une probabilité de sortie du champ de 2,4%, tandis que 242 correspondent à des personnes n'ayant pas pu être interrogées en vague 2 bien qu'a priori encore dans le champ (soit une probabilité d'autre non-réponse, conditionnellement au fait d'être dans le champ, de 4,4%).

Par ailleurs, ces deux dispositions du protocole n'expliquent par ailleurs que très partiellement le surcroît de mouvements d'entrée / sortie observés entre les vagues 5 et 6. En effet, si l'on répartit les mouvements ne pouvant être observés qu'entre les vagues 5 et 6 sur l'ensemble des transitions auxquelles ils auraient pu se produire¹⁵, le nombre d'entrées en vague 6 passerait de 1411 à 1231, et le nombre de sorties de 1493 à 1317 (tableau 5a), ce qui reste très supérieur aux nombres observés aux vagues précédentes et confirme que les principales causes du nombre élevé de mouvements en vague 6 doivent être recherchées ailleurs.

¹⁵ La ventilation est effectuée par statut d'activité et type de mouvement. Les entrées liées à la non-interrogation des logements imputés sont réparties entre la vague 6 et les autres de manière à obtenir (pour chaque motif de non-réponse et statut d'activité) une répartition des entrées entre la vague 6 et les autres identiques à celle observée sur les logements non imputés. Ceci permet de tenir compte du fait que pour les logements non imputés, la part des entrées observée en dernière vague est souvent supérieure à 30% (alors qu'elle aurait dû être de 20% si elle était répartie sur les cinq transitions possibles). Une méthode analogue est employée pour les sorties. Pour les nouveaux logements, en l'absence d'information comparable, la répartition revient simplement à répartir les entrées à l'identique sur les cinq transitions possibles.

Tableau 5 : Entrées et sorties après redressement des logements imputés et des logements neufs

a) par statut d'activité

vague initiale de la transition	Effectifs						taux d'entrée / probabilité de sortie							
	1	2	3	4	5	moy. 1 à 5	1	2	3	4	5	moy. 1 à 5	moy. 1 à 4	
Total des entrées														
actif occupé	535	383	386	322	516	428	9,6%	6,7%	6,7%	5,6%	9,2%	7,5%	7,1%	
chômeur	75	65	61	48	95	69	11,3%	9,8%	9,5%	7,6%	15,7%	10,7%	9,6%	
inactif	714	460	448	373	620	523	9,2%	5,7%	5,5%	4,6%	7,7%	6,5%	6,2%	
total	1325	909	895	742	1231	1020	9,4%	6,3%	6,2%	5,1%	8,6%	7,1%	6,7%	
dont : logement en '14' à la vague initiale (simulé)	18	19	18	18	34	21								
Total des sorties														
actif occupé	377	394	402	398	590	432	6,7%	6,9%	7,0%	7,0%	10,5%	7,6%	6,9%	
chômeur	74	70	68	71	80	73	11,2%	10,5%	10,6%	11,3%	13,3%	11,3%	10,9%	
inactif	429	457	441	438	647	482	5,5%	5,7%	5,5%	5,4%	8,0%	6,0%	5,5%	
inactif en '14' à la vague initiale (simulé)	44	44	44	44	86	52		2,4%	2,2%	2,1%	4,0%	3,2%		
inactif en '14' à aucune des vagues	385	413	398	394	561	430	6,5%	6,8%	6,6%	6,6%	9,4%	7,2%	6,6%	
total	881	921	911	906	1317	987	6,3%	6,4%	6,3%	6,3%	9,2%	6,9%	6,3%	

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, personnes présentes à la première (« vague initiale ») ou à la deuxième des deux vagues de la transition considérée.

Lecture : en moyenne un trimestre donné (et donc pour un sous-échantillon de rang 1 donné), 535 actifs occupés n'ayant pas eu leur statut d'activité collecté en vague 1 ont eu leur statut d'activité collecté en vague 2 ou l'auraient eu si le protocole de collecte théorique avait prévu la réinterrogation des ménages d'inactifs et des résidences secondaires ainsi que le repérage des nouveaux logements en vague intermédiaire. Cela représente un effectif équivalent à 9,6% des effectifs d'actifs occupés de vague 1.

b) Par motif détaillé

	effectifs						taux d'entrée						
	1	2	3	4	5	moy. 1 à 5	1	2	3	4	5	moy. 1 à 5	moy. 1 à 4
Entrées													
Entrées totales	1325	909	895	742	1231	1020	9,4%	6,3%	6,2%	5,1%	8,6%	7,1%	6,7%
Entrées dans le champ	320	303	324	288	548	357	2,3%	2,1%	2,2%	2,0%	3,8%	2,5%	2,1%
changement de ménage	76	80	93	74	156	96	0,5%	0,6%	0,6%	0,5%	1,1%	0,7%	0,6%
hors champ	152	133	131	120	261	159	1,1%	0,9%	0,9%	0,8%	1,8%	1,1%	0,9%
mobilité individuelle	65	63	74	66	105	75	0,5%	0,4%	0,5%	0,5%	0,7%	0,5%	0,5%
nouveau logement	27	27	27	27	27	27	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%
Entrées suite à non-réponse	1005	606	572	454	683	664	7,2%	4,2%	3,9%	3,1%	4,8%	4,6%	4,6%
absent de longue durée	245	105	101	90	104	129	1,7%	0,7%	0,7%	0,6%	0,7%	0,9%	0,9%
impossible à joindre	550	417	407	319	477	434	3,9%	2,9%	2,8%	2,2%	3,3%	3,0%	2,9%
inapte	36	18	11	11	31	21	0,3%	0,1%	0,1%	0,1%	0,2%	0,1%	0,1%
refus / activité inconnue	175	65	52	34	71	79	1,2%	0,4%	0,4%	0,2%	0,5%	0,6%	0,6%
Sorties	effectifs						probabilité de sortie (conditionnelles à la présence dans le champ pour les non-réponses)						
Sorties totales	881	921	911	906	1317	987	6,3%	6,4%	6,3%	6,3%	9,2%	6,9%	6,3%
Sorties du champ	300	329	352	317	636	387	2,1%	2,3%	2,4%	2,2%	4,4%	2,7%	2,3%
changement de ménage	66	77	85	70	173	94	0,5%	0,5%	0,6%	0,5%	1,2%	0,7%	0,5%
logement hors champ	150	167	173	158	297	189	1,1%	1,2%	1,2%	1,1%	2,1%	1,3%	1,1%
mobilité individuelle	84	85	94	89	166	104	0,6%	0,6%	0,7%	0,6%	1,2%	0,7%	0,6%
Sorties suite à non-réponse	537	548	516	545	857	601	3,9%	3,9%	3,7%	3,9%	6,3%	4,3%	3,8%
absent de longue durée	89	97	100	105	224	123	0,6%	0,7%	0,7%	0,7%	1,6%	0,9%	0,7%
impossible à joindre	379	386	359	383	444	390	2,8%	2,7%	2,5%	2,7%	3,3%	2,8%	2,7%
inapte	13	9	10	13	42	18	0,1%	0,1%	0,1%	0,1%	0,3%	0,1%	0,1%
refus / activité inconnue	57	57	47	43	147	70	0,4%	0,4%	0,3%	0,3%	1,1%	0,5%	0,4%

Source : Enquêtes Emploi, T1-2003 au T1-2007 et simulations. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005.

Note : le motif d'entrée (resp. sortie) correspond en premier lieu au motif de non-réalisation de l'enquête à la vague précédente (resp. suivante) : les changements de ménage ne correspondent qu'aux situations pour lesquelles un logement a répondu aux deux vagues concernées, avec un changement du ménage l'habitant entre les deux¹⁶.

Les modifications importantes de la liste des répondants entre les deux dernières vagues s'explique tant les mouvements dans et hors du champ que par les autres motifs d'entrées sorties

Pour mieux cerner ces causes, commençons par examiner les motifs de sortie ou d'entrée détaillés recueillis lors de l'enquête (tableau 5b). Toutes les catégories de mouvements sont plus fréquentes lors de la dernière transition. C'est le cas des mouvements d'entrées et sorties en lien avec des refus ou non-contacts (impossible à joindre...), qui sont en dernière vague supérieurs de moitié à ceux en général observé aux autres vagues. Cette spécificité de la dernière vague s'explique dans la mesure où cette dernière a un protocole différent des quatre

¹⁶ Ce choix de considérer le motif de non-réalisation de l'enquête avant l'éventuel changement de ménage permet un traitement plus symétrique des entrées et des sorties. En effet, dans de nombreux cas, la seule information disponible au moment où la sortie est constatée est l'impossibilité d'enquête, sans que l'on sache souvent avec certitude si c'est suite à un départ du ménage ou à une impossibilité à le joindre. A contrario, lorsqu'une (ré)entrée est observée, on sait si c'est avec un nouveau ménage. Accorder la primauté au changement de ménage sur les motifs reviendrait donc à observer un plus grand nombre d'entrées pour changement de ménage que de sorties.

précédentes : elle se déroule en face à face et avec un questionnement plus lourd pour l'enquêté, mais également plus rémunérateur pour l'enquêteur. De manière plus surprenante, les mouvements d'entrées et sorties du champ sont aussi plus nombreux entre les cinquième et sixième vagues : les taux d'entrées (ou de sortie) du champ sont de l'ordre de 4% à cette dernière vague, contre environ 2 % aux autres vagues. Tous les types de mouvements dans et hors du champ sont concernés (changements de ménage, mobilité individuelle, sortie du logement du champ...). Or, une fois corrigés les effets d'entrées et sorties du champ directement liés au protocole, les nombres (ou taux) d'entrées et sorties du champ moyens devraient peu varier selon la vague. Deux types d'explications peuvent être avancés pour expliquer l'ampleur des variations observées. Un premier type d'explication consisterait à se demander si le résultat surprenant sur les mouvements vis-à-vis du champ n'est pas simplement lié à une mauvaise classification de certains logements n'ayant pas pu être contactés. Il n'est en effet pas toujours évident de savoir si le non-contact résulte d'une sortie du champ (n'habite plus dans le logement) ou simplement d'insuccès dans les tentatives d'approches. Malgré les consignes données aux enquêteurs de se déplacer systématiquement en vague intermédiaire en cas d'absence de contact¹⁷, il est possible que les déplacements a priori plus nombreux en visite¹⁸, comme le format d'interrogation de cette dernière conduisent à des codifications différentes du motif en cas de doute, voire à une information plus précise sur le statut réel du logement vis-à-vis du champ (sollicitations facilitées des voisins à ce sujet, informations relevées sur les boîtes aux lettres...). Il paraît toutefois difficile de penser que la seule amélioration de l'information sur les motifs de mouvements explique une partie significative de la très forte hausse des entrées et sorties de champ en dernière vague, dans la mesure où l'on n'observe pas par ailleurs de diminution des effectifs d'enquêtés impossibles à joindre. On est donc amené à conclure à l'existence d'un deuxième effet : la manière dont l'enquête est réalisée sur le terrain en sixième vague conduit dans les faits à un plus grand nombre d'entrées dans le champ (probablement parce que les démarches de contact auprès de logement auparavant vides, ou ayant de nouveaux habitants sont plus poussées ou fructueuses lors des enquêtes en visites). Elle semble conduire également à un plus grand nombre de sorties de champ, ce qui reste plus délicat à expliquer sans avoir recours à des arguments sur la qualité de l'information sur le motif de mouvement (du type : le motif de sortie a d'autant plus de raisons d'être connu avec précision que les tentatives de contact auprès du logement sont nombreuses). On pourrait toutefois dans certains cas particuliers penser que des logements sortis du champ ont pu continuer à être interrogés à tort par téléphone en vague intermédiaire (dans le cas des ménages qui auraient déménagé sans changer de numéro de téléphone)¹⁹.

¹⁷ Cela a été précisé dans la version 2006 des instructions de collecte : « Les ré-interrogations se font normalement par téléphone sauf quelques cas particuliers où l'interrogation se fait en face à face. L'enquêteur tentera de joindre les ménages à différentes tranches horaires. Au bout de 5 tentatives infructueuses, il devra renouveler sa tentative au moins 2 fois en se rendant à domicile. Ces déplacements sur le terrain seront l'occasion de vérifier le statut des logements vacants au trimestre précédent ». « Si la ré-interrogation ne nécessite pas de déplacements particuliers, l'enquêteur vérifiera par une visite spécifique que les logements vacants le sont toujours. Dans le cas contraire, il les interrogera. »

¹⁸ Selon la version 2006 des instructions de collecte : « Pour les aires entrantes et les aires sortantes ainsi que les nouveaux ménages ou ceux qui n'ont jamais été interrogés l'enquêteur fera au moins 4 tentatives pour rencontrer les ménages en face à face. »

¹⁹ Les changements de ménage sont en effet beaucoup plus fréquents lors de la dernière transition. Pour plus d'un tiers des ménages, ces changements sont observés après une 5^{ème} interrogation où le logement était considéré comme vacant, et pour plus d'un autre tiers, après une cinquième interrogation réussie (le cas des logements impossibles à joindre en vague 5 constitue lui environ 20% des cas, et cette répartition est proche de celle observée pour les changements de ménage de la vague 4).

Une écriture paramétrique des probabilités de transitions permet une séparation et que quantification des différents effets décrits ci-dessus. Même si comme indiqué ci-dessus il est en pratique parfois difficile de séparer les mouvements vis-à-vis du champ et des mouvements liés à des « autres non réponses » (i.e. concernant des logements supposés dans le champ), ces deux catégories de mouvements, relevant de mécanismes différents, seront traités à part dans la formalisation. Ce traitement séparé est d'autant plus justifié que ces deux types de mouvements n'ont pas non plus la même incidence sur les résultats de l'enquête. En effet, les sorties du champ devraient a priori pouvoir être compensées par les entrées dans le champ (nouveau ménage remplaçant l'ancien, logement vacant qui ne l'est plus...), alors que les non-réponses de personnes dans le champ ne le sont pas (seules les procédures de redressement et de calage peuvent tenter d'en contrebalancer les effets).

L'impact de l'attrition sur le biais de rotation : un fort effet de la probabilité de non-réponse élevée des chômeurs

Commençons par étudier les probabilités de sortie. En moyenne, 6,9% des répondants à une vague donnée ne le sont plus à la vague suivante²⁰. Cette probabilité de sortie varie avec le statut d'activité de la personne au sens du B.I.T. à la vague initiale. C'est parmi les inactifs qu'elle est la plus faible (6%), en particulier chez les inactifs appartenant à des ménages d'inactifs de 65 ans ou plus (3,2%). Chez les actifs, la proportion de sortie est plus élevée chez les chômeurs (11,3%) que chez les actifs occupés (7,6%). Cette probabilité de sortie, qui varie peu entre les vagues 1 à 4 (autour de 6%)²¹, augmente de plusieurs points entre les vagues 5 et 6, ce type de profil par vague s'observant aussi pour chacune des catégories considérées²².

La probabilité de sortie moyenne (6,9%) correspond à la somme de la probabilité de sortir du champ (2,7%) et de la probabilité de ne pas répondre pour d'autres raisons (4,3% conditionnellement au fait d'être encore dans le champ)²³. Même après correction des logements imputés²⁴, ces probabilités varient selon la vague : la probabilité de sortie de champ atteint 4,4% lors de la dernière transition, contre 2,3% en moyenne lors des quatre premières transitions, tandis que la probabilité de sortie pour « autre raison » (conditionnellement au fait d'être dans le champ) est plus élevée lors de la dernière transition (6,3%) que lors des transitions précédentes (où elle varie peu autour de 3,8%).

²⁰ Les moyennes calculées ici le sont en rapportant la somme des sorties observées pour la catégorie et les vagues considérées sur les effectifs de personnes susceptibles de sortir. Cela correspond à une estimation des probabilités de sortie par maximum de vraisemblance (qui peut différer très légèrement de la moyenne arithmétique des probabilités de sorties par vague). De cette manière, les probabilités moyennes sur l'ensemble des vagues ne dépendent pas de la correction que l'on peut apporter aux logements d'inactifs de 65 ans ou plus (en effet, seul le nombre de sorties est pris en compte, et non les vagues auxquelles les sorties se sont produites).

²¹ Ce qui est une bonne nouvelle, car l'attrition pourrait être plus grande après la première vague (comme cela avait notamment été observé dans le panel européen des ménages).

²² Si l'on raisonne tous motifs de sortie confondus et sur les seules transitions commençant aux vagues 1 à 4, des tests du chi2 menés séparément sur les actifs occupés, les chômeurs, les inactifs (hors ceux imputés) ne permettent pas de rejeter l'hypothèse que la probabilité de sortir de l'échantillon est indépendante de la vague pour ces trois sous-populations. Cependant, si l'on effectue la même analyse par motif de sortie, on peut observer des petites variations entre vagues, avec jusqu'en vague 3, une légère tendance à la hausse des sorties du champ, compensée par une tendance à la baisse des autres non-réponses. Ces effets restent à expliquer (obtention d'une meilleure information ou stratégie de l'enquêteur ?).

²³ On choisit dans cette partie de mettre sur le même plan les deux risques de sortie possible (sortie du champ, non-réponse), et non de conditionner le second risque à la non-réalisation du premier.

²⁴ Cf. données complètes corrigées en annexe. On retrouve des résultats similaires en raisonnant sur les seuls logements non imputés lors de la vague initiale de la transition.

Compte tenu de ce qui précède, nous posons les hypothèses suivantes sur les probabilités de sorties $p(a,v,m)$ – où v est la vague v , a le statut d'activité initial et m le motif de sortie (m valant hc pour hors champ ou anr pour autre non réponse) :

- Les « vraies » probabilités de sorties de champ $\pi(a, hc)$ varient selon le statut d'activité initial mais non selon la vague. Les sorties de champ sont supposées imparfaitement observées lors des vagues téléphoniques : elles sont sous-estimées avec un taux $t(a)$. A la fin de la dernière transition, i.e. après la transition débutant en vague $v=5$, la situation réelle de la personne vis-à-vis du champ est supposée observée car ayant fait l'objet d'une enquête plus approfondie (en visite...). Les sorties de l'enquête recueillies lors de cette transition peuvent donc être des sorties ayant réellement eu lieu entre la cinquième et la sixième vague (se produisant avec la probabilité de sortie « réelle » $\pi(a, hc)$), mais elles peuvent aussi correspondre à des sorties antérieures observées avec retard. Deux types de sorties antérieures sont modélisées : celles correspondant à des sorties non-décelées lors des vagues téléphoniques, mais aussi celles concernant des sorties de ménages d'inactifs dont les données étaient précédemment imputées. Les probabilités associées à ces deux derniers types de sortie sont estimées en tenant compte du nombre de vagues téléphoniques ou du nombre d'imputations antérieures.
- Les autres non-réponses seront étudiées conditionnellement au fait d'être dans le champ, à partir des probabilités de base $\pi_t(a, anr)$. Ces probabilités de base correspondent à la probabilité d'« autre non-réponse » à une vague téléphonique (probabilité supposée ici indépendante de la vague) mais peuvent dépendre du statut d'activité initial a . La probabilité de non-réponse en sixième vague, est supposée plus élevée en raison des contraintes particulières liées à la dernière interrogation (visite, longueur...). Elle se déduit des probabilités d'autre non-réponse de base par un coefficient multiplicatif $1+vis(a)$. Le choix d'exprimer les probabilités d'« autre non-réponse » en se basant sur les probabilités obtenues lors des enquêtes téléphoniques (plutôt qu'en visite) se justifie notamment par la constance du protocole d'enquête téléphonique au cours de la période considérée, alors que le contenu de la dernière interrogation varie notamment d'une année à l'autre en raison des enquêtes complémentaires²⁵.

Mathématiquement, on a donc pour les transitions commençant aux vagues v 1 à 4 :

$$p(a,v, hc) = \pi(a, hc) * (1 - t(a)) \text{ et } p(a,v, anr) = \pi_t(a, anr)$$

Pour la transition entre les vagues 5 et 6, la probabilité d'autre non-réponse (pour les personnes dans le champ) s'écrit $p(a, 5, anr) = \pi_t(a, anr) * (1 + vis(a))$.

La formule donnant les probabilités de transition hors du champ entre les vagues 5 et 6 est plus complexe, car elle tient compte des sorties détectées en sixième vague mais ayant eu lieu plus tôt. Concrètement, pour cette transition, on commence par calculer la probabilité $ptel(a)$ que la personne soit en réalité sortie lors d'une des vagues téléphoniques immédiatement antérieures (en nombre n_{telav}) sans que cela ait été repéré lors de l'enquête par téléphone. On montre que²⁶ :

$$ptel(a) = 1 - q(a) * n_{telav} \text{ avec } q(a) = (1 - \pi(a, hc)) / (1 - \pi(a, hc) * (1 - t(a))).$$

²⁵ Citons deux autres raisons de ce choix : les probabilités de sortie lors des vagues téléphoniques peuvent être estimées à partir de données plus nombreuses, puisque portant sur quatre fois plus de transitions. Elles sont plus faibles, ce qui invite également à les prendre comme référence et à interpréter le coefficient $v(a)$ comme un surcroît de non-réponse lié aux particularités de la dernière vague.

²⁶ Le calcul s'obtient en constatant que la probabilité d'être effectivement dans le champ après la première vague téléphonique sachant que l'on est supposé être encore dans le champ est égale à $q(a)$.

On calcule ensuite la probabilité pimp qu'une personne toujours réellement dans le champ lors de sa dernière interrogation effective soit sortie lors une des nb14 vagues où ses données ont été imputées :

$$pimp(a)=1-(1-\pi(a,hc))^{nb14}.$$

La probabilité d'être observée comme sortant du champ en dernière vague est donc :

$$p(a,5,hc)=ptel(a)+(1-ptel(a))*pimp(a)+ (1-ptel(a))*(1-pimp(a))* \pi(a,hc).$$

Les paramètres utilisés dans les décompositions ci-dessus peuvent être estimés par maximum de vraisemblance. Les données utilisées pour l'estimation correspondent à l'empilement sur les cinq premières vagues des observations dont le statut d'activité est connu à la vague considérée et qui ne sont pas imputées à la vague suivante. Pour chacune de ces observations, une variable indiquant la situation observée à la vague suivante est construite. Cette variable (appelée sortie) prend la valeur 0 si la personne a un statut d'activité au sens du B.I.T. à la vague suivante, 1 si elle n'en a pas en raison d'une sortie de champ (changement de ménage, mobilité, logement hors champ) et 2 si elle n'en a pas en raison d'un autre motif de non-réponse (impossible à joindre, inapte absent de longue durée, refus, non-réponse partielle). La contribution de chaque observation à la log-vraisemblance ll est alors la suivante :

$$ll=(sortie=0)*\log[(1-p(a,v,hc))*(1-p(a,v,anr))]+ (sortie=1)*\log(p(a,v,hc)) + (sortie=2)*\log[(1-p(a,v,hc))*p(a,v,hc)]$$

La modélisation retenue - à visée essentiellement descriptive - conduit à des probabilités de sortie par statut d'activité, motif et vagues proches de celles effectivement observées²⁷. Elle facilite l'analyse comparée des probabilités de sortie des chômeurs et des actifs occupés et de leurs conséquences sur le biais de rotation.

²⁷ Cf. annexe pour une comparaison entre les probabilités de sortie par vague générées par le modèle et observées.

Tableau 6 – Probabilités de sortie (estimation par maximum de vraisemblance)

statut d'activité initial	sortie du champ			autre non-réponse (conditionnelle au fait d'être dans le champ)			taux de sortie total $\pi(a, hc) + (1 - \pi(a, hc)) * \pi_t(a, anr)$
	probabilité $\pi(a, hc)$	taux de sous-estimation en vague téléphonique	sous-estimation en points en vague téléphonique	probabilité $\pi_t(a, anr)$ (vague téléphonique)	taux d'augmentation en visite	augmentation en points en visite	
actif occupé	3,0%	10%	0,3%	4,4%	50%	2,2%	7,2%
chômeur	5,0%	5%*	0,2%	6,5%	18%*	1,1%	11,2%
Inactif	2,4%	6%	0,1%	4,3%	38%	1,6%	6,6%

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, personnes présentes à la première (« vague initiale ») et non imputées à la deuxième vague de la transition (et donc non imputées à la première non plus)..

Nombre d'observations = 764 152 ; -2 ln L= 493 850.

Tous les paramètres sont significatifs au seuil de 0,01% sauf les paramètres marqués '*' significatif au seuil de 1%.

La légère sous-estimation des sorties de champ lors des vagues téléphoniques a un impact négligeable sur le biais de rotation.

Commençons par étudier les sorties de champ. La probabilité de sortie de champ estimée $\pi(a, hc)$ diffère selon le statut d'activité de la personne à la vague initiale : elle est en moyenne de 3 % pour les actifs occupés, contre 5,0% pour les chômeurs, et ces valeurs sont très stables d'une année à l'autre²⁸. A priori, des sorties de champ plus fréquentes ne sont pas problématiques dans la mesure où elles sont supposées être remplacées par des entrées de champ également plus fréquentes (on verra plus tard ce qu'il en est). Mais actifs occupés et chômeurs diffèrent aussi par le taux de sous-estimation de leur sortie de champ lors des vagues téléphoniques : si en moyenne 10% des sorties de champ des actifs occupés ne semblent pas détectées en vague téléphonique, ce n'est le cas que pour 5% des sorties de champ des chômeurs²⁹, peut-être parce que ces derniers, lorsque leur ménage accepte de répondre, répondent moins souvent par proxy, ou peuvent consacrer plus de temps à l'enquêteur. En termes absolus toutefois, compte tenu de la plus forte probabilité de sortie de champ des chômeurs, ces sous-estimations se traduisent pour les deux catégories concernées par une sous-estimation moyenne comparable du taux de sortie de champ, d'environ 0,3% des

²⁸ Ainsi, si l'on réitère l'estimation sur différents sous-échantillon (correspondant à l'année de la vague courante), on obtient pour le taux de sortie de champ des actifs occupés des valeurs comprises entre 2,9% (sous-échantillon correspondant aux années 2003 et 2004) et 3,1% (sous-échantillon correspondant aux années 2006 et 2007). Pour les chômeurs, les probabilités de sortie de champ varient de 4,9% (années 2003 et 2004, année 2005) à 5,3% (année 2006 et 2007). Les variations selon le trimestre sont légèrement plus marquées pour les actifs occupés de 2,6% (deuxième trimestre) à 3,3% (troisième trimestre), tandis que le taux de sortie de champ des chômeurs reste quelque soit le trimestre égal à 5,0 ou 5,1%.

²⁹ La plus forte sous-estimation des taux de sortie de champ des actifs occupés s'observe également sur les sous-échantillons par année ou trimestre.

effectifs d'une vague téléphonique à l'autre, si bien que ce phénomène n'a en moyenne pas d'incidence sur le biais de rotation³⁰.

L'attrition liée aux autres non-réponses (refus, impossible à joindre...) est supérieure de 2,1 points chez les chômeurs pour les vagues téléphoniques, ce qui entraîne une baisse du taux de chômage même ordre entre deux vagues téléphoniques.

Les taux de sortie pour « autre non-réponse » (refus, impossible à joindre, absent de longue durée...) sont légèrement plus élevés, et sont eux aussi, variables selon la catégorie initiale au sens du B.I.T.. Entre deux vagues téléphoniques, ils sont en moyenne de 4,4% pour les actifs occupés de la vague initiale (restant dans le champ) contre 6,5% pour les chômeurs. Le surcroît de sorties chez les chômeurs s'observe quel que soit le motif de réponse détaillé, et n'est qu'à peine plus marqué pour les « impossibles à joindre » et les « refus ». Ces sorties - non compensées par des entrées – entraînent un biais de rotation. En effet³¹, les effectifs bruts de chômeurs diminuent à chaque vague d'environ 2,1 pourcents de plus que les effectifs d'actifs occupés, soit une baisse approximativement équivalente sur le taux de chômage entre deux vagues (ce qui représente -0,21 point pour un taux de chômage à 10 %). L'effet cumulé entre les vagues 1 à 6 est de près de 10%, soit un écart de taux de chômage entre les vagues extrêmes de -1% pour un taux de chômage initial de 10% (graphique 1). Cet écart est proche de celui observé, suggérant que cette attrition différentielle est un élément d'explication important de la tendance à la diminution du taux de chômage en fonction des vagues.

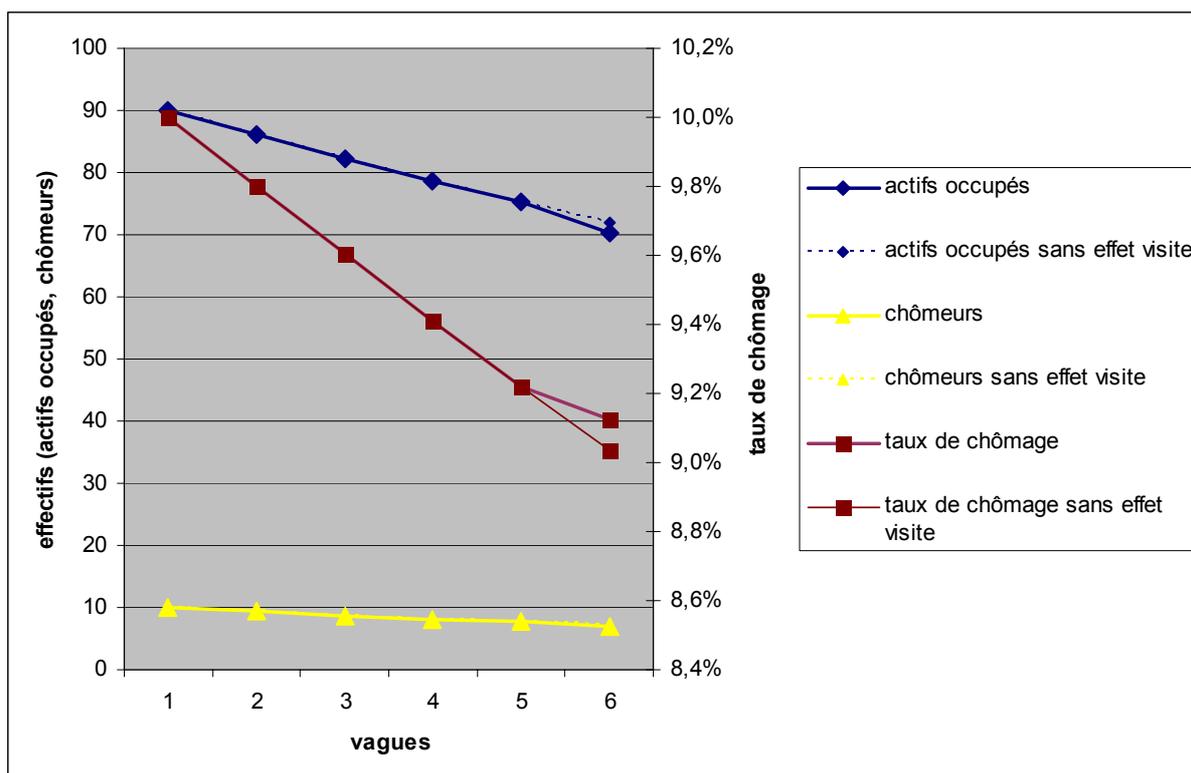
La probabilité de sortie pour « autre non-réponse » est plus élevée en dernière vague, en particulier chez les actifs occupés (+ 2,2 points), et elle varie selon les années

Les taux de sortie pour « autre non-réponse » sont particulièrement élevés en vague 6, car cette dernière enquête, en face à face, est plus longue et peut être suivie par une enquête complémentaire. L'augmentation de ces taux par rapport aux taux de réponses lors des vagues intermédiaires est particulièrement sensible pour les actifs occupés (hausse de 2,2 points du taux d' « autre non-réponse », soit + 50%), et est beaucoup plus faible pour les chômeurs (+1,1%, soit + 18%). Le différentiel de taux de sortie pour « autre non-réponse » entre actifs occupés et chômeurs est donc en moyenne deux fois plus faible en vague 6 qu'aux autres vagues. Son impact sur le biais de rotation est donc plus faible à cette vague là : il n'entraîne qu'une baisse de 1% du taux de chômage entre les vagues 5 et 6 (contre environ 2,1% aux autres vagues)(graphique 1).

Graphique 1 – Simulation : effet des taux de sorties pour autre non-réponse sur le taux de chômage

³⁰ La sous-estimation différentielle peut toutefois avoir un impact plus grand selon les périodes. Elle est notamment beaucoup plus élevée au troisième trimestre (du fait de sorties de champ plus nombreuses des actifs occupés, et d'une absence de sous-estimation des sorties pour les chômeurs à ce trimestre là), ce qui conduit à minorer de 0,04 point de pourcentage (pour un taux de chômage de 10%) l'évolution du taux de chômage ces trimestres-là.

³¹ En supposant qu'il n'y a pas de changement de catégorie d'activité et que les sorties d'enquête pour départ du champ sont parfaitement compensées.



Lecture : Supposons que l'on parte d'un effectif d'actif occupé en vague 1 de 90, et d'un effectif de chômeur de 10 (soit un taux de chômage de 10%), on aboutit au bout de cinq transitions sous l'effet des seuls motifs de non-réponses autres (estimés pour des vagues téléphoniques) à des effectifs de 72 et 7, et à un taux de chômage de 9,0% (courbes en pointillés). Si l'on tient compte en plus de la hausse de la probabilité de sortie pour autre non-réponse en 6^{ème} vague, on aboutit en sixième vague à des effectifs de 70 et 7, et à un taux de chômage de 9,1%

Compte tenu de la spécificité de la vague 6, et de la charge variable qu'elle implique pour l'enquête selon la présence et le thème de la complémentaire, il est intéressant d'examiner comment l'augmentation du taux pour « autre non-réponse » varie en fonction de l'année d'interrogation³².

³² Au-delà de l'hypothèse d'un effet désincitatif lié à la charge d'enquête supplémentaire pour l'enquêté, d'autres effets seraient possibles. Ainsi, des personnes pourraient choisir d'arrêter de répondre à l'enquête Emploi car se sentant peu concernées par le thème traité dans l'enquête complémentaire (annoncé dans la lettre-avis envoyée à tous les ménages). De manière plus indirecte, compte tenu des contraintes de temps pour la réalisation des enquêtes complémentaires (quelques semaines), certaines non-réponses pourraient aussi résulter de difficultés organisationnelles de l'enquêteur. Le nombre d'enquêtes que l'enquêteur peut réaliser en une journée est d'autant moins important que la durée moyenne de ces enquêtes est longue, mais aussi qu'elle est variable (l'enquêteur devant alors réserver a priori un temps important pour chaque rendez-vous, au cas où l'interrogation serait longue). Si l'enquêteur n'est alors pas en mesure d'utiliser l'ensemble de la période d'enquête prévue (en raison d'autres engagements en parallèle par exemple), un risque de non-réponse accru existe.

Tableau 7 – Probabilités de sortie pour autre non-réponse pour différents sous-échantillons (estimation par maximum de vraisemblance)

statut d'activité initial	2004				2005				2006 ou 2007			
	π_t (a,anr)	hausse en visite		non réponse totale en visite	π_t (a,anr)	hausse en visite		non réponse totale en visite	π_t (a,anr)	hausse en visite		
		relative	absolue			relative	absolue			relative	absolue	
actif occupé (1)	4,4%	41%	1,8%	6,2%	4,3%	63%	2,7%	7,1%	4,4%	45%	2,0%	
chômeur (2)	6,5%	6%	0,4%	6,9%	6,6%	16%	1,0%	7,7%	6,3%	29%	1,8%	
inactif	4,3%	30%	1,3%	5,6%	4,3%	43%	1,8%	6,1%	4,5%	37%	1,6%	
écart (2)-(1)	2,1%		-1,4%	0,7%	2,3%		-1,7%	0,6%	1,9%		-0,1%	

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, personnes présentes à la première (« vague initiale ») et non imputées à la deuxième vague de la transition (et donc non imputées à la première non plus). Restrictions supplémentaires selon les colonnes en fonction de l'année de collecte de la vague courante (et non de la vague de début).

C'est en 2004 que la non-réponse supplémentaire a été la plus limitée, et en 2005 qu'elle a été la plus forte. Signalons d'emblée que les raisons de ces variations ne sont probablement pas uniquement liées à l'enquête Emploi elle-même, mais peuvent être liées à la charge que les enquêteurs ou les divisions « enquête ménage » des directions régionales de l'Insee ont par ailleurs (ainsi, en raison du début du nouveau recensement rénové de la population en 2004, les enquêtes ménages ont été moins nombreuses que d'habitude en 2004, laissant aux enquêteurs plus de temps à consacrer à l'enquête Emploi). Par ailleurs, l'impact des enquêtes complémentaires elles-mêmes sur le taux de réponse dépend de plusieurs facteurs, comme leur champ, leur thème, leur longueur. Par exemple, les enquêtes complémentaires de 2004 s'adressaient quasi exclusivement aux actifs occupés, ce qui peut contribuer à expliquer la faible augmentation du taux de non-réponse des chômeurs par rapport aux années suivantes où ils étaient concernés par les complémentaires. La hausse des « autres non-réponses » en vague 6 a été maximale en 2005, année où deux enquêtes supplémentaires étaient réalisées en sixième vague, sur des thèmes assez différents. La hausse de la non-réponse a été particulièrement élevée chez les actifs occupés, concernés par les deux enquêtes. Des analyses plus complètes sur l'impact des complémentaires sur les taux de non-réponse seraient intéressantes à mener, mais sortent du cadre de cette étude. Du strict point de vue du biais de rotation, l'élément à étudier est le différentiel entre le taux de réponse des actifs et des chômeurs. C'est en 2005 que ce différentiel est le plus faible. Cette année-là en effet, la hausse de la non-réponse observée en vague 6 a été beaucoup marquée chez les actifs occupés (+2,7 points) que chez les chômeurs (+1%), ce qui a porté le taux de non-réponse des actifs occupés à 7,1%, et lui a presque permis de rejoindre celui des chômeurs (7,7%).

Tableau 8 – Questionnements supplémentaires en dernière vague selon l'année de cette dernière : enquêtes complémentaires et modules ad hoc de l'enquête Emploi

ANNEE	MODULE ADHOC	ENQUETE COMPLEMENTAIRE
2004	Organisation du temps de travail Perspective de croissance et de créations d'emploi (juin et juillet, 2500 personnes prévues, inférieure à 5 mn)	Perception du travail dans le secteur privé selon le type de contrat (CDI, CDD, intérim) -DARES (trois derniers trimestres de l'année)
2005	Conciliation vie familiale et vie professionnelle (personnes de 15 à 64 ans, 32 000 répondants)	Conditions de travail – DARES (personnes ayant un emploi salarié ou non ; 20 mn ; max. deux personnes par ménage, 19000 personnes)
2006	Passage à la retraite (personnes de 50 à 69 ans, 12 000 répondants)	Formation continue -CEREQ
2007		Accidents du travail et maladies professionnelles

Les entrées

S'ajoutant aux effets des règles du protocole d'enquête, l'entrée en vague 2 seulement de certains enquêtés conduit à des entrées 1,5 fois plus nombreuses à cette vague par rapport aux trois vagues suivantes ainsi qu'à une composition particulière des entrants à cette vague-là.

Au regard du protocole d'enquête « théorique », les entrées entre les vagues 1 et 2 devraient *a priori* être de même ordre de grandeur que lors des 3 transitions suivantes, dans la mesure où le repérage des personnes lors des quatre vagues téléphoniques se fait de la même manière. Les différences constatées confirment donc qu'au-delà des strictes règles du protocole, des effets liés au terrain sont visibles. Le taux d'entrées est ainsi 1,5 fois plus élevé en vague 2 qu'aux deux vagues suivantes. Ce surcroît d'entrées est pour l'essentiel dû à l'interrogation réussie en vague 2 de ménages supposés déjà dans le champ en première vague mais n'ayant pu être interrogés alors (impossibles à joindre...). C'est surtout pour les inactifs puis pour les actifs occupés que ce surcroît d'entrées dans l'enquête est prononcé, si bien que la composition par statut d'activité des nouveaux entrants varie en fonction de la vague (cf. **tableau en annexe**). Le taux de chômage des entrants en vague 2 (12,4%) est notamment plus faible que celui des entrants aux autres vagues³³. Ce résultat est cohérent avec celui des enquêtes non-répondants, montrant que les non-répondants de première vague sont moins souvent chômeurs.

Décomposition des probabilités d'entrées

³³ Ce taux de chômage reste par ailleurs supérieur à celui observé dans l'ensemble de la population, les chômeurs étant en moyenne plus nombreux parmi les entrants, en partie en raison de leur plus grande mobilité résidentielle.

Pour essayer de décomposer les mécanismes à l'œuvre en restant au plus près des étapes de la collecte, on modifie (temporairement) la perspective adoptée jusqu'à présent de plusieurs façons :

- en travaillant au niveau du logement et non plus de l'individu : en effet, la probabilité d'entrer dans les répondants n'est connue qu'au niveau du logement (car on ne connaît pas la composition des ménages restant non répondants) ;
- en séparant dans l'analyse quatre grands types/étapes d'entrées : entrée du logement dans les logements répondants, entrée d'un ménage dans les ménages répondants (pour un logement précédemment répondant, après un changement de ménage), entrée d'une personne dans un ménage précédemment répondant, logement neuf ou nouveau (dans les faits, on se s'intéressera en détail qu'au premier type d'entrée)
- en calculant des probabilités d'entrée (nombre d'entrées rapportées aux logements susceptibles d'entrer) et non des taux d'entrées (nombre d'entrées rapportées aux logements déjà présents).

Probabilité qu'un logement devienne répondant

Commençons par étudier la probabilité qu'un logement sans aucun répondant à la vague initiale d'une transition devienne répondant à la vague finale de cette transition^{34,35}.

En moyenne, toutes transitions confondues et par sous-échantillon, 2804 logements n'ont aucun répondant en vague initiale, se répartissant en 57% logements hors champ et 43% logements dans le champ mais n'ayant pas répondu³⁶. Le nombre de logements sans répondant en vague initiale est légèrement supérieur en vague 1 (3029 logements, soit 8% de plus que la moyenne) alors qu'il reste proche de 2750 aux quatre vagues suivantes. La part des logements hors champ est particulièrement faible en vague 1 (51%) et augmente jusqu'à la vague 5 (60%)³⁷.

Toutes vagues confondues, 13% des logements sans répondant à une vague deviennent répondants à la vague suivante³⁸. Mais là encore, cette moyenne cache d'importantes disparités, entre logements initialement hors champ (qui n'ont qu'une probabilité d'entrée de 5%) et logements initialement absents pour un autre motif de non-réponse (probabilité d'entrée de 25%), ou en fonction de la vague (probabilité d'entrée globale minimale - 9% - entre les vagues 4 et 5 et maximale - 18% - entre les vagues 1 et 2).

³⁴ On exclut à dessein dans cette première étape d'analyse des entrées le cas des logements neufs ou nouveaux repérés en sixième interrogation (on choisit de ne pas les considérer comme étant « sans enquêtés en vague initiale » du fait de leur absence de la liste des logements de cette vague initiale, ceci afin de rester au plus près du déroulement de la collecte en ne les incluant pas. On exclut également les logements neufs non tirés au sort de l'ensemble de l'analyse, car ceux-ci ne devant jamais être interrogés, leur inclusion conduirait à des probabilités d'entrée trop faibles.

³⁵ Pour cette analyse, le statut de répondant du logement n'a pas été défini à partir du code résultat ('10', '11'...) mais à partir de l'existence ou non de personnes répondants (i.e. avec statut d'activité défini). Ceci afin d'éviter les problèmes d'estimations susceptibles d'être posés dans les très rares cas où le logement est répondant au sens du code résultat, mais sans aucun individu avec code d'activité connu (le logement est alors reclassé en refus).

³⁶ Soit respectivement 1592 et 1212 logements en moyenne par échantillon.

³⁷ C'est pour tenir compte de ces variations qu'il a semblé préférable de baser la présente analyse sur des probabilités d'entrées (en rapportant donc les entrées au nombre de logements non-répondants, et non en les exprimant en pourcentage des logements répondants), et de la détailler en fonction du résultat d'enquête initial.

³⁸ Il est à noter que le pourcentage de logements devenant répondants dépend fortement du périmètre des logements pris en compte dans le calcul. Ici par exemple, tous les logements sont pris en compte (à l'exception des logements neufs non tirés au sort), et non les seuls logements figurant dans le carnet de tournée de l'enquêteur. La prise en compte des résidences secondaires dans le périmètre a un fort impact à la baisse sur le taux de réponse, qui sinon s'élèverait à 17% (et non 13 %).

Tableau 9 – Nombre moyen de logements sans répondant et nombre total d'enquêtés dans ces logements à la vague suivante, selon la vague de départ

vague de départ de la transition / statut du logement à la vague de départ	nombre de logements présents mais sans répondant à la vague de départ						nombre total d'enquêtés à la vague suivante dans ces logements					
	1	2	3	4	5	moy. 1 à 5	1	2	3	4	5	moy. 1 à 5
Ensemble	3029	2765	2738	2721	2767	2804	1147	729	693	567	969	821
hors champ	1530	1562	1597	1621	1650	1592	146	126	124	114	286	159
logement vacant	655	653	669	676	691	669	137	119	118	110	232	143
logement occasionnel	94	99	103	109	111	103	7	6	6	4	14	8
Résidence secondaire	707	737	751	763	773	746	0	0	0	0	38	8
<i>Logement en '14' à la vague finale</i>	<i>707</i>	<i>737</i>	<i>751</i>	<i>763</i>	<i>0</i>	<i>592</i>						
<i>Logement pas en '14' à la vague finale</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>773</i>	<i>155</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>38</i>	<i>8</i>
autre hors champ	73	73	73	74	75	74	1	1	0	0	2	1
<i>Logement en '14' à la vague finale</i>	<i>15</i>	<i>15</i>	<i>16</i>	<i>17</i>	<i>0</i>	<i>13</i>						
<i>Logement pas en '14' à la vague finale</i>	<i>58</i>	<i>57</i>	<i>57</i>	<i>56</i>	<i>75</i>	<i>61</i>	<i>1</i>	<i>1</i>	<i>0</i>	<i>0</i>	<i>2</i>	<i>1</i>
autre non-réponse	1499	1203	1141	1100	1117	1212	1001	602	568	453	683	661
Impossible à joindre	794	700	657	622	627	680	550	417	407	319	477	434
absent de longue durée	278	150	127	115	115	157	245	105	101	90	104	129
Inapte	109	102	101	107	109	106	36	18	11	11	31	21
Refus	318	251	256	256	267	269	171	62	49	32	71	77

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, logements présents dans l'échantillon à la vague n et à la vague n+1 hors logements neufs non tirés au sort, soit 168 232 logements ou 2804 en moyenne par vague et sous-échantillon.

Lecture : En moyenne sur les sous-échantillons introduits entre le 1^{er} trimestre 2003 et le 4^{ème} trimestre 2005, 3029 logements présents dans l'échantillon en vague 1 n'ont aucun répondant à cette vague. Dans ces mêmes logements, 1147 personnes de tous âges ont un statut d'activité connu en vague 2.

Les logements neufs n'étant pas pris en compte à cette étape de l'analyse, la seule disposition explicite du processus de collecte pouvant être à l'origine de variations en fonction de la vague est le traitement des résidences secondaires, qui, selon le protocole, ne peuvent changer de statut qu'en sixième vague³⁹. Il est toutefois très rare que ces résidences entrent dans le champ ou deviennent répondantes, ce qui explique que si l'on répartit les nombres de logements concernés sur toutes les vagues, les probabilités d'entrées selon les vagues sont peu modifiées⁴⁰. Bien que relativement mineure, cette correction sera conservée dans la suite de l'analyse pour pouvoir raisonner « à dispositif explicite de collecte constant ».

³⁹ En effet, les logements repérés comme tels en première vague sont supposés rester des résidences secondaires aux quatre vagues suivantes. Ils ne figurent donc pas dans la liste des logements confiés à l'enquêteur à ces vagues intermédiaires. Dans les fichiers de données, un code résultat égal à '14' leur est imputé à ces vagues. Dans les données figurent également quelques de logements ayant un résultat de collecte en vague 1 valant 'autre hors champ' et ayant un code '14' aux quatre vagues suivantes, probablement suite à des recodifications de code résultat datant du début de l'enquête Emploi. Ils feront par la suite l'objet des mêmes traitements que les résidences secondaires dans les statistiques descriptives.

⁴⁰ L'impact sur la probabilité moyenne d'être répondant (pour l'ensemble des logements sans répondants en vague initiale) est de -0,6 point entre les vague 5 et 6 de 0,14 point aux quatre autres transitions.

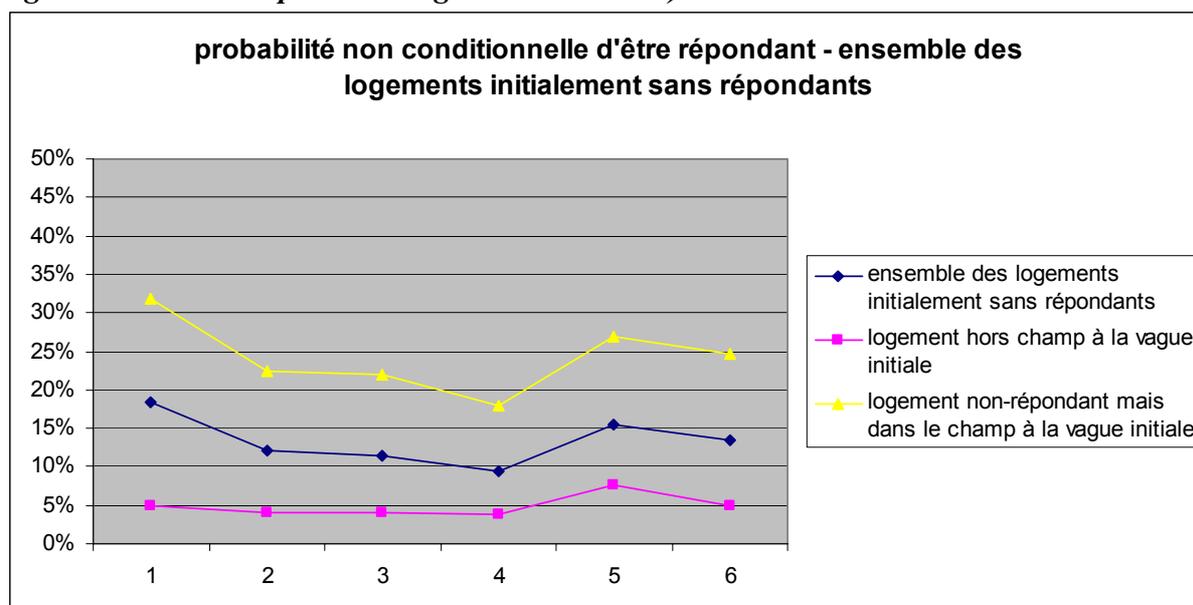
Tableau 10 – Probabilité pour qu'un logement sans répondant ait au moins un répondant à la vague suivante (après redressement des logements non ré-enquêtés en vague intermédiaire)

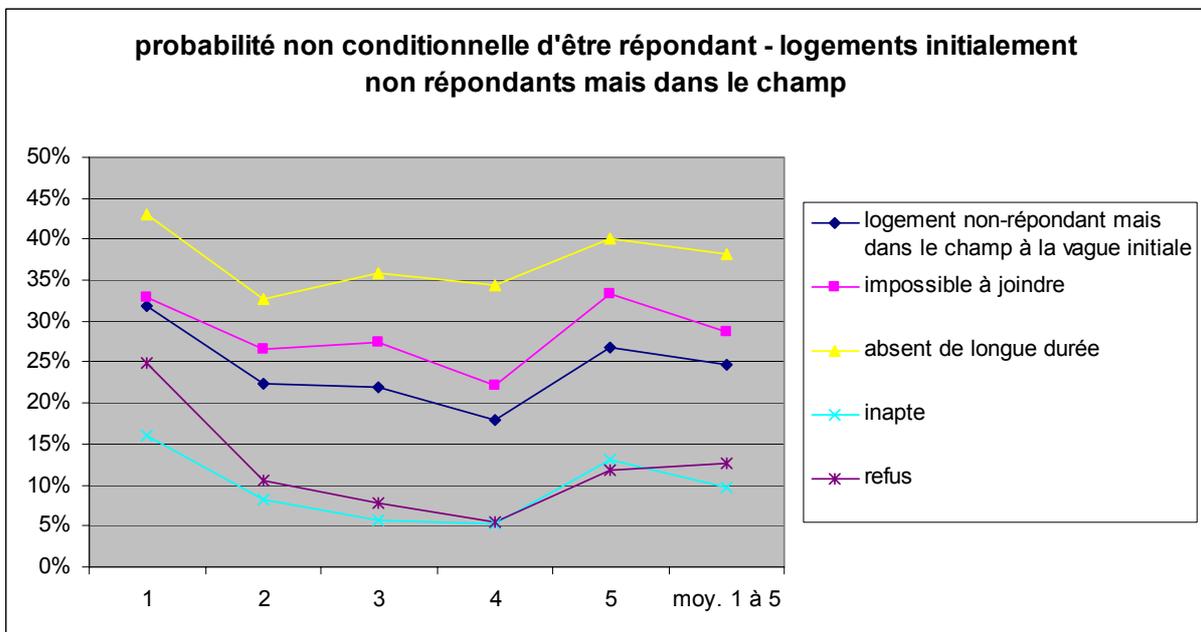
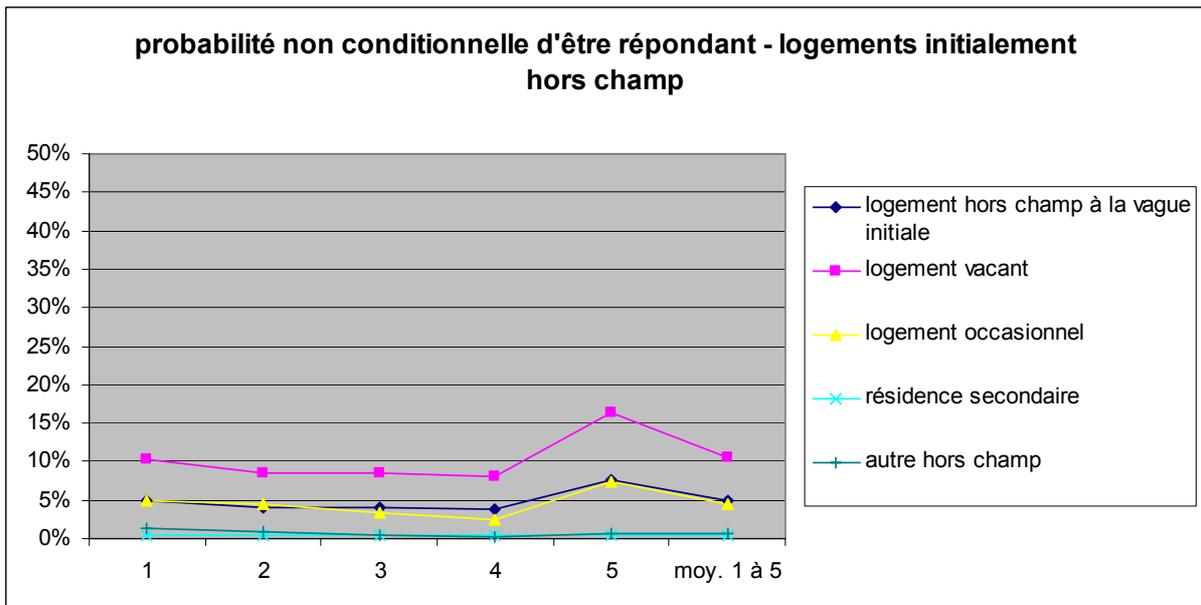
Vague de départ de la transition / statut du logement à la vague de départ	probabilité d'être répondant en n+1 (non conditionnellement à la présence dans le champ en n+1)					
	1	2	3	4	5	moy. 1 à 5
ensemble	18%	12%	12%	10%	15%	13%
hors champ	5%	4%	4%	4%	8%	5%
logement vacant	10%	8%	9%	8%	16%	10%
logement occasionnel	5%	5%	3%	3%	7%	5%
résidence secondaire	1%	1%	0%	0%	0%	1%
autre hors champ	1%	1%	1%	0%	1%	1%
autre non-réponse	32%	22%	22%	18%	27%	25%
impossible à joindre	33%	27%	27%	22%	33%	29%
absent de longue durée	43%	33%	36%	34%	40%	38%
Inapte	16%	8%	6%	5%	13%	10%
Refus	25%	11%	8%	5%	12%	13%

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, logements présents dans l'échantillon à la vague n et à la vague n+1 hors logements neufs non tirés au sort, soit 168 232 logements ou 2804 en moyenne par vague et sous-échantillon.

Graphiques 2 – Probabilité d'être répondant à la vague suivante (non conditionnellement à la présence dans le champ), en fonction de la vague de départ de la transition et du motif d'absence de réponse du logement à cette vague de départ (après redressement des logements non ré-enquêtés en vague intermédiaire)





Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, logements présents dans l'échantillon à la vague n et à la vague n+1 hors logements neufs non tirés au sort, soit 168 232 logements ou 2804 en moyenne par vague et sous-échantillon.

Le profil de la probabilité qu'un logement devienne répondant peut être caractérisé de la manière suivante :

- Quel que soit le motif d'absence de réponse initial, la probabilité d'être répondant est particulièrement élevée lors de la première transition, mais cet effet est beaucoup plus marqué pour les logements initialement non-répondants du champ ;
- La probabilité d'être répondant varie en général moins au cours des trois transitions suivantes : elle est, selon les motifs de non-réponse initiaux, décroissante (de manière marquée pour les logements en refus) ou relativement constante (en particulier pour les logements initialement vacants).

- Quel que soit le motif d'absence de réponse initial, la probabilité d'être répondant remonte très fortement lors de la dernière transition (enquête en visite), ceci étant particulièrement marqué pour les logements initialement absents du champ.

La décomposition de cette probabilité de réponse en deux étapes – probabilité de présence dans le champ puis probabilité de réponse conditionnellement à la probabilité de réponse – permet de préciser les mécanismes à l'œuvre.

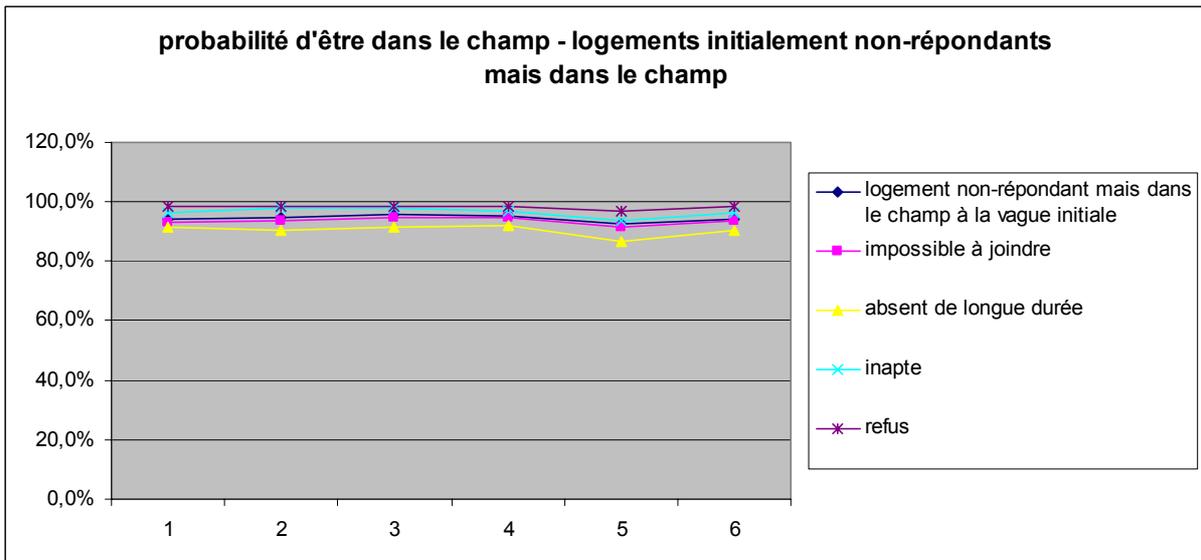
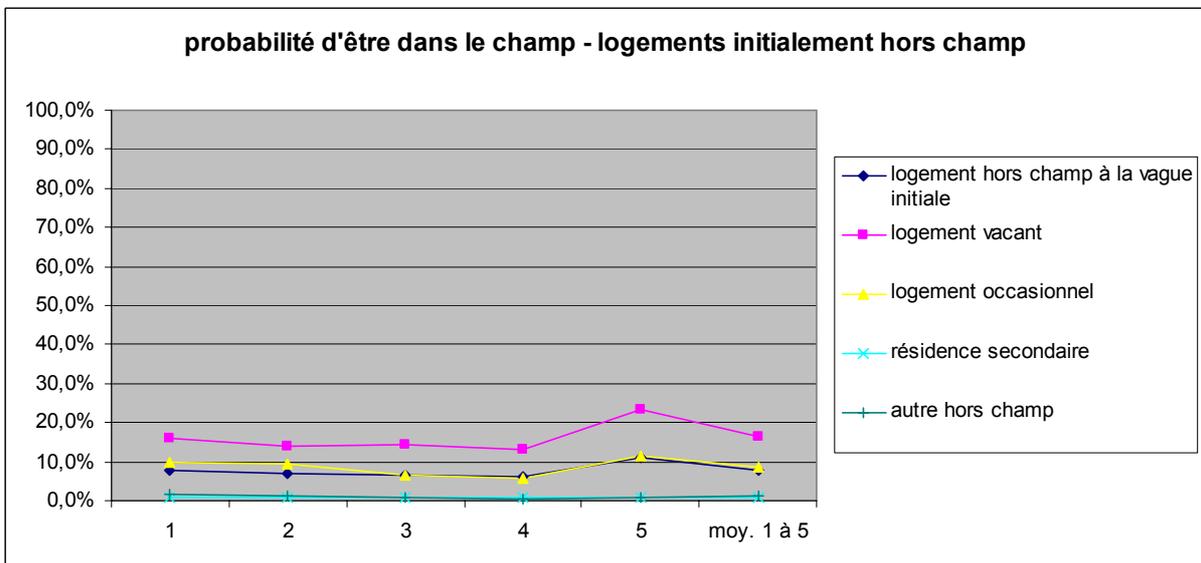
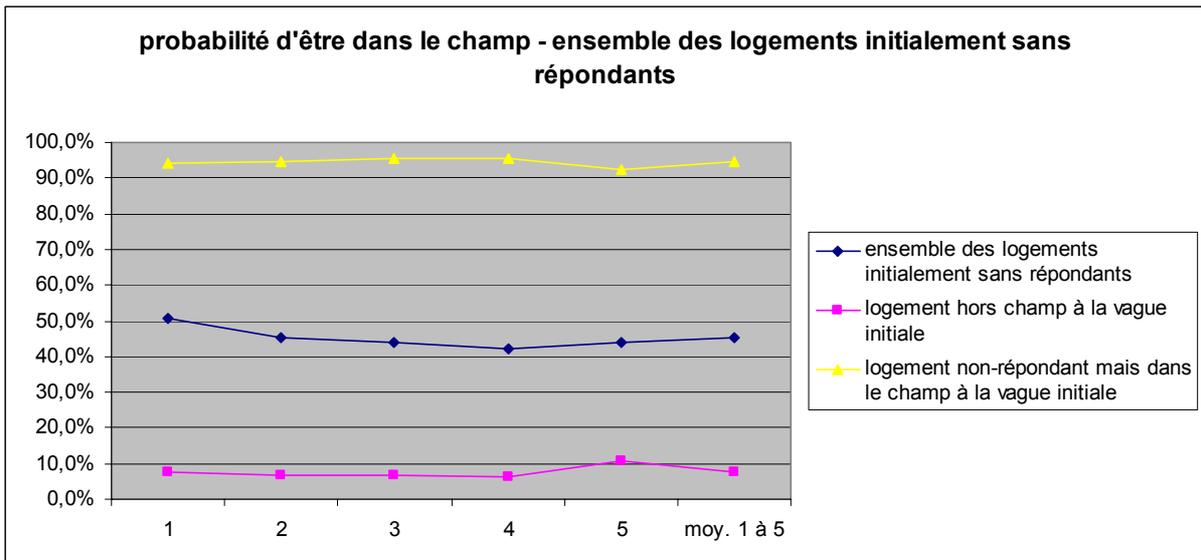
Tableau 11 – Probabilité pour qu'un logement sans répondant soit dans le champ à la vague suivante, ou soit répondant conditionnellement à sa présence dans le champ à la vague suivante (après redressement des logements non ré-enquêtés en vague intermédiaire)

vague de départ de la transition / statut du logement à la vague de départ	probabilité d'être dans le champ en n+1						probabilité d'être répondant en n+1 (conditionnellement à la présence dans le champ en n+1)					
	1	2	3	4	5	moy. 1 à 5	1	2	3	4	5	moy. 1 à 5
Ensemble	50,5%	45,1%	43,7%	42,3%	43,8%	45,2%	36%	27%	26%	23%	35%	30%
hors champ	7,8%	6,9%	6,7%	6,3%	10,9%	7,8%	64%	60%	61%	61%	70%	64%
logement vacant	15,9%	14,1%	14,2%	13,3%	23,4%	16,2%	65%	60%	61%	61%	70%	64%
logement occasionnel	9,8%	9,3%	6,6%	5,7%	11,5%	8,6%	50%	49%	52%	45%	65%	53%
résidence secondaire	0,7%	0,7%	0,7%	0,7%	0,7%	0,7%	71%	71%	71%	71%	71%	71%
autre hors champ	1,8%	1,4%	0,6%	0,4%	1,0%	1,0%	71%	63%	90%	55%	70%	70%
autre non-réponse	94,0%	94,7%	95,5%	95,4%	92,5%	94,4%	34%	24%	23%	19%	29%	26%
impossible à joindre	92,8%	93,9%	94,7%	94,4%	91,4%	93,4%	35%	28%	29%	24%	36%	31%
absent de longue durée	91,4%	90,2%	91,6%	91,8%	86,6%	90,6%	47%	36%	39%	37%	46%	42%
Inapte	96,1%	97,8%	98,0%	97,0%	93,7%	96,5%	17%	9%	6%	5%	14%	10%
Refus	98,5%	98,7%	98,6%	98,7%	97,0%	98,3%	25%	11%	8%	6%	12%	13%

Champ : logements présents dans l'échantillon à la vague n et à la vague n+1 hors logements neufs non tirés au sort, soit 168 232 logements ou 2804 en moyenne par vague et sous-échantillon.

Note : les probabilités de réponse en italique concernent chacune une dizaine de logements et sont donc non significatives.

Graphiques 3 – Probabilité d'être dans le champ à la vague suivante, en fonction de la vague de départ et du motif d'absence de réponse du logement à la vague de départ (après redressement des logements non ré-enquêtés en vague intermédiaire)



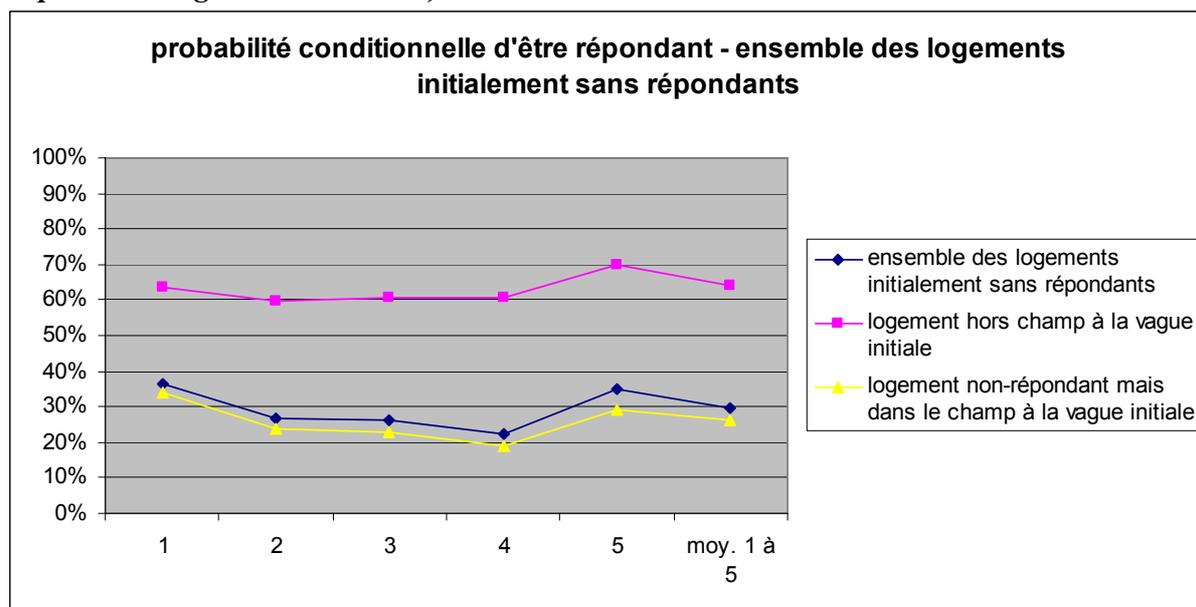
Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées
 Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, logements présents dans l'échantillon à la vague n et à la vague n+1 hors logements neufs non tirés au sort, soit 168 232 logements ou 2804 en moyenne par vague et sous-échantillon.

Normalement, une fois corrigés les effets des logements non-réinterrogés, la probabilité d'être dans le champ la vague suivante ne devrait plus varier selon la vague⁴¹. Même si elle reste effectivement assez stable, elle se caractérise néanmoins par les phénomènes suivants :

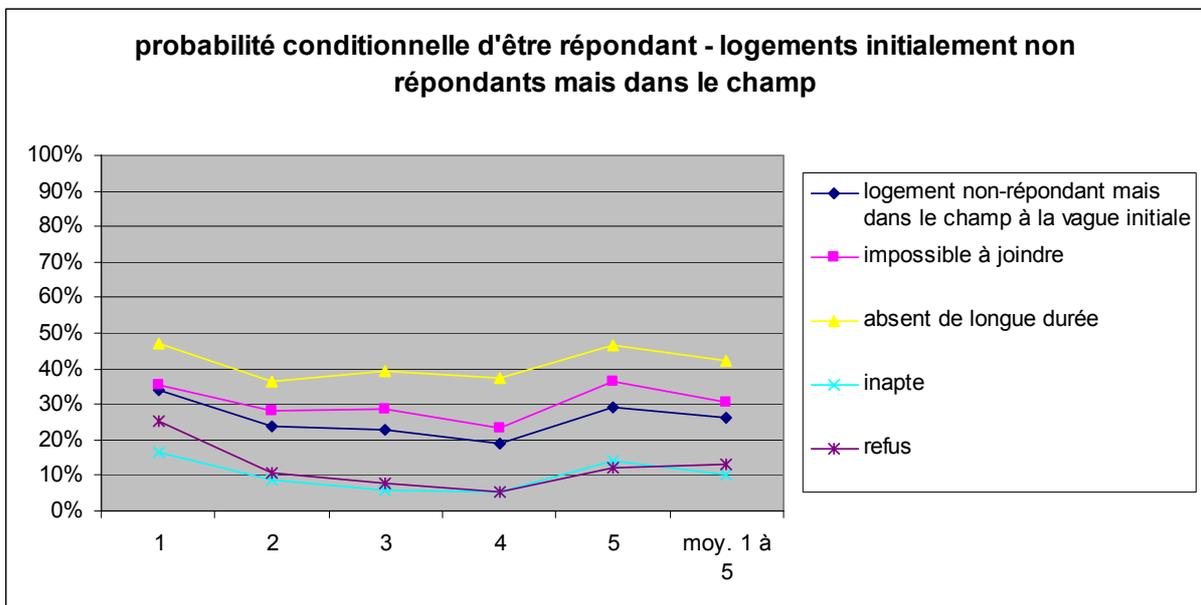
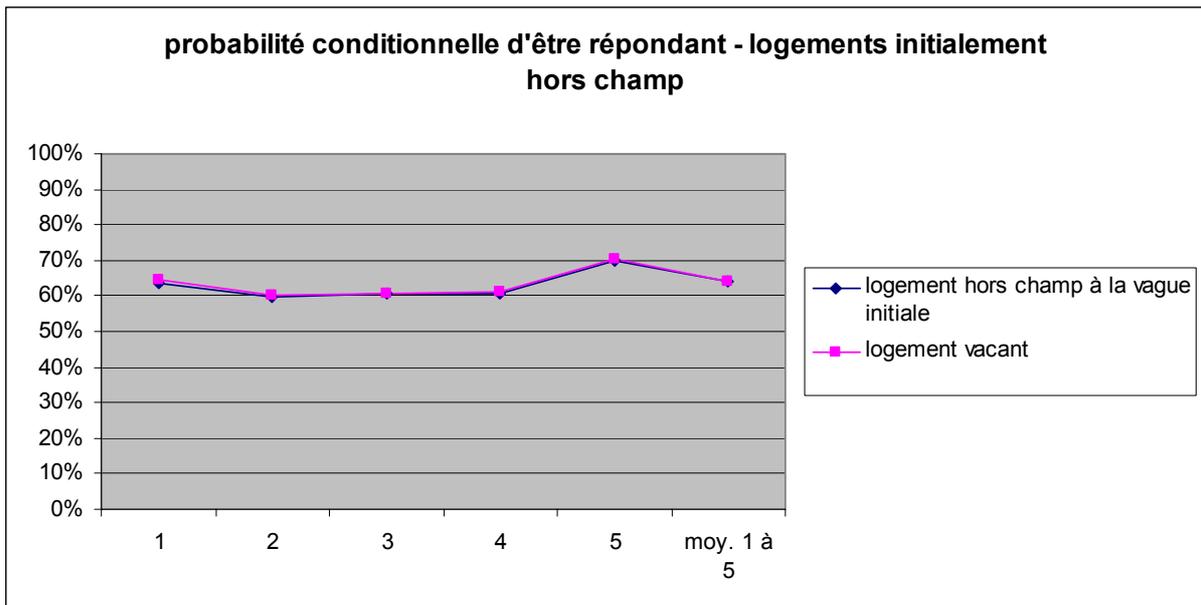
- La probabilité de changement de situation vis-à-vis du champ diminue au cours des quatre premières transitions, et ce pour la quasi-totalité des motifs initiaux d'absence de réponse : ainsi, la probabilité d'entrer dans le champ des logements initialement hors champ diminue, et la probabilité de quitter le champ pour les logements non-répondants du champ diminue également.
- Lors de la cinquième et dernière transition, la probabilité d'entrée dans le champ semble subir une correction : elle augmente fortement pour les logements initialement hors champ, et baisse fortement pour les autres.

Ces caractéristiques suggèrent que, comme pour les sorties, le statut vis-à-vis du champ est imparfaitement observé en vague téléphonique, avec une tendance à la sous-estimation des changements de statut, et donc une correction de ce statut lors de l'enquête par visite.

Graphiques 4 – Probabilité d'être répondant à la vague suivante (conditionnellement à la présence dans le champ), en fonction de la vague de départ et du motif d'absence de réponse du logement à la vague de départ (après redressement des logements non ré-enquêtés en vague intermédiaire)



⁴¹ Au premier ordre tout au moins (i.e. si l'on néglige les effets de compositions).



Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, logements présents dans l'échantillon à la vague n et dans le champ à la vague n+1.

Quant à elle, la probabilité de réponse conditionnelle à la présence dans le champ est :

- plus élevée lors de la première transition, en particulier pour les logements initialement non-répondants mais dans le champ ;
- globalement relativement stable lors des trois transitions suivantes mais légèrement décroissante pour les logements initialement non-répondants mais dans le champ ;
- en forte augmentation (de l'ordre d'une dizaine de points) lors de la dernière transition correspondant à l'enquête finale en visite.

Ce dernier résultat rappelle celui mis en évidence lors de l'analyse des sorties, à savoir que le fait qu'il y ait une enquête en visite avec éventuel questionnaire complémentaire a un impact sur les probabilités de réponse, qui dans le cas des entrées apparaît positif.

Quant à la baisse de probabilité de réponse pour les logements initialement non-répondants au cours des transitions intermédiaires, deux phénomènes (liés) pourraient l'expliquer :

- des effets de composition des logements sans répondant (plus on avance dans les vagues d'enquêtes, moins les logements non-répondants sont enclins à répondre) ;
- des effets de lassitude (ou d'apprentissage) des logements concernés ou des enquêteurs, qui conduiraient à une probabilité de réponse diminuant avec le nombre de non-réponses précédentes.

Dans la suite, pour simplifier et par symétrie avec les choix effectués pour décrire les sorties, on négligera les variations des probabilités d'entrée entre les deuxième et quatrième transitions.

Au vu de ce qui précède, les hypothèses retenues pour la modélisation des probabilités de réponse pour les logements initialement sans réponse seront les suivantes (cf. encadré pour la modélisation mathématique) :

- Comme pour les sorties, on sépare la probabilité d'être dans le champ en vague n+1 et la probabilité de réponse conditionnellement à la présence dans le champ afin d'interpréter plus facilement les effets modélisés ;
- Pour être cohérent avec cette modélisation en deux étapes, il est logique de supposer que les probabilités concernant les logements initialement hors champ et celles concernant les logements initialement sans réponse pour un autre motif puissent répondre à des paramètres différents (respectivement indicés par *lhc* et *anr*). On supposera de plus que la probabilité d'être dans le champ en n+1 puisse différer au sein des logements initialement hors champ entre les résidences secondaires (*log=sec*) et les autres⁴².
- Comme le statut vis-à-vis du champ (s'il était parfaitement observé) ne devrait pas varier en fonction de la vague à protocole donné, la probabilité d'être dans le champ à la vague suivante ne devrait varier qu'entre les vagues 5 et 6 (du fait de l'impossibilité d'entrer dans le champ avant la dernière transition pour les résidences secondaires). Pour tenir compte du fait que la probabilité d'être dans le champ à la vague suivante est dans les faits a-typique en vague 5 pour tous les motifs d'absence de réponse, on ajoute l'hypothèse que la présence dans le champ à la vague suivante est imparfaitement observée au téléphone (taux de sous-estimation de la présence dans le champ en vague suivantes de tel_{lhc} ou tel_{anr} ⁴³ selon le statut initial du logement, et que les mouvements vis-à-vis du champ qui seraient non détectés lors des vagues téléphoniques seraient repérés lors de la visite finale.
- La probabilité de réponse, conditionnellement à la présence dans le champ en n+1, est supposée varier en fonction du type d'enquête (taux d'augmentation vis_{lhc} ou vis_{anr} pour les enquêtes en visite). Compte tenu du profil par vague de cette probabilité de réponse, il est également supposé qu'elle est majorée d'un taux $v2_{lhc}$ ou $v2_{anr}$ lors de la transition entre les deux premières vagues. Comme signalé ci-dessus, on a en revanche renoncé à essayer d'introduire en supplément une attrition de la probabilité de réponse en fonction du nombre de non-réponse précédent⁴⁴.

⁴² Les autres paramètres concernant les résidences secondaires étant pris égaux à ceux des autres logements hors champ.

⁴³ tel_{lhc} peut être interprété comme le taux de logements initialement hors champ et entrant dans le champ dont l'entrée dans le champ n'a pas été repérée, et tel_{anr} comme la proportion de logements initialement absents pour un autre motif et sortant du champ dont la sortie n'a pas été repérée.

⁴⁴ Des tentatives d'estimations incluant des effets d'attrition (séparés selon le type de non-réponse initiale ou non) conduisent à des paramètres estimés peu réalistes (fort effet d'attrition, couplé avec un effet négatif de la vague 2 et un très fort effet visite). En effet, le modèle comporte alors trop de paramètres par rapport aux

Encadré – Modélisation des probabilités de réponse des logements non répondants

Probabilité que le logement soit dans le champ en n+1

La probabilité d'être dans le champ à la vague n+1 vaut :

. pour les quatre premières transitions, la probabilité observée d'entrer dans le champ s'obtient à partir de la probabilité « vraie » d'entrer dans le champ ($pc_x = pc_{lhc} * 1_{log=lhc \text{ et } log \neq sec} + pc_{sec} * 1_{log=sec} + pc_{anr} * 1_{log=anr}$) par prise en compte de taux de sous-estimation des probabilités d'être dans le champ à la vague suivante lors des vagues téléphoniques :

$pc = (pc_{lhc} * 1_{log=lhc \text{ et } log \neq sec} + pc_{sec} * 1_{log=sec}) * (1 - tel_{lhc}) + pc_{anr} * 1_{log=anr} * (1 - tel_{anr})$
où log est le statut du logement lors de la vague initiale de la transition (hors champ hors résidence secondaire/ résidence secondaire / autre motif de non-réponse) et $1_{log=x}$ les indicatrices permettant de séparer les différentes catégories x de logements.

. pour la dernière transition, il faut d'abord tenir compte des mouvements vis-à-vis du champ non repérés lors des vagues téléphoniques ayant précédé la vague en visite. Si nbnonreptelav est le nombre d'enquêtes téléphoniques restées sans réponse précédant immédiatement la vague ⁴⁵, et si l'on introduit : $q_x = (1 - pc_x) / (1 - pc_x * (1 - tel_x))$ ⁴⁶, alors la probabilité qu'un logement de catégorie x⁴⁷ soit déjà entré dans le champ lors d'une vague téléphonique sans que cela soit identifié est $dejaenttel = (1 - q_x * nbnonreptelav)$.

Il faut également tenir compte de la probabilité pour une résidence secondaire, d'être entrée dans le champ lors des vagues téléphoniques sans que cela puisse être observé compte tenu du protocole : $dejaent14 = 1 - (1 - pc_x) * nb14$, où nb14 est le nombre de vagues immédiatement précédentes lors desquelles le logement a été considéré comme résidence secondaire⁴⁸.

Au total, la probabilité d'être repéré comme entrant dans le champ lors de la dernière transition (que cette entrée dans le champ se fasse à ce moment là ou ait eu lieu sans être repérée avant) est :

$pc = dejaenttel + (1 - dejaenttel) * dejaent14 + (1 - dejaenttel) * (1 - dejaent14) * pc_x$

Probabilité de réponse du logement en n+1 sachant qu'il est dans le champ à cette date

informations disponibles (taux de réponse en fonction de la vague, du statut initial du logement, et de manière marginale, du nombre de réponses téléphoniques antérieures et du nombre de classements en '14' antérieurs. Même si ce jeu de paramètres est peu utilisable par la suite, ces estimations semblent indiquer que l'attrition dépend plus du nombre de non-réponses antérieures que du rang de la vague.

⁴⁵ On ne tient compte que des vagues téléphoniques à la fois sans réponse et immédiatement antérieures. En effet, si le logement a répondu par le passé à une vague, son statut vis-à-vis du champ est supposé avoir été bien repéré à cette vague. Le taux d'erreur sur la présence dans le champ ne s'applique qu'à la première vague sans réponse. Ce taux d'erreur est supposé se cumuler au fur et à mesure que les vagues passent.

⁴⁶ q est la probabilité que le logement soit vraiment hors champ en n+1 conditionnellement au fait qu'il soit considéré comme étant encore hors champ en n+1.

⁴⁷ parmi les logements sans réponse au début de la dernière transition

⁴⁸ C'est la présence de ce terme qui a conduit à distinguer pc_{sec} de pc_{lhc} . En effet, une première modélisation n'ayant pas introduit cette distinction conduisait à un pc_{lhc} moyen égal à 7,8% (contre et 14,1% pour les logements initialement hors champ hors résidences secondaires et 0,7% pour les résidences secondaires – cf. infra) et donc à des probabilités d'entrer dans l'échantillon très fortes pour les résidences secondaires en dernière vague du fait du passage à la puissance 4 (la plupart des résidences secondaires étant repérées en vague 1, et donc non interrogées durant les 4 suivantes). Le profil des taux d'entrée dans le champ en fonction des vagues simulé à partir de la modélisation était alors trop éloigné de l'observation.

La probabilité conditionnelle de réponse sachant que le logement est dans le champ en $n+1$ est égale à :

. lors de la première transition :

$$pr = pr_x * (1 + v_{2x})$$

. lors des deuxième à quatrième transitions :

$$pr = pr_x$$

. lors de la dernière transition :

$$pr = pr_x * (1 + vis_x)$$

L'ensemble des paramètres ci-dessus peut être estimé par maximum de vraisemblance. Seuls sont pris en compte pour l'estimation pour une vague donnée les logements dont le statut vis-à-vis du champ à la vague suivante n est pas imputé, ce qui revient à exclure les résidences secondaires en vague intermédiaire⁴⁹. L'estimation peut se faire soit en une étape⁵⁰, soit en deux étapes (avec estimation de la probabilité d'être dans le champ, puis dans une deuxième étape indépendante, estimation de la probabilité d'être répondant menée sur les seuls logements dans le champ). Dans la mesure où les probabilités d'être dans le champ et de répondre sont considérées comme indépendantes, les deux estimations mènent aux mêmes résultats.

Y a-t-il une corrélation entre la probabilité d'être dans le champ en $n+1$ et la probabilité de réponse à cette date et faut-il la prendre en compte ?

Par ailleurs, on peut se poser la question d'une éventuelle corrélation entre la probabilité d'être dans le champ en $n+1$ et la probabilité de réponse. En effet, on peut supposer que les logements qui ont la plus forte probabilité de réponse seront aussi ceux qui auront la plus forte probabilité d'être repérés comme étant dans le champ. Autrement dit, en l'absence de réponse (ou tout du moins de contact avec le ménage), il est plus difficile de voir qu'un logement est entré dans le champ. *A priori*, il est peu probable qu'une éventuelle corrélation induise un biais sur les variables explicatives retenues ici. En effet, celles-ci peuvent être considérées comme exogènes. Néanmoins, plusieurs estimations ont été réalisées en tenant compte de cette corrélation.

Pour introduire une corrélation entre probabilité de présence dans le champ et probabilité de réponse, on a supposé l'existence d'une loi normale bivariée entre variables latentes associées à chacune de ces probabilités. On a donc introduit $xbeta_x = \Phi^{-1}(pc_x)$ – variable latente de présence dans le champ en $n+1$ ⁵¹ – et dans un souci de simplicité, directement redéfini pr_x tel que formulé ci-dessus comme la variable latente déterminant le statut de réponse⁵². ($xbeta_x$, pr_x) sont alors supposés suivre une loi normale bivariée, avec un paramètre de corrélation ρ . Toutes les estimations supposant ρ non nulle ont été séquentielles (probabilité d'être dans le champ, puis dans une deuxième étape récupération du $xbeta_x$ et de pc_x et estimation de la probabilité de répondre). Un premier jeu d'estimation a modélisé la probabilité de réponse à

⁴⁹ Ainsi que les quelques logements classés comme 'autres hors champ' à la vague initiale de la transition et ayant un code résultat à '14' à la vague suivante (et donc n'étant pas enquêtés à cette vague-là).

⁵⁰ La vraisemblance pour une observation est alors : $ll = (entree=1) * \log(prob_n * prob_nn) + (entree=0) * \log(1 - prob_n) + (entree=2) * \log(prob_n * (1 - prob_nn))$;
if champpl=0 then entree=0; else if champpl=1 and totactmpl ne 0 and totactmpl ne . then entree=1; else entree=2;

⁵¹ Cela revient à créer une variable explicative unique $xbeta_x$ pour la probabilité d'être dans le champ, ce qui a paru la manière la plus facile de gérer le fait que pc_x ne pouvait aisément se mettre sous la forme d'une fonction des paramètres du modèle. Les probabilités pc_x obtenues par un modèle probit ayant pour seule variable explicative $xbeta_x$ restent de cette manière exactement celles prédites initialement.

⁵² soit un changement de paramétrage pour la probabilité de réponse, les nouveaux paramètres pr_x se déduisant des anciens par Φ^{-1}

partir de la loi normale bivariée, en écrivant pour la deuxième étape que la probabilité de répondre était égale à $\Phi_2(p_{r_x}, x_{beta_x}, \rho) / p_{c_x}$. et en écrivant la vraisemblance à partir de cette probabilité. Pour être plus robuste au cas où l'hypothèse de binormalité (cruciale) n'était pas vérifiée, un deuxième jeu d'estimations a été effectué en utilisant la méthode du maximum de vraisemblance simulée (décrite par exemple dans Lollivier XXXX). Cela consiste à tirer suite à la première étape pour toutes les observations un certain nombre de résidus u_i dans une loi normale, vérifiant $x_{beta} + u_i > 0$ si le logement est dans le champ à la vague suivante, et $x_{beta} + u_i < 0$ sinon (si le résidu tiré ne convient pas, on retire). La probabilité de répondre prise en compte dans la vraisemblance de deuxième étape est alors la moyenne sur 500 tirages u_i de $\Phi[(p_{r_x} + \rho * u_i) / \sqrt{1 - \rho * \rho}]$. Cette méthode, plus robuste, conduit à des résultats très proches, mais moins précis. Pour chaque méthode, deux hypothèses pour ρ ont été testées : soit ce coefficient de corrélation est supposé identique quelque soit le statut initial du logement, soit il a été supposé pourvoir prendre des valeurs distinctes selon la situation initiale du logement (hors champ, autre non-réponse).

Ces tentatives d'estimations ont montré que :

- Pour les logements initialement en « autre non réponse », la corrélation entre la probabilité d'être dans le champ et de répondre est difficilement identifiable. En effet, la probabilité d'être dans le champ varie très peu pour ces logements (autour de 94 %), et ces variations sont très corrélées avec le fait d'être en vague 5. Empiriquement, il est donc très difficile d'identifier dans la hausse de la probabilité de répondre en vague 5 ce qui relève d'un effet visite (vis_{anr}) de ce qui relève d'un effet d'une corrélation entre probabilité d'être dans le champ en $n+1$ et probabilité de réponse (ρ_{anr}). Cela ne constitue toutefois pas un problème car comme environ 94% de ces logements sont dans le champ à chaque vague, il n'y a guère de risque de sélection.
- Pour les logements initialement hors champ, la corrélation (résiduelle) entre la probabilité d'être dans le champ et de réponse est non significative ($\rho_{lhc} = 0,07$, écart-type = 0,06).

En conclusion, compte tenu des quelques variables explicatives déjà introduites dans le modèle, la corrélation résiduelle entre probabilité d'être dans le champ et probabilité de réponse est a priori faible. De plus, sa non prise en compte est peu susceptible de biaiser les paramètres estimés pour les variables explicatives d'intérêt, ces dernières par leur signification pouvant être considérées comme exogènes⁵³. Les modélisations retenues ne tiendront donc pas compte d'une telle corrélation.

L'estimation de ces paramètres donne les résultats suivants :

- La probabilité qu'un logement initialement en « autre non-réponse » soit dans le champ à la vague suivante est de 94% ; cette probabilité est sous-estimée de 0,5 points en vague téléphonique.
- La probabilité qu'un logement initialement hors champ entre dans le champ en vague suivante est de 14% hors résidence secondaire, et de 0,7% pour les résidences secondaires, ce qui légitime en passant le fait que pour ces dernières il ne soit pas demandé de vérifier le statut vis-à-vis du champ en vague téléphonique.
- La probabilité de réponse en vague téléphonique des logements initialement en « autre non-réponse » restant dans le champ est faible : 22%. En effet, les ménages qui ont été

⁵³ De fait, les paramètres estimés pour les logements initialement hors champ varient peu quand une corrélation est introduite dans le modèle

difficiles à joindre, absent, ou ont refusé à une vague donnée, sont susceptibles d'être dans la même situation à la vague suivante. Cette probabilité de réponse est toutefois supérieure de 12 points en vague 2 – confirmant que parmi les ménages non répondants en vague 1, certains auraient probablement pu être interrogés.

- La probabilité de réponse des logements venant d'entrer dans le champ est supérieure : 60% lors d'une vague téléphonique.
- En visite, les probabilité de réponse des logements du champ sont majorées de 7 points pour les logements initialement non répondant, et de 10 points pour les logements initialement hors champ.

Au total, la probabilité de devenir répondant reste toutefois plus forte pour les logements initialement en « autre non-réponse » (probabilité de 21% en vague téléphonique) que pour les logements initialement hors champ – ces derniers ayant en effet une faible probabilité d'entrer dans le champ.

Tableau 12 – Probabilité qu'un logement initialement sans réponse réponde à la vague suivante (estimation par maximum de vraisemblance)

statut initial du logement	présence dans le champ en n+1			réponse en n+1 (conditionnellement à présence dans le champ)					taux d'entrée total $\pi(a,c) * \pi_t(a,r)$
	probabilité $\pi(a,c)$	taux de sous-estimation en vague téléphonique	sous-estimation en points en vague téléphonique	probabilité $\pi_t(a,r)$ (vague téléphonique)	taux d'augmentation en vague 2	augmentation en points en vague 2	taux d'augmentation en visite	augmentation en points en visite	
autre non réponse	94,4%	-0,5%	-0,5%	21,9%	55,0%	12,0%	32,8%	7,2%	20,7%
hors champ (hors rés. sec.)	14,1%	7,3%	1,0%	59,9%	5,8%	3,5%	17,0%	10,2%	8,4%
résidence secondaire	0,7%		0,1%						0,4%

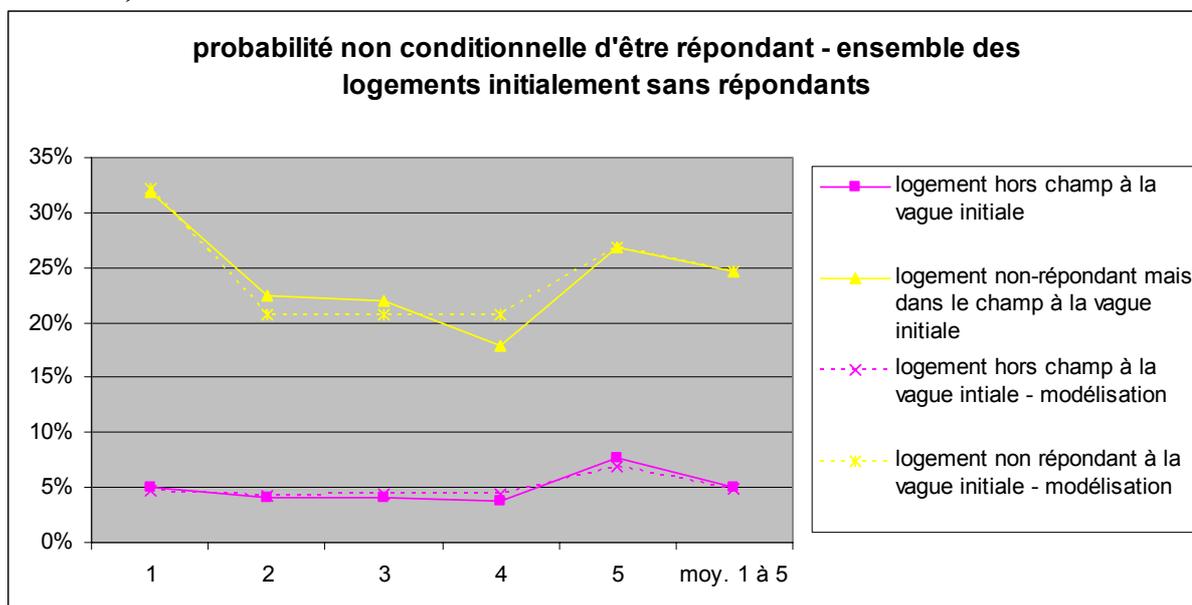
Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, logements présents dans l'échantillon à la vague n et à la vague n+1 hors logements neufs non tirés au sort, soit 168 232 logements ou 2804 en moyenne par vague et sous-échantillon.

Les paramètres estimés ci-dessous permettent de reconstituer de manière satisfaisante les probabilités de réponse en fonction des vagues et des statuts initiaux des logements sans réponse, et ce, sur l'ensemble des logements⁵⁴ et non uniquement ceux ayant servi aux estimations. Seule la baisse de la probabilité de réponse entre les vagues 2 et 4 des logements initialement en « autre non réponse » est mal reproduite, car non-prise en compte dans la modélisation paramétrique.

⁵⁴ L'estimation a en effet été réalisée sur les logements non classés en '14' à la vague suivante (c'est-à-dire sur les logements devant effectivement être interrogés à cette vague), et, pour la probabilité de réponse, sur les seuls logements dans le champ en n+1. Les prédictions à partir du modèle portent elles sur l'ensemble des logements, y compris ceux qui en réalité auraient été en '14' en n+1 (et qui là ne le sont plus forcément : résidence secondaire pouvant entrer dans le champ en vague intermédiaire...), et les probabilités conditionnelles de réponse s'appliquent aux probabilités prédites d'être dans le champ. Les données ainsi simulées sont ainsi définies de manière cohérente avec les données observées, i.e. après redressement des logements non réenquêtés en vague intermédiaire.

Graphique 5 – Probabilité qu'un logement initialement sans réponse réponde à la vague suivante : comparaison des probabilités prédites à partir des paramètres estimés et des probabilités observées (après redressement des logements non ré-enquêtés en vague intermédiaire)



Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, logements présents dans l'échantillon à la vague n et à la vague n+1 hors logements neufs non tirés au sort, soit 168 232 logements ou 2804 en moyenne par vague et sous-échantillon.

Nombre et statut d'activité des enquêtés entrants dans des logements initialement sans répondants

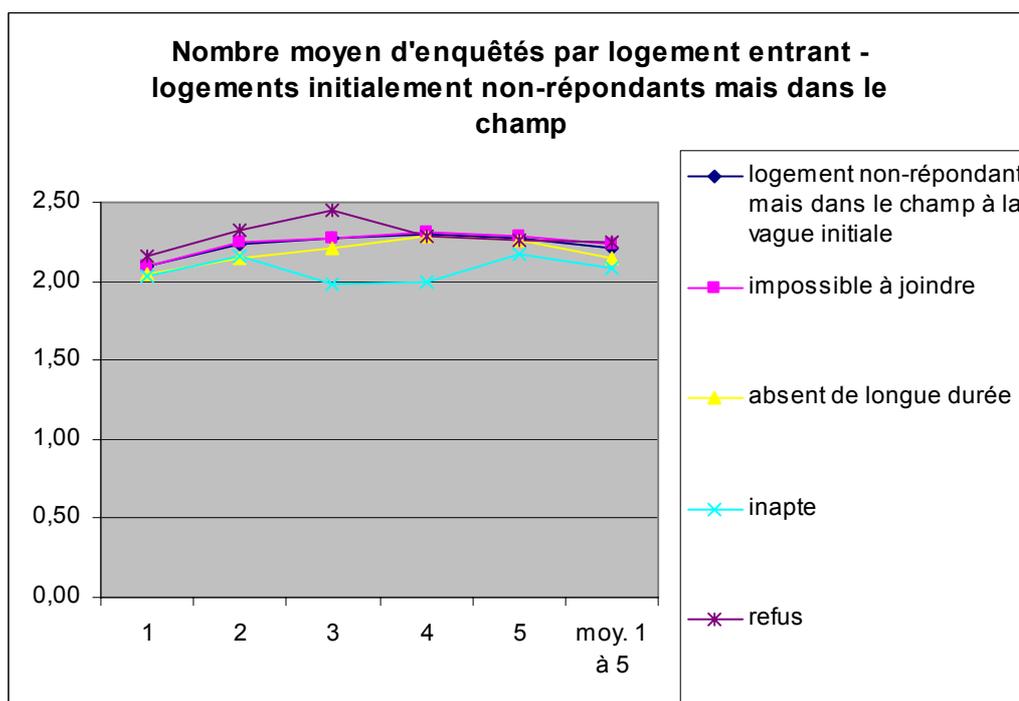
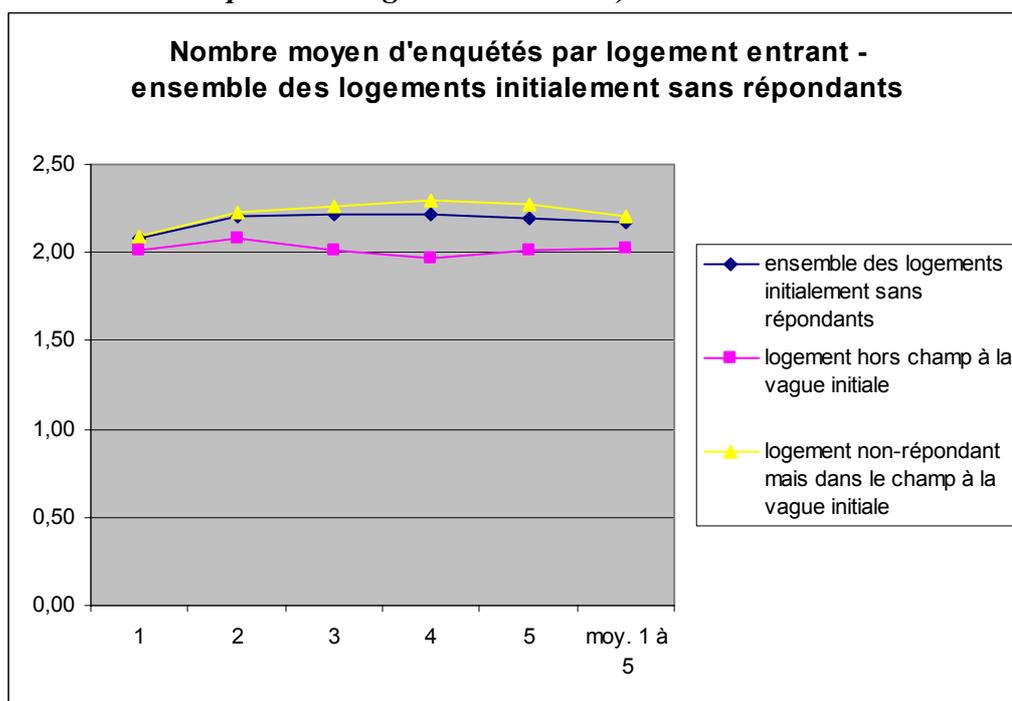
Le nombre de personnes répondantes (i.e. ayant un statut d'activité déclaré) dans un logement le devenant est également susceptible de varier en fonction du motif de non-réponse initial ou de la vague d'enquête. Ce nombre apparaît plus faible mais peu variable pour les logements initialement hors champ (2,02 en moyenne)⁵⁵. Pour les logements initialement en « autre non réponse », le nombre moyen de répondants est de 2,21 et les différences sont plus marquées entre motifs de non-réponse initiale : alors que ce nombre est de 2,23 pour les logements initialement impossibles à joindre, il est légèrement plus faible pour les logements initialement classés en absents de longue durée (2,15) ou en inaptés (2,08), bien que ces écarts soient peu significatifs.

Plus intéressant, le nombre moyen de répondants dans des logements initialement en « autre non-réponse » est plus faible en vague 1 (2,09) qu'aux vagues suivantes (2,23 en vague 2 et entre 2,27 et 2,29 aux vagues 3 à 5).

L'ampleur de ces variations reste toutefois faible devant l'ampleur des variations des probabilités de devenir répondant. C'est pourquoi on se contentera ici d'hypothèses plus simples sur les nombres d'entrants (2,02 pour les logements initialement hors champ, et pour les logements initialement en autre non-réponse 2,09 pour la première transition et 2,27 pour les autres transitions.

⁵⁵ Seuls les logements initialement en résidence occasionnelle auraient un nombre de répondant plus faible que les autres (1,60) mais la différence n'apparaît pas significative compte tenu du faible nombre de cas observés.

Graphique 6 – Nombre moyen d'enquêtés par logement entrant (après redressement des logements non ré-enquêtés en vague intermédiaire)

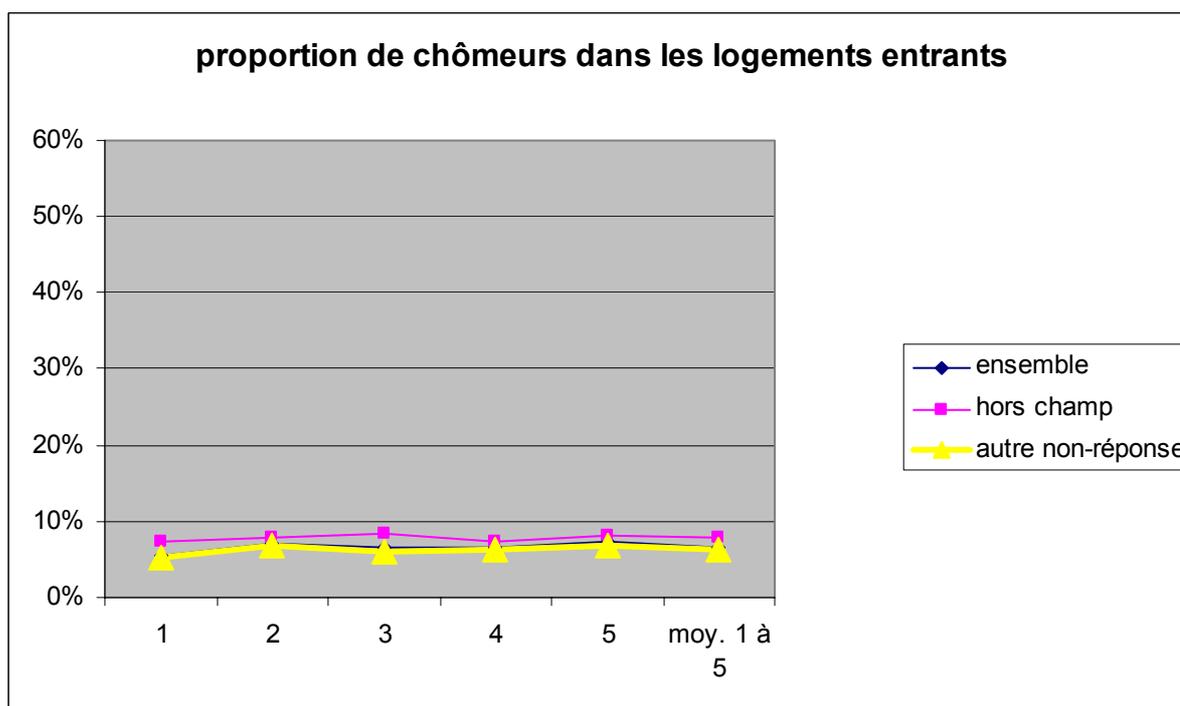
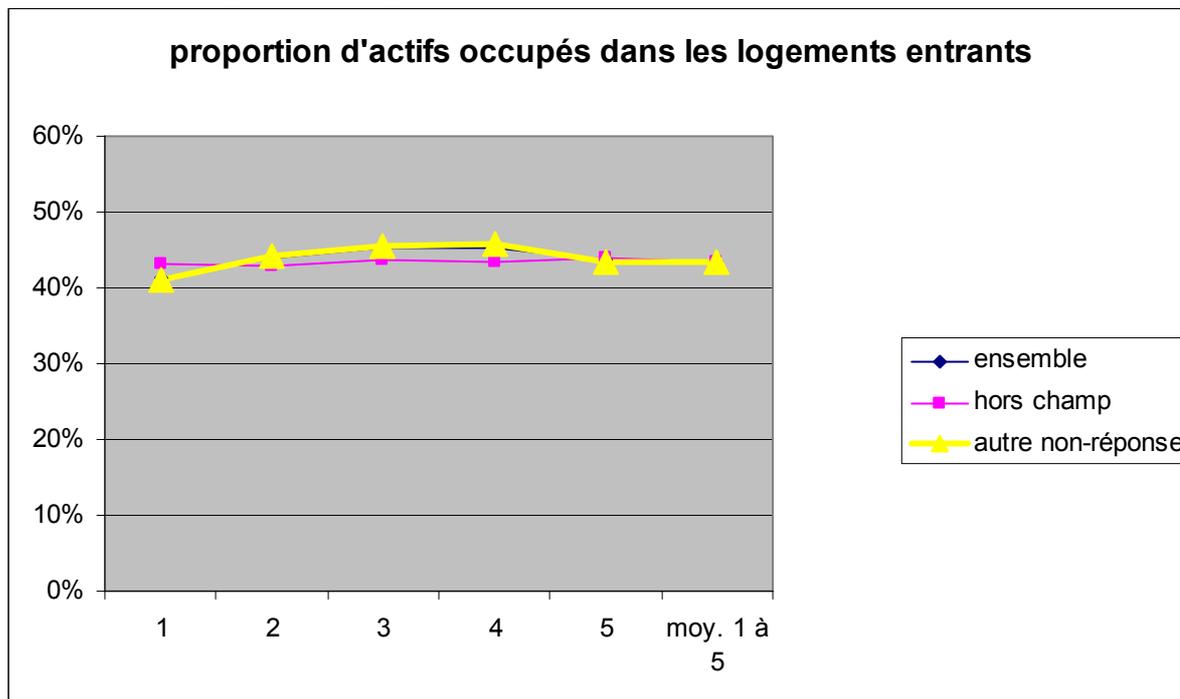


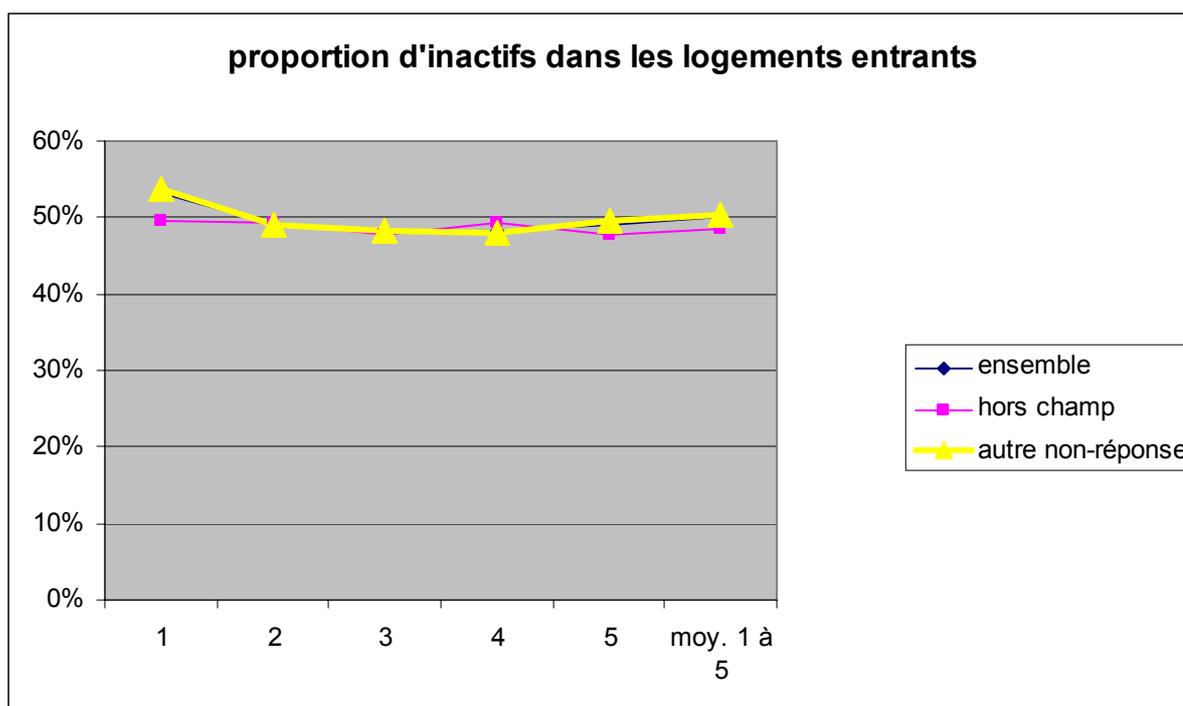
Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, logements présents dans l'échantillon à la vague n et répondants à la vague n+1, soit 22 684 logements ou 378 en moyenne par vague et sous-échantillon.

Pour les logements initialement hors champ, la répartition des entrants par statut d'activité ne varie pas de manière significative. Pour les logements initialement en « autre non-réponse », la principale différence significative est le poids plus fort lors de la première transition des inactifs.

Graphique 7 – Répartition par statut d'activité des entrants (après redressement des logements non ré-enquêtés en vague intermédiaire)





Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées
 Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, logements présents dans l'échantillon à la vague n et répondants à la vague n+1, soit 22 684 logements ou 378 en moyenne par vague et sous-échantillon.

Tableau 13 – Nombre et répartition par statut d'activité des répondants (estimation par maximum de vraisemblance)

	nombre total de répondants	effet V2	proportion actifs occupés	effet V2	proportion de chômeurs	effet V2	proportion d'inactifs	effet V2	taux de chômage	effet V2
statut initial du logement										
autre non réponse	2,27	-0,17	44,6	-3,6	6,6	-1,4	48,8	5,0	12,9	-1,7
hors champ (hors sec)	2,02	.	43,7	.	8,0	.	48,3	.	15,4	.
hors champ (avec rés. sec)	2,02		43,6		8,0		48,3		15,6	

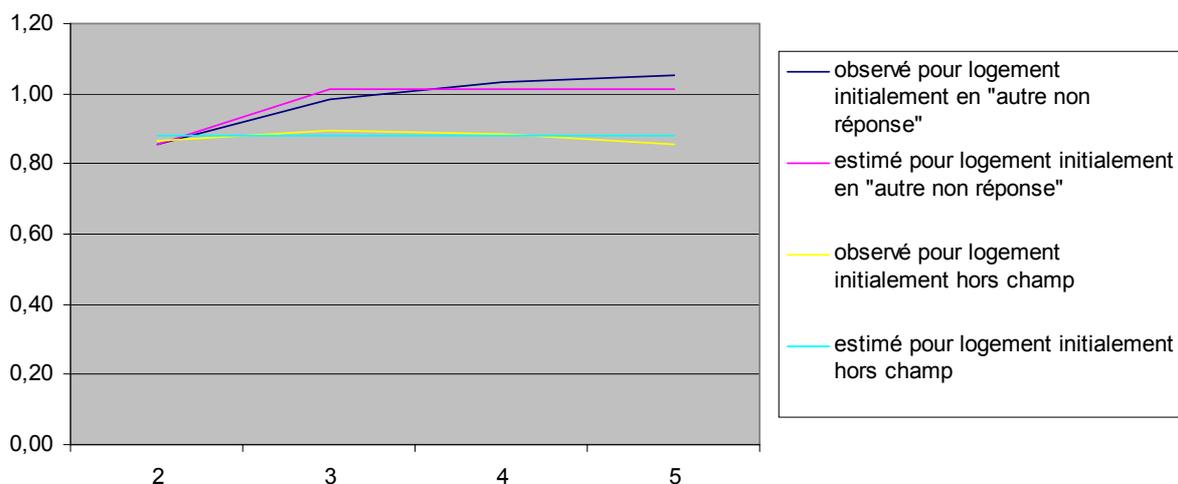
Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées
 Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, logements présents dans l'échantillon à la vague n et répondants à la vague n+1, soit 22 684 logements ou 378 en moyenne par vague et sous-échantillon.

Lecture : Le nombre moyen de répondants dans un logement répondant initialement en autre non réponse est de 2,27, sauf en vague 2 où il est inférieur de 0,17. La proportion d'actifs occupés parmi ces répondants est de 44,6%, sauf en vague 2 où il est inférieur de 3,6 points.

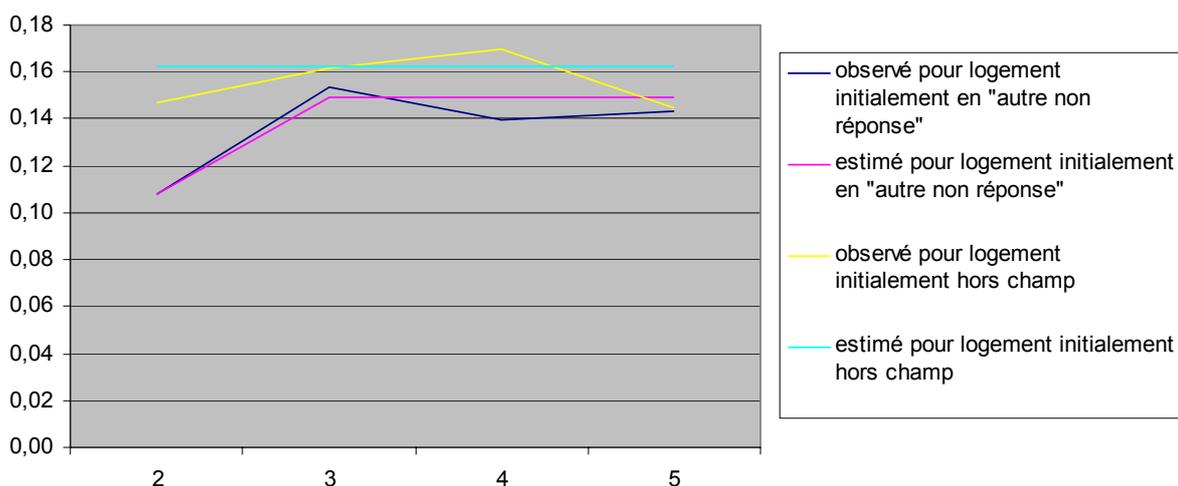
Les hypothèses ci-dessus permettent de recréer l'allure des nombres moyens de personnes par statut d'activité et par vague pour les deux catégories de logement étudiées.

Graphique 8 – Nombres d'entrants par statut d'activité (après redressement des logements non ré-enquêtés en vague intermédiaire)

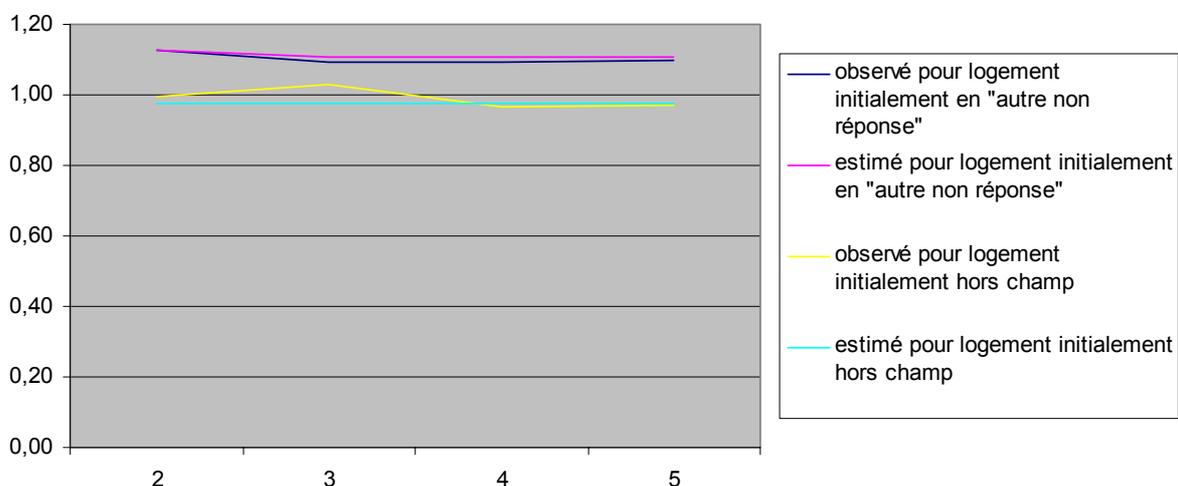
nombre moyen d'actifs occupés par logement entrant



nombre moyen de chômeurs par logement entrant



nombre moyen d'inactifs par logement entrant



Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, logements présents dans l'échantillon à la vague n et répondants à la vague n+1, soit 22 684 logements ou 378 en moyenne par vague et sous-échantillon.

Etude des présents – présents

L'idée sous-jacente est que le protocole de collecte pourrait avoir une incidence sur la possibilité d'observer ou non certaines transitions. Le caractère répété de l'interrogation pourrait notamment jouer : on peut par exemple penser à l'existence de phénomènes d'inertie (baisse du nombre de transition déclarées avec le nombre de vagues), qui pourraient être dus aux enquêteurs ou enquêtés. Les causes de ces éventuels phénomènes pouvant être multiples : moindre insistance des enquêteurs avec les réinterrogations de peur de lasser, apprentissage de la part des ménages des risques de questionnements supplémentaires induits par certaines réponses, normalisation des réponses fournies...

L'étude des transitions entre deux vagues données sera menée sur l'ensemble des personnes ayant un statut d'activité à ces deux vagues. Ce nouveau champ d'étude est un compromis entre le souhait de ne plus avoir à gérer à ce stade des entrées-sorties d'échantillon (déjà analysées), et celui d'éviter tout cylindrage inutile des données (susceptible de biaiser les transitions, les personnes toujours présentes aux bout de six fois pouvant être plus stables). Cette approche néglige la perspective temporelle en tant que telle, mais s'intéresse plutôt aux effets des caractéristiques de la collecte courante sur les résultats (visite ou téléphone, proxy...). L'idée sous-jacente n'est donc plus tant celle d'une éventuelle dérive des déclarations mais plutôt de différences de performances de protocoles vus en coupe.

Le nombre de transitions observés est faible (environ 600 transitions entre deux vagues consécutives par sous échantillon, soit 4,6%).

En moyenne, un peu plus de 13 000 personnes d'un sous-échantillon donné sont présentes à deux vagues d'interrogation consécutives de ce sous-échantillon, se répartissant environ entre 5 200 actifs occupés, 600 chômeurs et 7 500 inactifs. Toujours en moyenne, seules 615 de ces personnes connaissent entre ces deux vagues une transition entre les trois statuts d'activité principaux au sens du B.I.T., soit 4,6% de la population concernée⁵⁶ (tableau 13, dernières colonnes). Le taux de transition observé est plus important entre les vagues 1 et 2 et entre les vagues 5 et 6, c'est-à-dire entre les vagues par visite et par téléphone, et il diminue très légèrement au cours des vagues téléphoniques⁵⁷.

⁵⁶ Ce taux serait certes plus important si l'on limitait l'âge de la population étudiée, mais cela n'aurait pas d'incidence sur les effectifs absolus.

⁵⁷ Ces résultats subsistent si l'on se restreint aux personnes ayant répondu 6 fois, ou encore à celles ayant répondu six fois en personne, mais les effets sont parfois plus faibles.

Tableau 14 – Changements de statut d'activité au sens du B. I. T. entre deux vagues consécutives (en moyenne par sous-échantillon), selon la vague et les statuts de départ ou d'arrivée.

Statut initial ou final Vagues initiale / finale	actifs occupés			chômeurs			inactifs			Total transitions	
	solde	arrivées	départs	solde	arrivées	départs	solde	arrivées	départs	Effectifs	% situation départ
vague 1 à 2	26	243	-217	0	205	-204	-27	199	-225	646	4,89%
vague 2 à 3	13	228	-215	-16	195	-211	3	198	-196	621	4,56%
vague 3 à 4	9	214	-205	-7	194	-200	-2	189	-191	596	4,38%
vague 4 à 5	3	211	-208	0	191	-191	-3	186	-189	588	4,33%
vague 5 à 6	12	227	-216	-11	186	-197	-1	207	-208	621	4,85%
moyenne	13	225	-212	-7	194	-201	-6	196	-202	615	4,60%
Min	3	211	-217	-16	186	-211	-27	186	-225	588	4,33%
Max	26	243	-205	0	205	-191	3	207	-189	646	4,89%
Ecart	23	32	12	16	18	19	29	21	36	58	0,57%

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, personnes pour lesquelles le statut d'activité est disponible aux deux vagues étudiées.

Lecture : en moyenne pour un des 12 sous-échantillons introduits entre le T1 2003 et le T4 2005 : parmi les personnes dont le statut d'activité au sens du B.I.T (actif occupé, chômeur, inactif) est connu à la fois aux vagues 1 et 2, 646 personnes changent de statut d'activité entre ces deux vagues. Elles se répartissent en fonction de leur situation d'arrivée en 243 personnes devenant actives occupées, 205 devenant chômeuses et 199 devenant inactifs.

Et les soldes de ces transitions sont encore plus réduits, avec des variations d'effectifs des catégories B. I. T. entre deux vagues consécutives dépassant rarement la dizaine d'observations par sous-échantillon

615 observations changeant en moyenne de situation d'activité entre deux vagues, cela représente plus de 3000 changements entre deux échantillons trimestriels consécutifs de l'enquête emploi, ce qui est suffisant pour des études s'intéressant aux transitions sur l'ensemble de la population. Cependant, quand on souhaite comparer les transitions en fonction du rang d'interrogation, on ne dispose plus que des 615 observations par sous-échantillon. Les effectifs en jeu sont encore plus faibles si l'on considère que l'étude des transitions n'est qu'une étape dans une analyse, qui *in fine*, vise à comprendre le solde de ces transitions. Or, si on prend l'exemple du nombre de chômeurs, ce solde – résultante d'environ 200 entrées et 200 sorties du chômage entre deux vagues – dépasse rarement la dizaine d'observations par sous-échantillon. Dans une perspective méthodologique, le nombre d'observations peut être augmenté en agrégeant plusieurs sous-échantillons, ce que l'on a fait jusqu'à présent en raisonnant sur les 12 sous-échantillons introduits entre le T1 2003 et le T4 2005.

Même si les effectifs permettent difficilement de conclure, il semblerait que le nombre de nouveaux chômeurs diminue progressivement, en particulier après la vague 1 (diminution du reclassement d'inactifs de première vague en chômeurs) et la vague 5 (baisse du nombre d'actifs occupés devenant chômeurs).

Il semble y avoir un léger rattrapage en vague 6 des sorties du chômage (vers l'inactivité ou l'emploi) non décelées auparavant⁵⁸, et ce, même si l'on se restreint aux personnes ayant répondu 6 fois ou encore aux personnes ayant répondu 6 fois sans proxy.

Il se répartit à peu près équitablement entre les différents changements de statut d'activité possibles mais affecte donc proportionnellement plus fortement la catégorie des chômeurs

Si l'on s'intéresse à une seule des six transitions possibles (par exemple, actif occupé vers inactif), les effectifs pour un sous-échantillon tombent à une centaine d'observations environ (tableau 14). En effet, bien que les trois statuts au sens du B. I. T. aient des effectifs très différentes, les effectifs des personnes en transition varient peu en fonction des statuts de départ et d'arrivée considérés. De ce fait, les probabilités de transitions sont elles très différentes : c'est sur la catégorie des chômeurs que la centaine de transitions pèse plus lourd, les départs (ou arrivées) concernant à chaque fois 15 à 20% de la catégorie.

Tableau 15

	Transition 'actif occupé' -> 'inactif'					Transition 'inactif -> actif occupé'				
	Effectifs	% situation	% situation	chi2 départ	chi2 arrivée	Effectifs	% situation	% situation	chi2 départ	chi2 arrivée
		départ	arrivée				départ	arrivée		
vague 1 à 2	114	2,19%	1,49%	0,6	1,7	124	1,67%	2,31%	39,3	32,5
vague 2 à 3	109	2,04%	1,42%	3,1	2,4	106	1,38%	1,98%	0,8	1,3
vague 3 à 4	106	1,98%	1,37%	7,9	8,7	97	1,26%	1,82%	17,5	16,3
vague 4 à 5	109	2,05%	1,50%	2,6	4,2	97	1,25%	1,92%	18,4	14,6
vague 5 à 6	125	2,48%	1,72%	31,3	29,7	113	1,56%	2,24%	11,2	12,0
moyenne 1 à 5	112	2,14%	1,50%	45,5	46,7	107	1,42%	2,05%	87,1	76,7
min	106	1,98%	1,37%			97	1,25%	1,82%		
max	125	2,48%	1,72%			124	1,67%	2,31%		
écart	19	0,50%	0,35%			27	0,42%	0,49%		
	Transition 'actif occupé -> chômeur'					Transition 'chômeur -> actif occupé'				
	Effectifs	% situation	% situation	chi2 départ	chi2 arrivée	Effectifs	% situation	% situation	chi2 départ	chi2 arrivée
		départ	arrivée				départ	arrivée		
vague 1 à 2	103	1,98%	17,32%	2,2	0,4	120	20,21%	2,23%	0,8	0,7
vague 2 à 3	105	1,96%	18,29%	1,4	0,6	122	20,43%	2,28%	0,2	0,3
vague 3 à 4	100	1,86%	17,90%	0,3	0,2	117	20,41%	2,21%	0,3	0,4
vague 4 à 5	99	1,86%	18,98%	0,4	0,0	114	20,51%	2,27%	0,1	2,1
vague 5 à 6	91	1,81%	17,84%	2,3	0,0	115	22,00%	2,27%	5,2	0,5
moyenne 1 à 5	100	1,90%	18,05%	6,6	1,2	117	20,68%	2,25%	6,6	4,1
min	91	1,81%	17,32%			114	20,21%	2,21%		
max	105	1,98%	18,98%			122	22,00%	2,28%		
écart	14	0,17%	1,66%			8	1,79%	0,07%		

⁵⁸ Sur la sous-population des chômeurs en vague n dont le statut d'activité est connu en vague n+1, un test du chi2 de l'association entre le numéro de vague et la situation d'activité en n+1 rejette l'hypothèse d'indépendance du fait des transitions observées entre les vagues 5 et 6.

	Transition 'chômeur -> inactif'					Transition 'inactif -> chômeur'				
	Effectifs	% situation départ	% situation arrivée	chi2 départ	chi2 arrivée	Effectifs	% situation départ	% situation arrivée	chi2 départ	chi2 arrivée
vague 1 à 2	85	14,32%	1,11%	0,6	1,5	102	1,37%	17,05%	9,6	0,3
vague 2 à 3	89	14,95%	1,16%	0,4	2,7	90	1,18%	15,66%	4,6	7,5
vague 3 à 4	83	14,45%	1,07%	0,3	0,6	94	1,22%	16,95%	0,7	0,3
vague 4 à 5	77	13,91%	1,07%	2,7	9,0	92	1,20%	17,71%	2,6	0,3
vague 5 à 6	83	15,84%	1,14%	5,8	0,9	95	1,32%	18,68%	2,5	11,8
moyenne 1 à 5	83	14,68%	1,11%	9,7	14,7	95	1,25%	17,17%	19,9	20,2
min	77	13,91%	1,07%			90	1,18%	15,66%		
max	89	15,84%	1,16%			102	1,37%	18,68%		
écart	12	1,93%	0,09%			12	0,20%	3,01%		

Simulations

A partir des éléments ci-dessus, on peut faire une simulation pour essayer de reconstituer les profils par vagues observés (et le biais de rotation) et faire des simulations. La simulation ne tient pas compte des logements neufs, et s'affranchit des effets de protocoles « voulus » (pas de non-interrogation en vague intermédiaire des ménages d'inactifs)

La simulation s'appuie sur les hypothèses suivantes :

- Etape 0 : les effectifs de répondants par statut d'activité sont ceux observés en vague 1 (après retraitement des logements non interrogés en vague intermédiaire)
- Etape 1 : Entrées d'enquêtés appartenant à des logements n'ayant jamais répondu avant. Les logements initialement hors champ ou en « autre non-réponse » ont une probabilité d'être dans le champ puis d'être répondant à la vague suivante données par les paramètres estimés du tableau 12 ; s'ils sont répondants, les nombres de personnes entrant et leur répartition par statut d'activité sont obtenus grâce aux paramètres du tableau 13. A partir des effectifs de logements hors champ ou en « autre non-réponse » observés en vague 1, on prédit ainsi le nombre d'entrées d'enquêtés en vague 2, puis aux vagues suivantes⁵⁹.
- Etape 2 : ajout des types entrées impliquant des effectifs réduits et que l'on supposera donc dans un souci de facilité exogènes, à savoir entrées suite à un changement de ménage, et entrées suite à une mobilité individuelle. Les nombres d'entrées par vague et statut d'activité sont ceux observés.
- Etape 3 : pour chaque vague : estimations des sorties des individus répondants à la vague initiale de la transition, en utilisant les paramètres estimés du tableau 6⁶⁰ ; estimations des entrées des personnes déjà enquêtés précédemment (mais non répondantes à la vague initiale de la transition) en utilisant les paramètres du tableau 12⁶¹ ; estimation des changements d'activité des présents-présents en utilisant les taux

⁵⁹ Pour simplifier, chaque logement garde son motif de non-réponse initial tout au long de la procédure.

⁶⁰ Sauf pour le taux de non-réponse des inactifs, sur-estimé compte tenu du fait que l'on n'a pas considéré que parmi les inactifs deux catégories existaient avec des taux de réponse différents : les inactifs de plus de 65 ans et les autres. Le taux de non-réponse retenu est de 3,5% (au lieu de 4,3%)

⁶¹ On fait l'hypothèse que les paramètres sont valides aussi bien pour les entrées de logements que d'individus.

de transition du tableau 15 ; calcul du nombre final d'individu par catégorie (actif occupé, chômeur, inactifs, hors champ, non-répondant) et itération.

Les données simulées sont proches des données observées, tant pour ce qui est des effectifs que du profil des taux de chômage.

Tableau 16.

Données observées (après traitement des '14' et hors logements neufs)

	1	2	3	4	5	6	moy
effectifs initiaux							
actif occupé	5591	5763	5753	5733	5648	5573	5677
chômeur	666	666	645	630	606	609	637
inactif	7788	8033	8025	8016	7935	7895	7949
total	14044	14462	14423	14380	14189	14076	14262

Données simulées (après traitement des '14' et hors logements neufs)

	1	2	3	4	5	6	moy
effectifs initiaux							
actif occupé	5591	5747	5723	5684	5602	5517	5644
chômeur	666	665	638	621	604	597	632
inactif	7788	8014	7994	7961	7928	7860	7924
total	14044	14427	14355	14266	14134	13974	14200
Taux de chômage							
observé	10,6%	10,4%	10,1%	9,9%	9,7%	9,8%	
simulé	10,6%	10,4%	10,0%	9,8%	9,7%	9,8%	

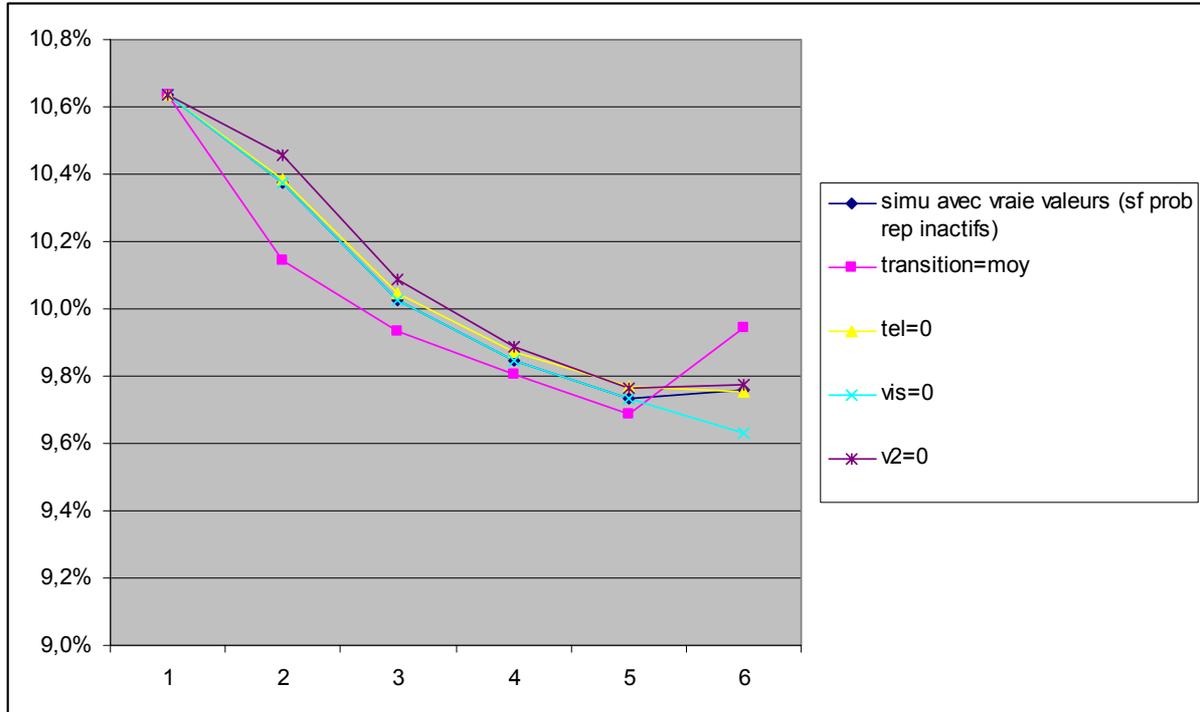
Tableau 17 – taux de chômage dans différents scénarii

	1	2	3	4	5	6	min	max	biais	
simu avec vraie valeurs (sf prob rep inactifs)		10,6%	10,4%	10,0%	9,8%	9,7%	9,8%	9,7%	10,6%	0,9%
transition=moy		10,6%	10,1%	9,9%	9,8%	9,7%	9,9%	9,7%	10,6%	1,0%
tel=0		10,6%	10,4%	10,0%	9,9%	9,8%	9,8%	9,8%	10,6%	0,9%
vis=0		10,6%	10,4%	10,0%	9,8%	9,7%	9,6%	9,6%	10,6%	1,0%
v2=0		10,6%	10,5%	10,1%	9,9%	9,8%	9,8%	9,8%	10,6%	0,9%
effectifs v2		10,4%	10,2%	9,9%	9,8%	9,7%	9,7%	9,7%	10,4%	0,7%
effectifs v3		10,1%	10,1%	9,8%	9,7%	9,7%	9,7%	9,7%	10,1%	0,4%
effectifs v4		9,9%	10,0%	9,8%	9,7%	9,7%	9,7%	9,7%	10,0%	0,3%
effectifs v5		9,7%	9,9%	9,7%	9,7%	9,7%	9,7%	9,7%	9,9%	0,2%
effectifs v6		9,9%	10,0%	9,8%	9,7%	9,7%	9,8%	9,7%	10,0%	0,3%
effectifs moyens		10,1%	10,1%	9,9%	9,8%	9,7%	9,7%	9,7%	10,1%	0,4%
effectifs v1 et autres paramètres de perturbation à 0		10,6%	10,2%	10,0%	9,9%	9,7%	9,8%	9,7%	10,6%	0,9%
effectifs v2 et autres paramètres de perturbation à 0		10,4%	10,1%	9,9%	9,8%	9,7%	9,8%	9,7%	10,4%	0,6%
effectifs v5 et autres paramètres de perturbations à 0		9,7%	9,7%	9,7%	9,7%	9,7%	9,8%	9,7%	9,8%	0,1%

Un premier jeu de simulation vise à tester un à un les effets de différents paramètres de « perturbations »: transitions des présents-présents différents des probabilités de transition moyenne par statut d'activité, mauvaise appréciation de la présence dans le champ pour les

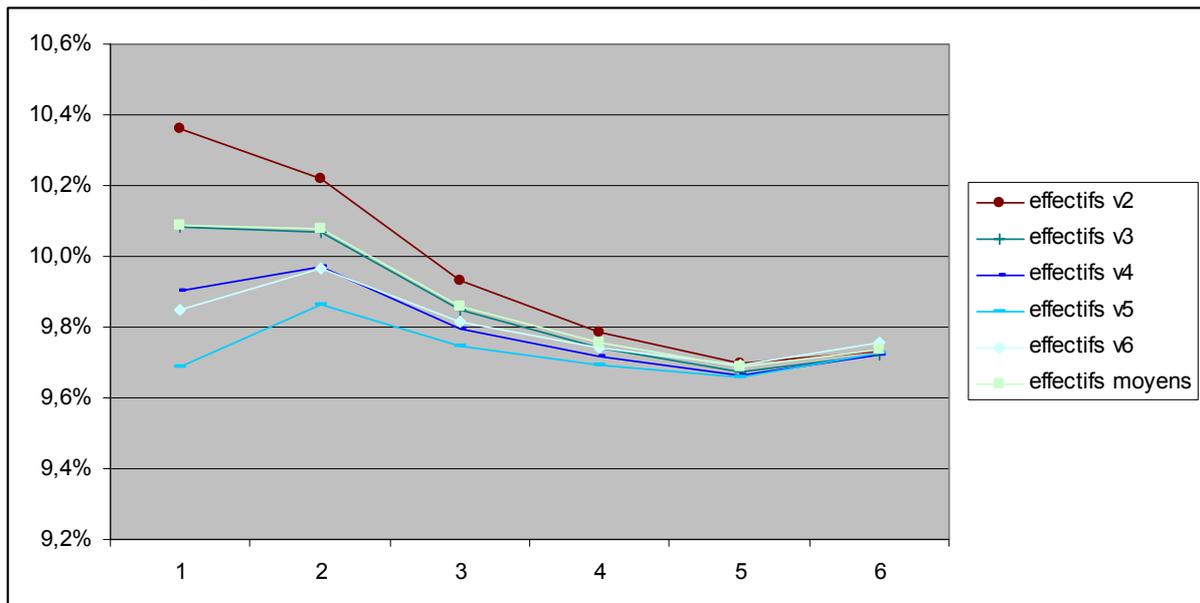
enquêtes intermédiaires (téléphone), probabilité de devenir répondant conditionnellement à la présence dans le champ supérieure en visite (et probabilité d'abandon aussi), spécificité de la vague 2 (pour un grand nombre de paramètres). Jouer sur ces paramètres a peu d'effet sur le profil des taux de chômage et le biais de rotation.

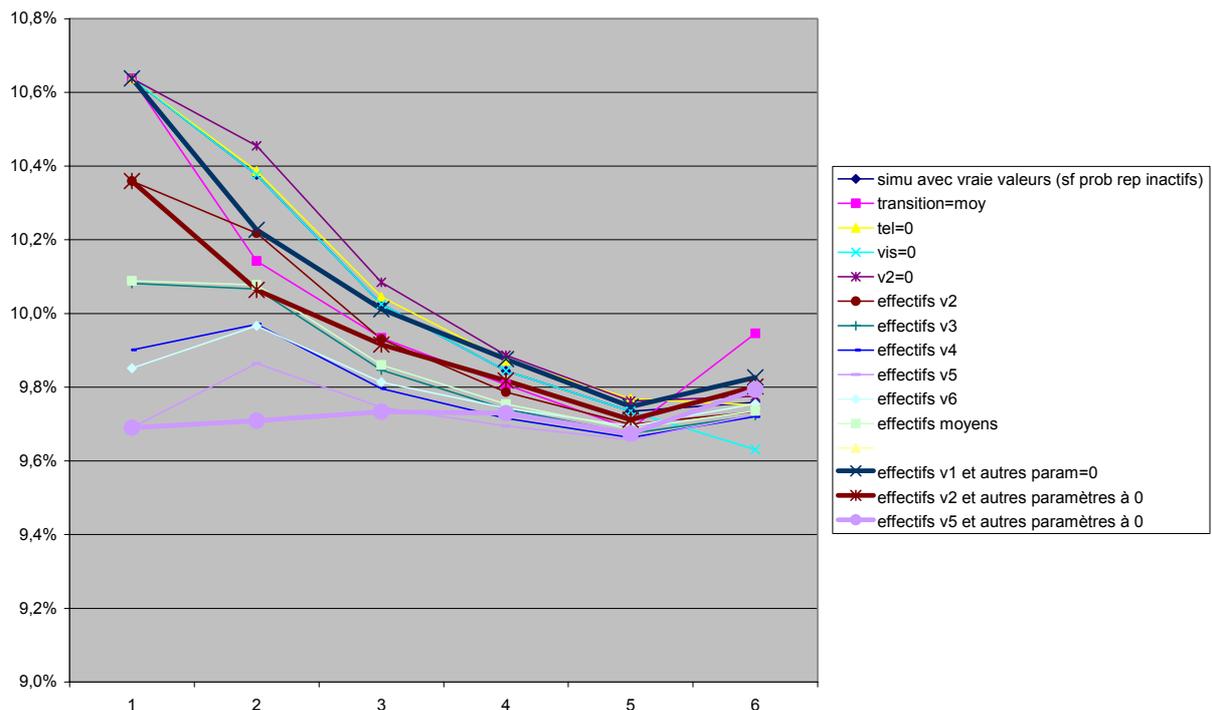
Taux de chômage



En revanche, modifier les effectifs de départ de la simulation a un effet plus fort. Partir des effectifs de vague 5 réduit ainsi à 0,2 point le biais de rotation. Si en plus on annule les paramètres de perturbations, le biais de rotation se réduit à 0,1 point.

Taux de chômage





Ainsi, finalement, les problèmes de protocole examinés ici (traitement des logements neufs et des logements d'inactifs de 65 ans ou plus, mauvaise estimation de la probabilité d'être dans le champ en vague intermédiaire, effet visite, transitions entre actif occupé, chômeur, inactif) semblent expliquer assez peu le biais de rotation. C'est bien l'association des effectifs initiaux par statut d'activité, et des probabilités de transition moyennes entre différents états (actif occupé, chômeur, inactif, non réponse) qui semble être à l'origine de ce dernier. On ne peut toutefois sans doute pas déduire du fait que la répartition des effectifs par statut d'activité en vague 5 se traduise par un biais de rotation le plus faible que c'est la « vraie » répartition. C'est probablement seulement celle qui est la plus compatible avec les probabilités de transitions moyennes entre états (qui sont probablement aussi biaisées en raison de la non-réponse plus forte des chômeurs).

Annexe

Tableau 2 (complet) – Changements de situation individuels entre deux vagues successives

actif occupé		effectifs selon la transition						répartition sur les présents à au moins une des deux dates selon la transition					
		1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy.	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy.
toujours présent	encore actif occupé	4997	5147	5131	5107	4813	5039	79%	81%	81%	82%	75%	80%
	ne va plus être actif occupé	217	215	205	208	216	212	3%	3%	3%	3%	3%	3%
	va devenir actif occupé	243	228	214	211	227	225	4%	4%	3%	3%	4%	4%
actif occupé sortant échantillon		377	394	402	398	590	432	6%	6%	6%	6%	9%	7%
actif occupé entrant dans l'échantillon		516	364	366	302	595	428	8%	6%	6%	5%	9%	7%
total vague initiale		5591	5756	5738	5712	5619	5683	100%	100%	100%	100%	100%	100%
total vague finale		5756	5738	5712	5619	5635	5692						
chômeur		effectifs selon la transition						répartition sur les présents à au moins une des deux dates selon la transition					
		1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy.	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy.
toujours présent	encore chômeur	387	385	374	365	324	367	41%	42%	42%	42%	36%	41%
	ne va plus être chômeur	204	211	200	191	197	201	22%	23%	22%	22%	22%	22%
	va devenir chômeur	205	195	194	191	186	194	22%	21%	22%	22%	21%	22%
chômeur sortant échantillon		74	70	68	71	80	73	8%	8%	8%	8%	9%	8%
chômeur entrant dans l'échantillon		73	63	58	45	104	69	8%	7%	7%	5%	12%	8%
total vague initiale		666	665	643	627	601	640	100%	100%	100%	100%	100%	100%
total vague finale		665	643	627	601	615	630						
inactif		effectifs selon la transition						répartition sur les présents à au moins une des deux dates selon la transition					
		1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy.	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy.
toujours présent	encore inactif	7177	7458	7505	7536	7041	7343	83%	86%	86%	87%	78%	84%
	ne va plus être inactif	225	196	191	189	208	202	3%	2%	2%	2%	2%	2%
	va devenir inactif	199	198	189	186	207	196	2%	2%	2%	2%	2%	2%
inactif sortant échantillon		385	413	398	394	822	482	4%	5%	5%	5%	9%	6%
inactif entrant dans l'échantillon		691	437	425	350	712	523	8%	5%	5%	4%	8%	6%
total vague initiale		7788	8067	8093	8119	8072	8028	100%	100%	100%	100%	100%	100%
total vague finale		8067	8093	8119	8072	7961	8062						
total		effectifs selon la transition						répartition sur les présents à au moins une des deux dates selon la transition					
		1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy.	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy.
tjrs présent – même statut		12561	12989	13010	13007	12179	12749	79%	81%	82%	83%	75%	80%
tjrs présent – autre statut		646	621	596	588	621	615	4%	4%	4%	4%	4%	4%
tjrs présent – total		13208	13610	13607	13595	12800	13364	83%	85%	85%	86%	78%	84%
va sortir de l'échantillon		837	877	867	862	1493	987	5%	5%	5%	5%	9%	6%
va entrer dans l'échantillon		1280	864	850	697	1411	1020	8%	5%	5%	4%	9%	6%
total vague initiale		14044	14488	14474	14457	14292	14351	100%	100%	100%	100%	100%	100%
total vague finale		14488	14474	14457	14292	14211	14384						

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, personnes présentes à au moins une des deux vagues étudiées.

Annexe :

Tableau – Les entrées en cours d'enquête en fonction du motif du statut d'activité à la vague courante et du rang d'interrogation, après redressement des effets du protocole lié aux logements d'inactifs imputés et aux logements neufs

vague initiale de la transition	effectifs						taux d'entrée						
	1	2	3	4	5	moy. 1 à 5	1	2	3	4	5	moy. 1 à 5	moy. 1 à 4
total													
actif occupé	535	384	386	322	516	428	9,6%	6,7%	6,7%	5,6%	9,2%	7,5%	7,1%
chômeur	75	65	61	48	95	69	11,3%	9,8%	9,5%	7,6%	15,7%	10,7%	9,6%
inactif	714	460	448	373	620	523	9,2%	5,7%	5,5%	4,6%	7,7%	6,5%	6,2%
total	1325	909	895	742	1230	1020	9,4%	6,3%	6,2%	5,1%	8,6%	7,1%	6,7%
entrées "normales" (correspondant à des entrées dans le champ)													
actif occupé	124	117	126	114	219	140	2,2%	2,0%	2,2%	2,0%	3,9%	2,5%	2,1%
logement en '14' à la vague initiale	7	7	7	7	13	9							
logement pas en '14' à la vague initiale (ni finale)	117	110	119	107	206	132							
chômeur	23	24	26	19	47	28	3,5%	3,6%	4,0%	3,1%	7,9%	4,4%	3,5%
logement en '14' à la vague initiale	1	1	1	1	3	1							
logement pas en '14' à la vague initiale (ni finale)	22	23	24	18	45	26							
inactif	174	164	173	155	282	189	2,2%	2,0%	2,1%	1,9%	3,5%	2,4%	2,1%
logement en '14' à la vague initiale	10	10	10	10	18	12							
logement pas en '14' à la vague initiale (ni finale)	165	154	163	145	264	178							
total	322	305	325	288	549	358	2,3%	2,1%	2,2%	2,0%	3,8%	2,5%	2,2%
logement en '14' à la vague initiale	18	19	18	18	34	21							
logement pas en '14' à la vague initiale (ni finale)	304	287	306	270	514	336							
changement ménage													
actif occupé	33	36	42	35	67	43	0,6%	0,6%	0,7%	0,6%	1,2%	0,8%	0,6%
logement en '14' à la vague initiale	4	4	4	4	7	5							
logement pas en '14' à la vague initiale (ni finale)	30	32	39	31	60	38							
chômeur	7	7	8	5	15	8	1,0%	1,1%	1,2%	0,7%	2,5%	1,3%	1,0%
logement en '14' à la vague initiale	1	1	1	1	1	1							
logement pas en '14' à la vague initiale (ni finale)	6	6	7	4	14	8							

inactif	36	37	43	35	74	45	0,5%	0,5%	0,5%	0,4%	0,9%	0,6%	0,5%
logement en '14' à la vague initiale	5	5	5	5	10	6							
logement pas en '14' à la vague initiale (ni finale)	31	32	37	30	64	39							
total	76	80	93	74	156	96	0,5%	0,6%	0,6%	0,5%	1,1%	0,7%	0,6%
logement en '14' à la vague initiale	10	10	10	10	18	11							
logement pas en '14' à la vague initiale (ni finale)	67	71	83	65	137	85							
hors champ (y compris anciennes résidences secondaires)													
actif occupé	65	57	57	52	115	69	1,2%	1,0%	1,0%	0,9%	2,0%	1,2%	1,0%
logement en '14' à la vague initiale	2	3	2	2	5	3							
logement pas en '14' à la vague initiale (ni finale)	63	54	55	50	110	66							
chômeur	11	10	11	9	21	13	1,7%	1,6%	1,7%	1,4%	3,6%	2,0%	1,6%
logement en '14' à la vague initiale	0	0	0	0	1	0							
logement pas en '14' à la vague initiale (ni finale)	11	10	11	8	21	12							
inactif	76	65	63	59	124	77	1,0%	0,8%	0,8%	0,7%	1,5%	1,0%	0,8%
logement en '14' à la vague initiale	3	3	3	3	6	4							
logement pas en '14' à la vague initiale (ni finale)	72	62	59	56	118	73							
total	152	133	131	120	261	159	1,1%	0,9%	0,9%	0,8%	1,8%	1,1%	0,9%
logement en '14' à la vague initiale	6	6	6	6	12	7							
logement pas en '14' à la vague initiale (ni finale)	146	126	124	114	249	152							
mobilité individuelle													
actif occupé	13	12	14	14	25	16	0,2%	0,2%	0,3%	0,2%	0,4%	0,3%	0,2%
logement en '14' à la vague initiale	0,9	1,0	0,9	0,9	1,6	1,07							
logement pas en '14' à la vague initiale (ni finale)	12	11	14	13	23	15							
chômeur	4	5	6	5	10	6	0,6%	0,8%	0,9%	0,8%	1,6%	0,9%	0,8%
logement en '14' à la vague initiale	0	0	0	0	0	0							
logement pas en '14' à la vague initiale (ni finale)	4	5	5	5	9	6							
inactif	50	48	54	47	71	54	0,6%	0,6%	0,7%	0,6%	0,9%	0,7%	0,6%

logement en '14' à la vague initiale	1	1	1	1	2	1								
logement pas en '14' à la vague initiale (ni finale)	48	47	53	46	69	53								
total	67	65	74	67	105	76	0,5%	0,5%	0,5%	0,5%	0,7%	0,5%	0,5%	
logement en '14' à la vague initiale	2	3	2	2	4	3								
logement pas en '14' à la vague initiale (ni finale)	64	63	72	64	102	73								
nouveau logement														
actif occupé	13	13	13	13	13	13	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%
chômeur	1	1	1	1	1	1	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%
inactif	13	13	13	13	13	13	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%
total	27	27	27	27	27	27	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%
entrées suite à non-interrogation de ménages supposés dans le champ														
actif occupé	411	266	260	208	296	288	7,4%	4,6%	4,5%	3,6%	5,3%	5,1%	5,0%	
chômeur	52	42	35	28	47	41	7,9%	6,3%	5,5%	4,5%	7,9%	6,4%	6,1%	
inactif	540	296	276	218	339	334	6,9%	3,7%	3,4%	2,7%	4,2%	4,2%	4,1%	
total	1003	604	571	454	682	663	7,1%	4,2%	3,9%	3,1%	4,8%	4,6%	4,6%	
ald														
actif occupé	82	41	40	37	42	49	1,5%	0,7%	0,7%	0,7%	0,8%	0,9%	0,9%	
chômeur	10	7	5	6	6	7	1,5%	1,0%	0,8%	0,9%	1,0%	1,0%	1,0%	
inactif	153	57	56	48	56	74	2,0%	0,7%	0,7%	0,6%	0,7%	0,9%	1,0%	
total	245	105	101	90	104	129	1,7%	0,7%	0,7%	0,6%	0,7%	0,9%	0,9%	
iaj														
actif occupé	260	195	197	151	218	204	4,6%	3,4%	3,4%	2,6%	3,9%	3,6%	3,5%	
chômeur	33	30	27	20	36	29	4,9%	4,4%	4,2%	3,1%	5,9%	4,5%	4,2%	
inactif	257	193	183	148	223	201	3,3%	2,4%	2,3%	1,8%	2,8%	2,5%	2,4%	
total	550	417	407	319	476	434	3,9%	2,9%	2,8%	2,2%	3,3%	3,0%	2,9%	
inapte														
actif occupé	10	7	4	4	10	7	0,2%	0,1%	0,1%	0,1%	0,2%	0,1%	0,1%	
chômeur	3	1	0	1	2	1	0,4%	0,1%	0,1%	0,1%	0,4%	0,2%	0,2%	
inactif	22	11	7	7	19	13	0,3%	0,1%	0,1%	0,1%	0,2%	0,2%	0,1%	
total	36	18	11	11	31	21	0,3%	0,1%	0,1%	0,1%	0,2%	0,1%	0,1%	
refus / statut d'activité inconnu														
actif occupé	59	24	19	15	26	29	1,1%	0,4%	0,3%	0,3%	0,5%	0,5%	0,5%	
chômeur	7	4	3	3	4	4	1,0%	0,7%	0,5%	0,4%	0,6%	0,6%	0,6%	
inactif	107	35	29	16	41	46	1,4%	0,4%	0,4%	0,2%	0,5%	0,6%	0,6%	
total	173	63	51	34	70	78	1,2%	0,4%	0,4%	0,2%	0,5%	0,5%	0,6%	

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005. Pour les effectifs : personnes dont le statut d'activité est connu à la deuxième des deux vagues étudiées et non à la précédente (dite « vague initiale »).

Tableau – Les sorties en cours d'enquête en fonction du motif du statut d'activité à la vague précédente et du rang d'interrogation, après redressement des effets du protocole lié aux logements d'inactifs imputés

	effectifs	probabilité de sortie
--	-----------	-----------------------

vague initiale	1	2	3	4	5	moy. 1 à 5	1	2	3	4	5	moy. 1 à 5	moy. 1 à 4
total													
actif occupé	377	394	402	398	590	432	6,7%	6,9%	7,0%	7,0%	10,5%	7,6%	6,9%
chômeur	74	70	68	71	80	73	11,2%	10,5%	10,6%	11,3%	13,3%	11,3%	10,9%
inactif	429	457	441	438	647	482	5,5%	5,7%	5,5%	5,4%	8,0%	6,0%	5,5%
<i>inactif en '14' à la vague initiale</i>	44	44	44	44	86	52		2,4%	2,2%	2,1%	4,0%	3,2%	
<i>inactif en '14' à aucune des vagues</i>	385	413	398	394	561	430	6,5%	6,8%	6,6%	6,6%	9,4%	7,2%	6,6%
total	881	921	911	906	1317	987	6,3%	6,4%	6,3%	6,3%	9,2%	6,9%	6,3%
sorties (correspondant à des sorties du champ)													
actif occupé	135	148	168	149	237	167	2,4%	2,6%	2,9%	2,6%	4,2%	2,9%	2,6%
chômeur	31	32	29	29	37	32	4,7%	4,8%	4,6%	4,7%	6,2%	5,0%	4,7%
inactif	156	172	176	161	276	188	2,0%	2,1%	2,2%	2,0%	3,4%	2,3%	2,07%
<i>inactif en '14' à la vague initiale</i>	22	22	22	22	49	27		1,2%	1,1%	1,0%	2,3%	1,7%	
<i>inactif en '14' à aucune des vagues</i>	134	150	155	139	227	161	2,3%	2,5%	2,6%	2,3%	3,8%	2,7%	2,4%
total	301	330	352	318	637	387	2,1%	2,3%	2,4%	2,2%	4,5%	2,7%	2,3%
changement ménage													
actif occupé	34	36	43	34	67	43	0,6%	0,6%	0,8%	0,6%	1,2%	0,8%	0,6%
chômeur	4	6	5	5	7	5	0,6%	0,9%	0,8%	0,7%	1,2%	0,9%	0,8%
inactif	34	41	42	36	76	46	0,4%	0,5%	0,5%	0,4%	0,9%	0,6%	0,5%
<i>inactif en '14' à la vague initiale</i>	6	6	6	6	11	7		0,3%	0,3%	0,3%	0,5%	0,4%	0,4%
<i>inactif en '14' à aucune des vagues</i>	28	35	36	31	64	39	0,5%	0,6%	0,6%	0,5%	1,1%	0,7%	0,5%
total	66	77	85	70	173	94	0,5%	0,5%	0,6%	0,5%	1,2%	0,7%	0,5%
hors champ													
actif occupé	66	76	81	72	111	81	1,2%	1,3%	1,4%	1,3%	2,0%	1,4%	1,3%
chômeur	13	13	12	12	15	13	1,9%	1,9%	1,9%	1,9%	2,5%	2,0%	1,9%
inactif	81	88	90	84	130	95	1,0%	1,1%	1,1%	1,0%	1,6%	1,2%	1,1%
<i>inactif en '14' à la vague initiale</i>	10	10	10	10	23	13		0,6%	0,5%	0,5%	1,1%	0,8%	0,7%
<i>inactif en '14' à aucune des vagues</i>	71	78	80	73	107	82	1,2%	1,3%	1,3%	1,2%	1,8%	1,4%	1,3%
total	150	167	173	158	297	189	1,1%	1,2%	1,2%	1,1%	2,1%	1,3%	1,1%
mobilité individuelle													
actif occupé	35	36	44	43	60	44	0,6%	0,6%	0,8%	0,7%	1,1%	0,8%	0,7%
chômeur	14	13	12	13	15	13	2,1%	2,0%	1,9%	2,0%	2,5%	2,1%	2,0%
inactif	41	42	44	41	70	47	0,5%	0,5%	0,5%	0,5%	0,9%	0,6%	0,5%
<i>inactif en '14' à la vague initiale</i>	6	6	6	6	14	7		0,3%	0,3%	0,3%	0,7%	0,4%	0,4%
<i>inactif en '14' à aucune des vagues</i>	35	37	38	35	56	40	0,6%	0,6%	0,6%	0,6%	0,9%	0,7%	0,6%
total	85	86	94	91	167	104	0,6%	0,6%	0,7%	0,6%	1,2%	0,7%	0,6%
sorties liées à une non-interrogation de personnes supposées dans le champ													
actif occupé	242	247	234	249	353	265	4,3%	4,3%	4,1%	4,4%	6,3%	4,7%	4,3%
chômeur	43	38	39	41	43	41	6,4%	5,7%	6,0%	6,6%	7,2%	6,4%	6,2%
inactif	273	285	265	277	371	294	3,5%	3,5%	3,3%	3,4%	4,6%	3,7%	3,4%

inactif en '14' à la vague initiale	22	22	22	22	37	25		1,2%	1,1%	1,1%	1,7%	1,6%	
inactif en '14' à aucune des vagues	251	263	243	254	334	269	4,2%	4,3%	4,0%	4,2%	5,6%	4,5%	4,2%
total	536	547	515	544	856	600	3,8%	3,8%	3,6%	3,8%	6,0%	4,2%	3,7%
ald													
actif occupé	38	40	41	45	79	49	0,7%	0,7%	0,7%	0,8%	1,4%	0,9%	0,7%
chômeur	7	4	5	7	9	6	1,1%	0,6%	0,8%	1,1%	1,6%	1,0%	0,9%
inactif	52	61	62	62	101	68	0,7%	0,8%	0,8%	0,8%	1,3%	0,8%	0,7%
inactif en '14' à la vague initiale	9	9	9	9	15	10		0,5%	0,4%	0,4%	0,7%	0,6%	0,6%
inactif en '14' à aucune des vagues	44	53	53	54	87	58	0,7%	0,9%	0,9%	0,9%	1,5%	1,0%	0,8%
total	89	97	100	105	224	123	0,6%	0,7%	0,7%	0,7%	1,6%	0,9%	0,7%
iaj													
actif occupé	176	179	170	181	209	183	3,2%	3,1%	3,0%	3,2%	3,7%	3,2%	3,1%
chômeur	31	30	29	31	24	29	4,6%	4,5%	4,5%	4,9%	4,0%	4,5%	4,6%
inactif	179	185	168	179	182	179	2,3%	2,3%	2,1%	2,2%	2,3%	2,2%	2,2%
inactif en '14' à la vague initiale	7	7	7	7	8	7		0,4%	0,4%	0,4%	0,4%	0,5%	
inactif en '14' à aucune des vagues	172	177	161	172	174	171	2,9%	2,9%	2,7%	2,9%	2,9%	2,9%	2,8%
total	379	386	359	383	444	390	2,7%	2,7%	2,5%	2,7%	3,1%	2,7%	2,6%
inapte													
actif occupé	5	5	4	5	11	6	0,1%	0,1%	0,1%	0,1%	0,2%	0,1%	0,1%
chômeur	1	0	0	2	2	1	0,1%	0,1%	0,1%	0,2%	0,4%	0,2%	0,1%
inactif	10	6	7	9	20	10	0,1%	0,1%	0,1%	0,1%	0,2%	0,1%	0,1%
inactif en '14' à la vague initiale	2	2	2	2	5	3		0,1%	0,1%	0,1%	0,3%	0,2%	
inactif en '14' à aucune des vagues	8	4	5	7	14	8	0,1%	0,1%	0,1%	0,1%	0,2%	0,1%	0,1%
total	13	9	10	13	42	18	0,1%	0,1%	0,1%	0,1%	0,3%	0,1%	0,1%
refus / statut d'activité inconnu													
actif occupé	23	23	19	18	55	27	0,4%	0,4%	0,3%	0,3%	1,0%	0,5%	0,4%
chômeur	5	3	4	2	8	4	0,7%	0,5%	0,6%	0,4%	1,3%	0,7%	0,6%
inactif	32	33	28	26	68	37	0,4%	0,4%	0,3%	0,3%	0,8%	0,5%	0,4%
inactif en '14' à la vague initiale	4	4	4	4	9	5		0,2%	0,2%	0,2%	0,4%	0,3%	
inactif en '14' à aucune des vagues	28	29	24	22	59	32	0,5%	0,5%	0,4%	0,4%	1,0%	0,5%	0,4%
total	56	56	47	42	146	69	0,4%	0,4%	0,3%	0,3%	1,0%	0,5%	0,3%

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 200, personnes dont le statut d'activité est connu à la première des deux vagues étudiées (« vague initiale ») et (pour les sorties) non à la suivante.

Lecture : en moyenne un trimestre donné (et donc pour un sous-échantillon de rang 1 donné), et après redressement des effets du protocole de collecte liés à l'imputation des ménages d'inactifs, 429 inactifs répondant en vague 1 ne répondent plus en vague 2, soit 5,5% des effectifs d'inactifs de vague 1. Ces 429 sorties se décomposent en 385 sorties effectivement enregistrées en vague 2 dans l'enquête, et un nombre estimé de 44 sorties de personnes appartenant à des ménages d'inactifs de 65 ans qui auraient pu être observées à cette vague-là si les données des personnes concernées n'avaient pas été imputées jusqu'en vague 6. Ce nombre de 44 sorties est estimé en répartissant les 261 sorties observées en vague 6 pour des inactifs de 65 ans ou plus dont les données ont été imputées en vague 5 sur l'ensemble des cinq vagues auxquelles elles ont pu se produire.

Annexe

Tableau –estimation des probabilités de sorties par vague en fonction des paramètres du modèle et comparaison avec les probabilités observées.

a) Probabilités de sortie totale

statut d'activité initial	probabilités de sortie en fonction de la vague					
	totale					
	estimation à partir des paramètres du modèle					
	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy
actif occupé	6,9%	6,9%	6,9%	6,9%	10,3%	7,8%
chômeur	11,0%	11,0%	11,0%	11,0%	12,9%	11,7%
inactif	6,5%	6,5%	6,5%	6,5%	10,6%	7,7%
	observées					
	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy
actif occupé	6,7%	6,9%	7,0%	7,0%	10,5%	7,6%
chômeur	11,2%	10,5%	10,6%	11,3%	13,3%	11,3%
inactif	4,9%	5,1%	4,9%	4,8%	10,2%	6,0%
<i>imputé</i>		0,0%	0,0%	0,0%	12,3%	3,2%
<i>non imputé</i>	6,5%	6,8%	6,6%	6,6%	9,4%	7,2%
	écart estimé-observé					
	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy
actif occupé	0,2%	0,1%	-0,1%	0,0%	-0,2%	0,1%
chômeur	-0,2%	0,5%	0,4%	-0,3%	-0,4%	0,3%
inactif	1,5%	1,4%	1,6%	1,6%	0,4%	1,6%

b) Probabilités de sortie du champ et pour autre non-réponse

statut d'activité initial	probabilités de sortie en fonction de la vague												
	pour sortie de champ						pour autre non-réponse (conditionnellement à la présence dans le champ)						
	estimation à partir des paramètres du modèle												
	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy	moy 1 à 4
actif occupé	2,7%	2,7%	2,7%	2,7%	4,0%	2,9%	4,4%	4,4%	4,4%	4,4%	6,6%	4,8%	4,4%
chômeur	4,8%	4,8%	4,8%	4,8%	5,7%	5,0%	6,5%	6,5%	6,5%	6,5%	7,6%	6,7%	6,5%
inactif	2,3%	2,3%	2,3%	2,3%	4,9%	2,9%	4,3%	4,3%	4,3%	4,3%	6,0%	4,7%	4,3%
	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy	moy 1 à 4
actif occupé	2,4%	2,6%	2,9%	2,6%	4,2%	2,9%	4,4%	4,4%	4,2%	4,5%	6,6%	4,8%	4,4%
chômeur	4,7%	4,8%	4,6%	4,7%	6,2%	5,0%	6,8%	6,0%	6,3%	6,9%	7,6%	6,7%	6,5%
inactif	1,7%	1,9%	1,9%	1,7%	4,5%	2,3%	3,6%	3,6%	3,3%	3,5%	4,8%	3,8%	3,5%
<i>imputé</i>		0,0%	0,0%	0,0%	6,3%	1,7%							
<i>non imputé</i>	2,3%	2,5%	2,6%	2,3%	3,8%	2,7%							
	écart estimé-observé												
	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy	moy 1 à 4
actif occupé	0,3%	0,1%	-0,2%	0,1%	-0,2%	0,0%	-0,1%	0,0%	0,2%	-0,1%	0,0%	0,0%	0,0%
chômeur	0,1%	0,0%	0,2%	0,1%	-0,4%	0,0%	-0,3%	0,5%	0,2%	-0,4%	0,0%	0,0%	0,0%
inactif	0,5%	0,4%	0,3%	0,5%	0,4%	0,6%	0,7%	0,7%	1,0%	0,9%	1,2%	1,0%	0,8%

Le modèle d'ajuste parfaitement (de manière mécanique) sur l'observation s'agissant de la probabilité moyenne de sortie de champ et de la probabilité conditionnelle d'autre non-

réponse moyenne sur les vagues téléphoniques, et ce, pour les actifs occupés et les chômeurs. Pour les inactifs, il aurait fallu également traiter dans les prédictions le cas des logements imputés (qui ne participent pas aux estimations pour les premières vagues), ce qui n'a pas été jugé utile ici, les inactifs n'intervenant pas dans les estimations de taux de chômage. Toujours pour les actifs occupés et les chômeurs, les écarts entre le modèle et l'observé ne dépassent jamais 0,5 points, quelle que soit la vague ou le type de sortie considéré.

Tableau 6 – Statut d’activité des personnes entrant ou sortant de l’échantillon de répondants entre deux vagues consécutives

répartition des entrées d'échantillon par statut d'activité final	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy.
actif occupé	40,3%	42,1%	43,1%	43,3%	42,1%	42,0%
chômeur	5,7%	7,3%	6,9%	6,5%	7,4%	6,7%
Inactif	54,0%	50,6%	50,0%	50,2%	50,5%	51,3%
Total	100%	100%	100%	100%	100%	100%
<i>taux de chômage</i>	12,4%	14,8%	13,8%	13,1%	14,9%	13,8%
répartition des sorties d'échantillon par statut d'activité initial	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy.
actif occupé	45,1%	45,0%	46,3%	46,1%	39,5%	43,8%
chômeur	8,9%	7,9%	7,9%	8,2%	5,4%	7,3%
Inactif	46,1%	47,1%	45,8%	45,7%	55,1%	48,9%
Total	100%	100%	100%	100%	100%	100%
<i>taux de chômage</i>	16,5%	15,0%	14,5%	15,1%	11,9%	14,4%

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, personnes présentes à au moins une des deux vagues étudiées

Tableau 3 – Variations d’effectifs par statut d’activité entre deux vagues consécutives, ventilées selon leur cause (changement de statut d’activité au sens du B.I.T. / entrée ou sortie de l’échantillon de répondants)

soldes des entrées et sorties d'échantillon	1 à 2	2 à 3	3 à 4	4 à 5	5 à 6	moy.
actif occupé	-165	18	27	93	-16	-9
changement de statut B.I.T.	-26	-13	-9	-3	-12	-13
entrées sorties échantillon	-139	31	35	96	-4	4
chômeur	1	22	16	25	-13	10
changement de statut B.I.T.	0	16	7	0	11	7
entrées sorties échantillon	1	7	10	25	-24	4
inactif	-279	-26	-26	47	111	-35
changement de statut B.I.T.	27	-3	2	3	1	6
entrées sorties échantillon	-306	-24	-28	44	110	-41
total - solde entrées sorties échantillon	-443	13	17	165	82	-33

Source : Enquêtes Emploi, T1-2003 au T1-2007. Données non pondérées

Champ : France métropolitaine, échantillons entrés entre le T1 2003 et le T4 2005, personnes présentes à au moins une des deux vagues étudiées

total - solde (valeur absolue) entrées sorties échantillon	2116	1741	1718	1559	2904	2008
---	-------------	------	------	------	-------------	------