

# PERSEVERANCE, ACCESSIBILITE ET ACCEPTATION :

## UNE NOUVELLE MODELISATION DE LA NON-REPONSE

*Xavier d'HAULTFOEUILLE (\*), Philippe FEVRIER<sup>1</sup> (\*)*

*(\*) INSEE, Unité Méthodes Statistiques*

### Introduction

Le premier indicateur de la réussite d'une enquête est son taux de non-réponse. Un faible taux de non-réponse permet d'avoir d'une part des estimations plus précises, puisque le nombre de répondants est plus important, et d'autre part des estimations moins biaisées, puisque le phénomène de non-réponse peut être lié aux variables d'intérêt de l'enquête. La non-réponse, son impact sur les biais des estimations ainsi que sa correction par repondération ont été étudiés par de nombreux auteurs (Cheng, 1998 ; Colombo, 1992 ; Deville et Särndal, 1992 ; Drew et Fuller, 1980 ; Ellis et alii, 1970 ; Fillion, 1967 ; Fitzgerald et Fuller, 1982 ; Foster, 1997 ; Fuller, 1974 ; Lin et Schaeffer, 1995 ; Smith, 1984 ; Stinchcombe et al., 1981). Ces études sur la non-réponse utilisent souvent le nombre total d'essais, réalisés par l'enquêteur pour obtenir un entretien avec le ménage, dans l'estimation de la probabilité de réponse de ce ménage. Mais les essais réalisés avant et après un premier contact avec le ménage sont de natures très différentes : leur issue dépend pour les premiers de la présence ou de l'accessibilité du ménage, pour les seconds de son bon vouloir. Cet indicateur mélange donc la facilité de contact du ménage et sa propension à accepter l'entretien. Lynn et al. (2002) étudient ces deux effets séparément et montrent qu'il n'y a pas d'interaction entre ceux-ci.

En fait, le phénomène de non-réponse est encore plus complexe, puisque la facilité de contact du ménage peut elle-même être décomposée en deux aspects, qui, à notre connaissance, n'ont pas été étudiés dans la littérature sur le sujet. En effet, pour contacter un ménage, il faut d'une part que l'enquêteur décide de faire un essai et d'autre part que le ménage soit accessible au moment de l'essai. La première dimension, que l'on peut considérer comme un effet enquêteur, est trop souvent oubliée et le concept de contact réduit à son deuxième aspect. Notons que ce nouvel « effet enquêteur » est très différent de l'effet enquêteur habituellement étudié. Jusqu'à présent et depuis Hanson et Marks en 1958, il a surtout été analysé pour comprendre l'impact de l'enquêteur sur la non-réponse partielle et sur les modalités "Ne sait pas" des questions d'opinions (Bailar et alii. (1977), Hox, De Leeuw et Kreft (1991), Pickery et Loosveldt (2000)). Ces études ont montré que l'enquêteur n'était pas "neutre" mais avait une influence sur la façon dont les personnes interrogées répondaient. Le nouvel « effet enquêteur » que nous introduisons diffère fondamentalement du précédent, puisqu'il ne résulte pas de l'interaction entre enquêteur et enquêté, mais de l'effort que l'enquêteur est prêt à fournir pour contacter le ménage.

Le phénomène de non-réponse doit donc être modélisé en trois étapes. Dans la première étape, l'enquêteur décide du nombre d'essais qu'il veut réaliser pour contacter le ménage. Son effort dépend à la fois de ses propres caractéristiques, mais peut également dépendre des caractéristiques du ménage qu'il cherche à contacter. Dans un deuxième temps, le ménage est plus ou moins facilement accessible. A chaque essai de l'enquêteur, il a donc une certaine probabilité d'être contacté, qui dépend bien sûr de ses caractéristiques. Enfin, une fois le contact établi (s'il y a contact), le ménage peut décider ou non de répondre à l'enquête.

Cette modélisation permet d'obtenir des estimations très fines du phénomène de non-réponse et de répondre à de nombreuses questions jusqu'alors sans réponse : L'enquêteur différencie-t-il son effort en fonction des caractéristiques du ménage qu'il connaît ? Si oui, quels sont les ménages pour lesquels il ne fait pas d'effort ? Ces ménages correspondent-ils à des ménages qui sont facilement accessibles ou au contraire à des ménages qu'il est difficile d'atteindre ? Les ménages appartenant à

---

<sup>1</sup> Nous tenons à remercier Thierry Magnac, Daniel Verger pour leurs remarques sur cette étude. Nous remercions tout particulièrement Jean-Marc Robin pour ses conseils et le suivi de ce travail.

la catégorie qui refuse plus facilement l'enquête sont-ils moins contactés ? Cette modélisation en trois étapes permet donc de décomposer les différents effets des caractéristiques des ménages : un effet indirect sur la persévérance des enquêteurs et deux effets directs distincts : l'un concernant l'accessibilité et l'autre le refus. Elle permet également de distinguer deux effets enquêteurs : un effet dans la probabilité de refus et un effet de persévérance et d'effort, peu étudié usuellement.

Le papier est organisé comme suit. La première partie présente les données. La deuxième partie est consacrée à l'effet enquêteur, c'est-à-dire au choix de son effort pour contacter un ménage. La troisième partie se consacre à l'étude de l'accessibilité des ménages. La quatrième partie enfin analyse le phénomène de refus.

## 1. Les données

L'étude, basée sur les enquêtes PCV d'octobre 2000 et 2001, utilise trois sources de données : les fichiers des deux enquêtes, les fiches-adresses associées à l'édition d'octobre 2001 et des données relatives aux enquêteurs ayant participé à ces deux enquêtes.

Les enquêtes PCV comportent trois éditions annuelles, réalisées en janvier, mai et octobre et portant chacune sur un thème précis<sup>2</sup>. Elles sont réalisées auprès de 8000 ménages environ, qui sont interrogés deux fois, à un an d'intervalle. Ne sont considérés, pour étudier le contact, que les ménages en 1<sup>ère</sup> interrogation. En effet, en 2<sup>ème</sup> interrogation, le comportement de l'enquêteur et du ménage lors de la phase de contact dépendent de l'issue de la première interrogation. Les efforts de l'enquêteur peuvent dépendre du contact ou non la première fois, du refus éventuel ou des caractéristiques observées du ménage ; quant au ménage, il peut développer des stratégies d'évitement ou au contraire se rendre davantage disponible. Ceci rendrait par conséquent l'analyse particulièrement complexe. L'étude se restreint également aux logements dans le champ de l'enquête, excluant par exemple les logements vacants, détruits ou les résidences secondaires pour lesquelles la notion de contact et de persévérance de l'enquêteur n'ont plus de sens. Cette restriction est possible dans la mesure où le résultat de collecte (acceptation, refus, hors-champ...) est connu pour chaque logement dans l'échantillon.

Finalement, en octobre 2001, l'échantillon initial des premières interrogations était composé de 4407 logements ; 645 étaient hors-champ et parmi les 3762 restants, 2894 ménages ont accepté de répondre, soit un taux d'acceptation de 76,9%.

Les données de PCV sont appariées avec celles de la fiche-adresse de l'enquête d'octobre 2001. Une fiche-adresse est un document de quatre pages remis à l'enquêteur pour chaque logement qu'il doit enquêter. Elle indique en première page l'adresse précise du logement et certaines de ses caractéristiques, qui permettent en théorie de l'identifier avec certitude. Y figurent notamment le type du logement (individuel ou collectif) et le nombre de pièces du logement. On dispose également d'informations sur le ménage qui occupait le logement lors du RP : en particulier, sa taille, ainsi que le nom, le prénom, l'âge et la profession<sup>3</sup> de la personne de référence. Ces variables ne sont pas remplies pour les logements vacants au moment du RP ou pour les logements neufs<sup>4</sup>. Sur les pages 2 et 3, l'enquêteur renseigne quelques variables générales sur le logement et le ménage au moment de l'enquête. Enfin, en quatrième page de la fiche-adresse, l'enquêteur indique dans un tableau la chronologie des essais qu'il a réalisés en vue d'obtenir un entretien avec le ménage. Pour chacun de ces essais, il indique le jour et l'heure de l'essai, le moyen utilisé (téléphone ou visite) et le résultat de l'essai : par exemple, ménage non contacté, ménage contacté et refus, prise de rendez-vous pour un prochain entretien ou encore entretien réalisé. Il est donc possible de différencier un ménage non contacté au bout de sept essais d'un ménage non contacté après une seule tentative de contact, c'est-à-dire d'isoler l'effet du ménage de l'effet de l'enquêteur dans la phase de contact.

---

<sup>2</sup> Qualité de l'habitat et de l'environnement en janvier, santé, logement et endettement des ménages en mai et participations et contacts sociaux en octobre. A ces thèmes fixes s'ajoute souvent une partie variable différente d'une enquête à l'autre.

<sup>3</sup> malheureusement inexploitable du fait du grand nombre de valeurs manquantes

<sup>4</sup> Ceux-ci peuvent être repérés par une autre variable

Enfin, les informations disponibles sur les enquêteurs sont le nombre d'années d'expérience, le sexe, le statut matrimonial et la pratique d'une autre activité professionnelle. Le diplôme est également disponible mais n'a pas été utilisé car la variable est souvent non renseignée. Enfin, la région sera parfois utilisée comme une variable enquêteur.

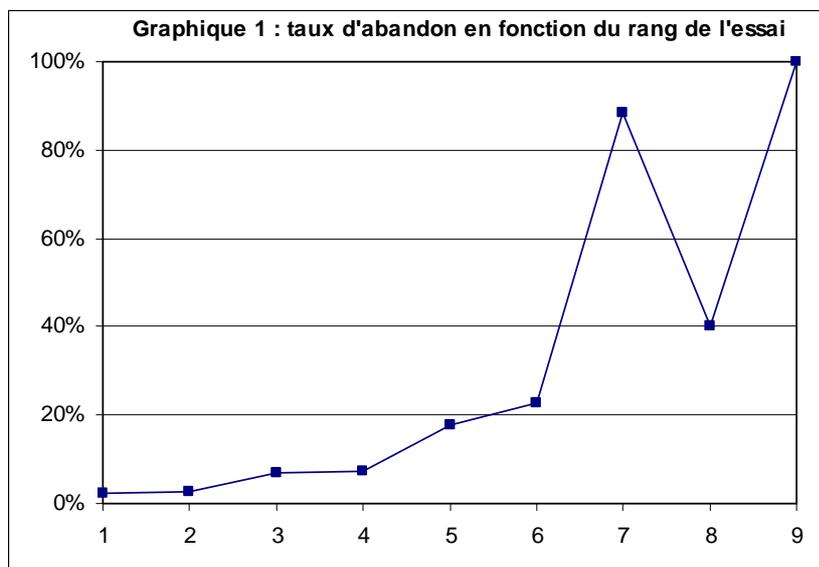
Ces données permettent donc de distinguer les deux phases principales du processus de collecte, à savoir le contact et l'acceptation du ménage. Le contact dépend lui-même de la persévérance de l'enquêteur, définie comme le nombre d'essais de contact<sup>5</sup> qu'il est prêt à réaliser, et de l'accessibilité du ménage, qui correspond au nombre d'essais nécessaire pour le contacter.

## 2. La persévérance des enquêteurs

La persévérance des enquêteurs est cruciale à plus d'un titre<sup>6</sup>. Elle conditionne la qualité de l'enquête : une faible persévérance entraîne une non-réponse importante, et donc une augmentation de la variance et du biais. Par ailleurs, l'institut peut l'utiliser comme levier pour diminuer la non-réponse. En effet, comme on le verra en 2.1, la consigne sur le nombre d'essais à réaliser avant abandon est plutôt bien suivie par les enquêteurs. Une modification adéquate de celle-ci est donc susceptible d'entraîner une diminution de la non-réponse.

### 2.1. Statistiques descriptives

La persévérance des enquêteurs peut être mesurée globalement par le taux d'abandon, c'est-à-dire la probabilité que l'enquêteur abandonne après chaque essai infructueux<sup>7</sup>. Ce taux peut être comparé à la consigne de l'INSEE, qui est de réaliser au minimum 5 essais de contact avant d'abandonner le ménage.



Source : fiches-adresses de l'enquête PCV d'octobre 2001

Champ : ménages en 1<sup>ère</sup> interrogation dans le champ de l'enquête

Lecture : après le 7<sup>ème</sup> essai de contact, les enquêteurs ont abandonné dans 89% des cas.

L'estimation du taux d'abandon permet d'évaluer le nombre d'essais que les enquêteurs sont prêts à réaliser en moyenne : il atteint 5,92 en 2001.

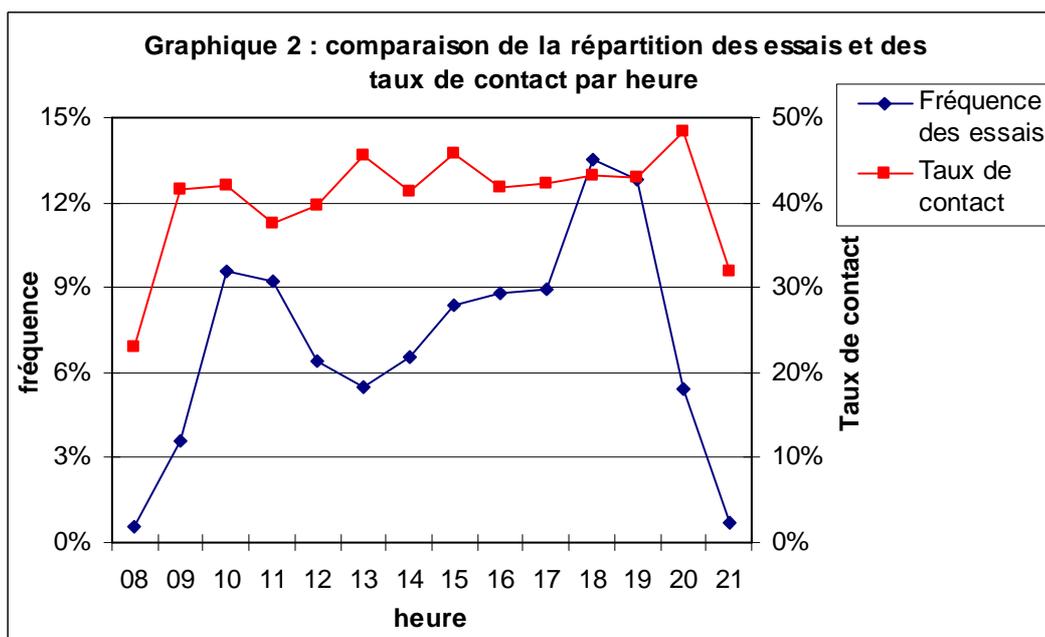
<sup>5</sup> On ne tient pas compte des essais réalisés après un premier contact car ces essais ne dépendent pas seulement de l'enquêteur mais également de la réaction du ménage lors de ce premier contact.

<sup>6</sup> Il existe un cas où la persévérance est inutile voire coûteuse, celui des ménages « absents pour une longue durée », c'est-à-dire pendant toute la durée de la collecte. En toute rigueur, il serait nécessaire de ne pas considérer ces ménages dans l'analyse de la persévérance et de l'accessibilité.

<sup>7</sup> Le taux d'abandon est donc un taux de hasard en temps discret.

Par ailleurs, le graphique 1 montre que la consigne est globalement bien suivie par les enquêteurs : le taux d'abandon reste faible après 4 essais ou moins. Le fort taux d'abandon observé après 7 essais s'explique peut-être par le fait que le tableau d'essais de contact ne laisse la place qu'à 7 essais. Il est certes demandé aux enquêteurs de noter les essais supplémentaires sur une feuille à part, mais cette consigne contraignante n'est probablement pas toujours respectée. Il est également possible que ce tableau de 7 essais joue comme une consigne implicite de ne pas aller au-delà de ce chiffre.

Il est à noter que la persévérance n'est pas le seul facteur dépendant de l'enquêteur et influant sur le taux de contact des ménages. Le choix des jours et des heures auxquels sont réalisés les essais de contact, ainsi que la dispersion des moments choisis, influent sur le taux de contact comme le montre le graphique ci-dessous :



Source : fiche-adresses de l'enquête PCV d'octobre 2001

Champ : essais de contact pour les ménages en 1<sup>ère</sup> interrogation dans le champ de l'enquête

Lecture : 9% des essais ont été réalisés à 10 heures, et le taux de contact obtenu a été de pratiquement 40%.

Les enquêteurs réalisent donc la plupart des essais entre 10 et 11 heures le matin, 18 et 19 heures l'après-midi. Par ailleurs, on constate que le taux contact reste remarquablement constant en fonction des heures de la journée (mis à part pour 8 heures et 21 heures). On se serait pourtant attendu à ce qu'il soit beaucoup plus faible pendant la journée qu'en soirée. Cela indique très certainement que les enquêteurs choisissent les heures des essais en fonction des caractéristiques qu'ils connaissent des ménages, et réussissent ainsi à obtenir de bons taux de contact même aux heures « creuses ». Par contre, ils réalisent peu d'essais à 13 heures et à 20 heures alors que le taux de contact y est élevé.

## 2.2. Modélisation

Un enquêteur est donc prêt en moyenne à réaliser 5,9 essais, les étale au cours de la journée et respecte correctement la consigne de l'INSEE de réaliser plus de 5 essais en cas de non-contact.

Nous essayons maintenant d'analyser plus finement le comportement de persévérance des enquêteurs et de comprendre comment l'enquêteur choisit le nombre d'essais qu'il va réaliser pour contacter un ménage. Ce choix dépend évidemment des caractéristiques propres de l'enquêteur (hétérogénéité des enquêteurs), mais également des caractéristiques du ménage à contacter. L'enquêteur observe en effet les caractéristiques indiquées sur la fiche-adresse du ménage occupant le logement au RP. Il peut donc faire dépendre son effort de ces caractéristiques. Cet effort traduit en quelle que sorte la stratégie de l'enquêteur pour contacter un ménage, sachant que les

caractéristiques du ménage reflètent en partie la probabilité de contact ainsi que la probabilité de refus après un contact du ménage considéré.

L'estimation de la persévérance a été réalisée sur les données de l'enquête PCV 2001 en se restreignant aux logements dans le champ de l'enquête que l'enquêteur doit effectivement essayer de contacter. Comme précédemment, nous limitons également notre échantillon aux personnes en première interrogation. Enfin, sont exclus de l'analyse les ménages classés « absent de longue durée », c'est-à-dire ceux qui étaient absents du logement durant toute la période de collecte. En effet, la persévérance n'est pas souhaitable dans ce cas puisque le ménage a une probabilité nulle d'être contacté.

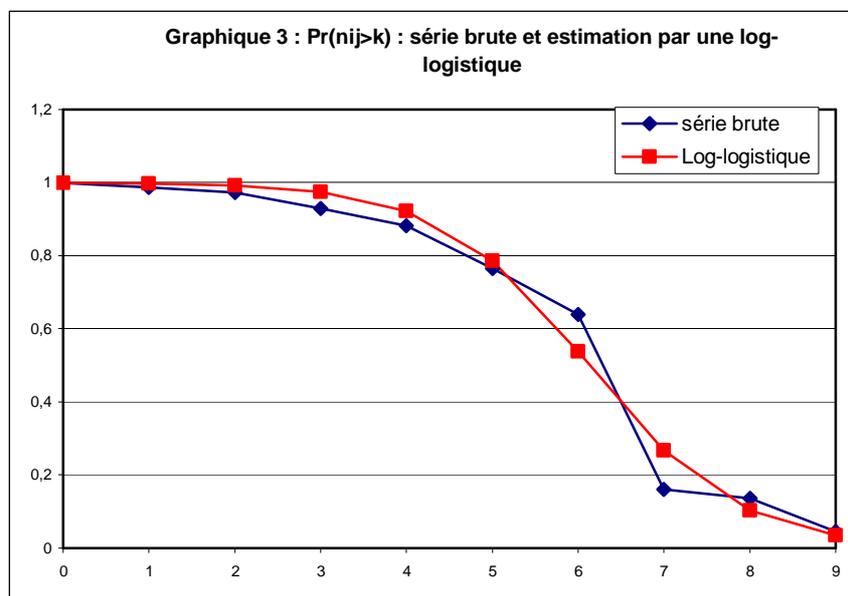
Soit  $\tilde{n}_{ij}$  le nombre d'essais réalisé par un enquêteur  $i$  pour contacter un ménage  $j$  et observé dans les données. Ce nombre d'essais correspond soit à la persévérance de l'enquêteur, s'il n'y a pas eu de contact, soit à une borne inférieure de cette persévérance en cas de contact. Ainsi, en notant  $n_{ij}$  la persévérance et  $c_{ij}$  l'indicatrice indiquant s'il y a eu contact ou non, nous obtenons :  $n_{ij} = \tilde{n}_{ij}$  s'il n'y a pas eu de contact ( $c_{ij} = 0$ ) et  $n_{ij} \geq \tilde{n}_{ij}$  en cas de contact ( $c_{ij} = 1$ ).

Nous modélisons par ailleurs la persévérance de l'enquêteur grâce au modèle log-logistique suivant<sup>8</sup> :

$$\ln(n_{ij}^*) = X_i \beta + X_j^* \gamma + \sigma \varepsilon_{ij}$$

où  $\varepsilon_{ij}$  suit une loi logistique.  $X_i$  représente les caractéristiques de l'enquêteur tandis que  $X_j^*$  correspond aux variables « ménage » indiquées sur la fiche adresse. Il est important de noter que l'enquêteur ne peut décider du nombre d'essais qu'il va effectuer qu'à partir de ces seules caractéristiques puisque les caractéristiques réelles du ménage lui sont inconnues tant qu'il ne l'a pas contacté.

Nous avons retenu un modèle log-logistique pour modéliser la persévérance de l'enquêteur, car c'est celui qui s'ajuste le mieux aux données (sans introduction de variables exogènes) comme le montre le graphique suivant où la série brute est la série estimée par la méthode de Kaplan-Meier.



Source : fiches-adresses de l'enquête PCV d'octobre 2001

Champ : essais de contact des ménages en 1<sup>ère</sup> interrogation dans le champ de l'enquête

<sup>8</sup> Il est possible de mettre des effets individuels pour les enquêteur dans la modélisation. Nous n'analysons ici que le modèle le plus simple. Il est également à noter que cette modélisation ne permet pas d'introduire la contrainte de l'INSEE, ni la « consigne implicite » des 7 essais.

Enfin, puisque le nombre d'essais  $n_{ij}$  effectué par l'enquêteur doit être entier, nous supposons que  $n_{ij}$  (qui varie entre 0 et 9) est obtenu à partir de  $n_{ij}^*$  à partir de la transformation suivante :

$$\begin{aligned} n_{ij} &= E[n_{ij}^*] + 1 \quad \text{si } n_{ij}^* < 8 \\ n_{ij} &= 9 \quad \text{si } n_{ij}^* \geq 8 \end{aligned}$$

La vraisemblance du modèle s'écrit donc :

- si  $c_{ij} = 0$  et  $\tilde{n}_{ij} < 9$ ,

$$\begin{aligned} l(\tilde{n}_{ij}, c_{ij}) &= l(\tilde{n}_{ij}, 0) = P(n_{ij} = \tilde{n}_{ij}) \\ &= P(\tilde{n}_{ij} - 1 \leq n_{ij}^* < \tilde{n}_{ij}) \\ &= \frac{1}{1 + e^{-\frac{1}{\sigma}[\ln(\tilde{n}_{ij}) - X_i\beta - X_i^*\gamma]}} - \frac{1}{1 + e^{-\frac{1}{\sigma}[\ln(\tilde{n}_{ij}-1) - X_i\beta - X_i^*\gamma]}} \end{aligned}$$

- si  $c_{ij} = 1$  et  $\tilde{n}_{ij} < 9$ ,

$$\begin{aligned} l(\tilde{n}_{ij}, c_{ij}) &= l(\tilde{n}_{ij}, 1) = P(n_{ij} \geq \tilde{n}_{ij}) \\ &= P(n_{ij}^* \geq \tilde{n}_{ij} - 1) \\ &= \frac{e^{-\frac{1}{\sigma}[\ln(\tilde{n}_{ij}-1) - X_i\beta - X_i^*\gamma]}}{1 + e^{-\frac{1}{\sigma}[\ln(\tilde{n}_{ij}-1) - X_i\beta - X_i^*\gamma]}} \end{aligned}$$

- enfin, si  $\tilde{n}_{ij} = 9$ ,

$$\begin{aligned} l(\tilde{n}_{ij}, c_{ij}) &= l(9, c_{ij}) = P(n_{ij} = 9) \\ &= P(n_{ij}^* \geq 8) \\ &= \frac{e^{-\frac{1}{\sigma}[\ln(8) - X_i\beta - X_i^*\gamma]}}{1 + e^{-\frac{1}{\sigma}[\ln(8) - X_i\beta - X_i^*\gamma]}} \end{aligned}$$

Les paramètres du modèle sont finalement estimés par la méthode du maximum de vraisemblance. Les résultats sont donnés dans le tableau 1 page suivante : un coefficient positif correspond à une plus grande persévérance, et les modalités en gras ont un coefficient significatif à 5%.

**Tableau 1 : estimation des facteurs de la persévérance des enquêteurs**

Variable	Estimation	Ecart-type
<b>Constante</b>	<b>1,95</b>	<b>0,13</b>
Logement recensé au RP90	Réf.	/
Logement neuf	-0,07	0,11
Logement d'une pièce	Réf.	
Logement de 2 pièces	-0,09	0,07
Logement de 3 pièces	-0,06	0,08
Logement de 4 pièces	-0,02	0,08
Logement de 5 pièces ou plus	-0,04	0,09
Pas d'indication sur le nombre de pièces	-0,15	0,12
Logement individuel	Réf.	
<b>Logement collectif</b>	<b>-0,14</b>	<b>0,06</b>
Logement en zone rurale	-0,07	0,07
Logement en petite zone urbaine	0,07	0,08
Logement en zone urbaine de moyenne taille	0,07	0,06
Logement en grande zone urbaine	Réf.	
<b>Région Ile de France</b>	<b>-0,19</b>	<b>0,09</b>
Régions du centre	-0,08	0,09
Région Nord Pas de Calais	Réf.	
Régions de l'est	-0,04	0,11
Régions de l'ouest	0,05	0,10
Région du sud-ouest	0,02	0,10
Régions Rhône Alpes et Auvergne	0,01	0,10
<b>Régions Paca et Languedoc-Roussillon</b>	<b>-0,21</b>	<b>0,09</b>
Personne de référence âgée de moins de 30 ans	0,14	0,08
Personne de référence âgée de 30 à 65 ans	-0,01	0,07
Personne de référence âgée de plus de 65 ans	Réf.	
Pas d'indication sur l'âge de la personne de réf.	-0,02	0,09
Enquêteur féminin	Réf.	
Enquêteur masculin	0,00	0,06
Ancien enquêteur (expérience ≥ 4 ans)	Réf.	
<b>Nouvel enquêteur (expérience &lt; 4 ans)</b>	<b>-0,10</b>	<b>0,05</b>
<b>Jeune enquêteur (35 ans ou moins) marié</b>	<b>-0,25</b>	<b>0,12</b>
Autre situation	Réf.	/
Pratique d'une autre activité professionnelle	0,00	0,04
Aucune autre activité professionnelle	Réf.	/
<b>Sigma</b>	<b>0,21</b>	<b>0,01</b>

Source : données enquêteur et fiche-adresses de l'enquête PCV d'octobre 2001

Champ : ménages en 1<sup>ère</sup> interrogation dans le champ de l'enquête

Les changements de comportements de l'enquêteur sont donc principalement liés au statut du logement : l'enquêteur insiste davantage auprès des ménages habitant en logement individuel. Ceci s'explique sans doute par les barrières à l'entrée des logements collectifs, comme les digicodes et interphones, qui peuvent dissuader de persévérer dans la mesure où elles interdisent l'accès direct au logement. Par ailleurs, il semble que les enquêteurs soient prêts à faire plus d'effort pour contacter les jeunes<sup>9</sup>.

Pour les variables «enquêteur», si le tableau montre que le sexe et le fait d'avoir une autre activité ne joue pas sur la persévérance, il indique au contraire qu'il semble y avoir un effet régional fort, ainsi qu'un rôle important de l'expérience des enquêteurs. Deux régions se distinguent : la région parisienne et la région sud (PACA et Languedoc-Roussillon). Dans ces régions, les enquêteurs ont tendance à faire moins d'effort que dans le reste de la France. Enfin, les « nouveaux » enquêteurs qui ont moins de quatre ans d'expérience ont également tendance à moins persévérer que les autres. L'INSEE doit donc particulièrement insister sur l'importance de la persévérance pour les jeunes embauchés.

<sup>9</sup> La modalité est significative à 10%

### 3. Accessibilité des ménages

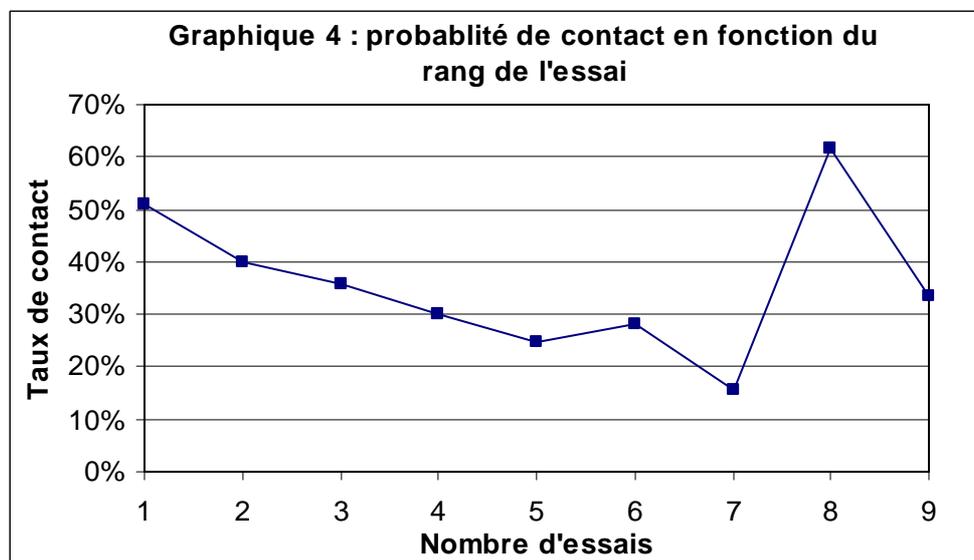
L'accessibilité du ménage est le deuxième aspect crucial dans la phase de contact. La persévérance est en effet inutile si le ménage est impossible à contacter. Cette accessibilité mélange en fait deux aspects : l'accès au ménage et sa présence au moment du passage de l'enquêteur. L'accès peut être compromis par des barrières à l'entrée du logement comme les digicodes ou les interphones, et par l'impossibilité d'obtenir un numéro de téléphone du fait des listes rouges et de la multiplication des portables. La présence du ménage dans son logement dépend a priori principalement de son activité professionnelle.

Les situations en terme d'accessibilité sont donc très diverses, et cette hétérogénéité produit une dispersion dans les pondérations finales. Une analyse des déterminants de l'accessibilité permet de corriger cet effet en amont et en aval de l'enquête. En amont, les ménages identifiés comme difficile à joindre peuvent être surreprésentés lors du tirage, afin de réduire la dispersion finale des poids. En aval, le modèle présenté ici peut être utilisé pour calculer la probabilité de contacter le ménage.

Par ailleurs, l'étude séparée de l'accessibilité et de la persévérance des enquêteurs permet de savoir si les enquêteurs insistent davantage sur les ménages difficiles à joindre ou au contraire concentrent leurs efforts sur les ménages qui présentent des caractéristiques favorables a priori au contact. Leur comportement d'anticipation peut alors être comparé au nombre d'essais optimal qui devrait être fait étant donné les caractéristiques du ménage interrogé. Une politique adéquate (formation, incitation monétaire...) doit permettre d'approcher cet optimum.

#### 3.1. Statistiques descriptives

L'accessibilité peut être mesurée à partir de la probabilité de contacter le ménage lors d'un essai. Elle s'élève en moyenne à 42%. On observe cependant une grande disparité suivant le rang de l'essai de contact.



Source : fiche-adresses de l'enquête PCV d'octobre 2001

Champ : essais de contact des ménages en 1<sup>ère</sup> interrogation dans le champ de l'enquête

Lecture : au 2<sup>ème</sup> essai de contact, les ménages ont été contactés dans 40% des cas

La baisse du taux de contact en fonction du rang de l'essai provient de l'hétérogénéité des ménages : les ménages faciles à joindre sont contactés lors des premiers essais et les ménages restant sont de plus en plus difficiles à contacter. La remontée du taux de contact au 8<sup>ème</sup> essai pourrait provenir d'une sélection endogène : les enquêteurs, qui réalisent rarement 8 essais de contact, n'y consentent que lorsqu'une information auxiliaire leur donne bon espoir de contacter le ménage. Notons que ce

comportement, certes marginal, de persévérance de l'enquêteur n'a pas été pris en compte dans le modèle présenté en 2.2.

L'étude de la persévérance des enquêteurs et de l'accessibilité des ménages permet d'introduire formellement la notion de contact : celui-ci a lieu lorsque la persévérance des enquêteurs est supérieure à l'accessibilité des ménages. Dans l'enquête PCV d'octobre 2001, 90,4% des ménages en première interrogation dans le champ de l'enquête ont été contactés. Les ménages non contactés sont théoriquement soit des « impossibles à joindre », soit des « absents de longue durée »<sup>10</sup>. En pratique, on observe un certain flou dans le classement : quelques ménages non contactés ont été classés en refus, d'autres en inapte.

**Tableau 2 : résultats de collecte pour les ménages non contactés**

Résultat de collecte	Fréquence
Enquête réalisée	0,0%
Refus	5,0%
Impossible à joindre	76,9%
Absent de longue durée	17,1%
Inapte	1,1%

*Source : fiche-adresses de l'enquête PCV d'octobre 2001  
Champ : essais de contact des ménages en 1<sup>ère</sup>  
interrogation dans le champ de l'enquête*

### 3.2. Modélisation

Comme nous venons de le voir, les ménages sont hétérogènes et il est important de comprendre quels sont les ménages facilement contactables. En effet, ces différences d'accessibilité entre ménages conduisent à des biais de sélection dans l'échantillon qui peuvent ensuite introduire des biais dans les estimations. Une partie de ces biais est corrigée par l'emploi de procédure de calage telle que CALMAR (procédure de calage sur marge utilisée à l'INSEE), mais ce n'est vraisemblablement pas suffisant. Une connaissance approfondie de l'accessibilité des ménages permettrait de diminuer les biais de deux manières. La première, au niveau même de la collecte, consiste à inciter les enquêteurs à faire plus d'efforts sur les ménages difficilement accessibles. La seconde, au niveau des repondérations, concerne la correction de la non-réponse. En effet, l'accessibilité est une étape importante dans la compréhension de la non-réponse, dont il faut tenir compte pour repondérer correctement les enquêtes.

On cherche donc à expliquer l'accessibilité des ménages par les caractéristiques de ces ménages et des logements qu'ils occupent. Malheureusement, ces informations ne sont observées que sur les répondants<sup>11</sup>. Il est donc nécessaire d'estimer un modèle de vraisemblance conditionnelle, où l'on a conditionné par le fait que le ménage est répondant i.e par le fait que le ménage a été contacté et qu'il a accepté de répondre à l'enquête.

Nous supposons qu'à chaque essai de l'enquêteur  $i$ , un ménage  $j$  est contacté avec une probabilité  $p(X_i, X_j)$  qui dépend des caractéristiques  $X_j$  du ménage mais aussi  $X_i$  de l'enquêteur. En effet, un enquêteur connaissant bien la zone d'enquête est susceptible de choisir les heures des essais de contact de manière adéquate, et d'augmenter ainsi la probabilité de contacter le ménage. Par contre, cette probabilité est considérée comme indépendante du rang de l'essai : l'enquêteur n'améliore pas

<sup>10</sup> La différence entre les deux notions est que l'enquêteur ne peut classer un ménage comme absent de longue durée que s'il sait de source sûre que le ménage sera absent de son logement durant la totalité de la collecte.

<sup>11</sup> Une autre possibilité consistait à se limiter à l'ensemble des individus qui n'ont pas déménagé depuis le Recensement de la Population. Cette variable est en effet renseignée par l'enquêteur, que le ménage soit interrogé ou non. Malheureusement, elle est de très mauvaise qualité et il n'est pas possible de l'utiliser dans l'exploitation de l'enquête.

sa stratégie au fur et à mesure des essais infructueux<sup>12</sup>. Si on note  $z_{ij}$  le nombre d'essais nécessaires à l'enquêteur  $i$  pour contacter le ménage  $j$ ,  $z_{ij}$  suit donc une loi géométrique de paramètre  $p(X_i, X_j)$ .

Soit encore  $\tilde{n}_{ij}$  le nombre d'essais, réalisé par un enquêteur  $i$  pour contacter un ménage  $j$ , observé dans les données. Ce nombre d'essais correspond soit au nombre  $z_{ij}$  d'essais nécessaires pour contacter le ménage, s'il y a eu contact, soit à une borne inférieure de ce nombre d'essais, lorsqu'il n'y a pas eu de contact. Ainsi, en notant toujours  $c_{ij}$  l'indicatrice indiquant s'il y a eu contact ou non, nous obtenons :

$$z_{ij} > \tilde{n}_{ij} \text{ s'il n'y a pas eu de contact } (c_{ij} = 0)$$

et :

$$z_{ij} = \tilde{n}_{ij} \text{ en cas de contact } (c_{ij} = 1).$$

Comme expliqué précédemment, nous n'observons les caractéristiques  $X_j$  du ménage qu'en cas de contact et d'acceptation. Nous verrons plus loin (cf. graphique 5) qu'il est possible de supposer l'indépendance entre  $z_{ij}$  et le fait d'accepter, mais il est nécessaire d'étudier plus précisément le contact. Il y a contact si et seulement si le nombre d'essais  $n_{ij}$  entrepris par l'enquêteur  $i$  (qui dépend des caractéristiques  $X_i$  de l'enquêteur ainsi que des caractéristiques du ménage observées sur la fiche adresse que nous avons notées  $X_j^*$  dans la partie précédente) est supérieur au nombre d'essais nécessaires pour contacter le ménage, i.e.  $n_{ij} \geq z_{ij}$ . Observer les caractéristiques du ménage revient donc ici à conditionner par le fait que  $n_{ij} \geq z_{ij}$ .

Nous pouvons enfin écrire la vraisemblance des observations conditionnellement au fait de les observer :

$$\begin{aligned} P(z_{ij} = k | X_j, X_i, X_j^*, z_{ij} \leq n_{ij}) &= \frac{P(z_{ij} = k, z_{ij} \leq n_{ij} | X_j, X_i, X_j^*)}{P(z_{ij} \leq n_{ij} | X_j, X_i, X_j^*)} \\ &= \frac{P(z_{ij} = k | X_i, X_j) P(k \leq n_{ij} | X_i, X_j^*)}{\sum_{l=1}^9 P(z_{ij} = l | X_i, X_j) P(l \leq n_{ij} | X_i, X_j^*)} \\ &= \frac{(1 - p(X_i, X_j))^{k-1} P(k \leq n_{ij} | X_i, X_j^*)}{\sum_{l=1}^9 (1 - p(X_i, X_j))^{l-1} P(l \leq n_{ij} | X_i, X_j^*)} \end{aligned}$$

où la deuxième égalité vient de l'indépendance entre  $z_{ij}$  et  $n_{ij}$  (qui est toujours inférieur à 9) conditionnellement aux variables  $X_i$  et  $X_j^*$  sur lesquelles l'enquêteur décide de l'effort qu'il va fournir, tandis que la troisième égalité vient de la loi géométrique de  $z_{ij}$ .

L'étude de la persévérance des enquêteurs nous a fourni une estimation des probabilités  $P(k \leq n_{ij} | X_i, X_j^*)$  que nous noterons  $\hat{p}_k(X_i, X_j^*)$ <sup>13</sup>. En remplaçant cette estimation dans l'équation précédente, nous obtenons :

$$\hat{P}(z_{ij} = k | X_j, X_i, X_j^*, z_{ij} \leq n_{ij}) = \frac{(1 - p(X_i, X_j))^{k-1} \hat{p}_k(X_i, X_j^*)}{\sum_{l=1}^9 (1 - p(X_i, X_j))^{l-1} \hat{p}_l(X_i, X_j^*)}$$

<sup>12</sup> cela revient à supposer que l'enquêteur n'intègre pas dans son choix de l'heure de l'essai l'information des horaires précédents ayant conduit à un échec

<sup>13</sup> Ces probabilités sont en fait estimées à partir d'une version simplifiée du modèle présenté en 2, incluant les variables âge (moins de 30 ans / autres), type de logement (collectif / individuel), région (Paris, Languedoc-Roussillon, PACA / autres), ancienneté (moins de 4 ans d'expérience / plus de 4 ans) et statut matrimonial (jeune marié / autres).

En faisant l'hypothèse que les probabilités de contact  $p(X_i, X_j)$  à chaque essai sont issues d'un modèle logistique, nous obtenons enfin :

$$p(X_i, X_j) = \frac{1}{1 + e^{-(X_i \delta_1 + X_j \delta_2)}}$$

et les paramètres  $\delta = (\delta_1, \delta_2)$  peuvent être estimés par maximum de vraisemblance, où la vraisemblance est estimée d'après les équations précédentes. Les résultats sont présentés dans le tableau 3<sup>14</sup> :

**Tableau 3 : estimation des déterminants l'accessibilité des ménages**

Variables	Estimation	Ecart-type
Constante	0,15	0,16
Enquêteur féminin	Réf.	/
Enquêteur masculin	-0,02	0,11
Ancien enquêteur (expérience $\geq$ 4 ans)	Réf.	/
Nouvel enquêteur (expérience $<$ 4 ans)	-0,05	0,10
Jeune enquêteur (35 ans ou moins) marié	0,40	0,29
Autre situation	Réf.	/
Pratique d'une autre activité professionnelle	-0,12	0,07
Aucune autre activité professionnelle	Réf.	/
Logement recensé au RP90	Réf.	/
Logement neuf	-0,19	0,11
Logement individuel	Réf.	/
Logement collectif	-0,08	0,09
<b>Logement en zone rurale</b>	<b>0,22</b>	<b>0,10</b>
<b>Logement en petite zone urbaine</b>	<b>0,22</b>	<b>0,10</b>
Logement en zone urbaine moyenne	-0,01	0,11
Logement en grande zone urbaine	Réf.	/
Région Ile de France	-0,18	0,17
Régions du centre	-0,18	0,14
Région Nord Pas de Calais	Réf.	/
Régions de l'est	-0,28	0,18
Régions de l'ouest	-0,24	0,16
Région du sud-ouest	-0,23	0,16
Régions Rhône Alpes et Auvergne	-0,25	0,16
Régions Paca et Languedoc-Roussillon	-0,31	0,16
Ménage d'une personne	Réf.	/
Ménage de 2 personnes	0,08	0,08
<b>Ménage de 3 personnes</b>	<b>0,32</b>	<b>0,10</b>
<b>Ménage de 4 personnes ou plus</b>	<b>0,65</b>	<b>0,14</b>
<b>Personne de réf. âgée de moins de 30 ans</b>	<b>-0,41</b>	<b>0,13</b>
Personne de réf. âgée de 30 à 59 ans	Réf.	/
<b>Personne de réf. âgée de 60 ans ou plus</b>	<b>0,38</b>	<b>0,09</b>
Revenus inférieurs à 100 000 F	Réf.	/
Revenus compris entre 100 000 F et 180 000 F	0,04	0,09
Revenus supérieurs à 180 000 F	0,05	0,10
Non-réponse sur la question des revenus	-0,13	0,17
Aucune sortie au cinéma par an	Réf.	/
Entre 1 et 6 sorties au cinéma par an	-0,04	0,09
Entre 7 et 12 sorties au cinéma par an	-0,19	0,15
<b>Plus de 12 sorties au cinéma par an</b>	<b>-0,26</b>	<b>0,10</b>

Source : enquête PCV d'octobre 2001  
 Champ : ménages répondants en 1<sup>ère</sup> interrogation

<sup>14</sup> Les écart-types donnés dans la table ne prennent pas en compte le fait que la vraisemblance a été estimée par l'intermédiaire des  $\hat{p}_k(X_i, X_j^*)$ .

En premier lieu, les caractéristiques observées des enquêteurs, notamment le nombre d'années de pratique, n'ont pas d'effet significatif sur la probabilité de contact des ménages. Les différences de stratégie d'un enquêteur à l'autre sont donc a priori négligeables, ou du moins n'ont pas de conséquences visibles sur les résultats de collecte.

De même, les variables liées au logement ont peu d'impact puisque le seul facteur discriminant est la strate d'habitation : les logements en zone rurale et en petite zone urbaine sont plus accessibles que les autres. Mais ni le type de logement (individuel ou collectif), ni la région n'influent sur l'accessibilité du ménage : en particulier, les problèmes d'accessibilité ne sont pas plus grand en Ile de France, toutes choses égales par ailleurs.

Par contre, les résultats obtenus confirment la grande hétérogénéité des ménages. Les ménages de grande taille sont plus accessibles que les ménages de petite taille ; et les caractéristiques du chef de ménage semblent également jouer un rôle important. Les personnes âgées sont facilement joignables, à l'inverse des jeunes. Les personnes qui vont très souvent au cinéma sont moins accessibles. Cette variable permet de prendre en compte une partie des activités en dehors du logement. Evidemment, moins le ménage est dans le logement, et moins il est facile de le contacter. Par contre, le revenu n'a pas d'effet, toutes choses égales par ailleurs.

Si les résultats semblent pour la plupart intuitifs, il est important de remarquer que cette estimation correspond à une mesure de l'accessibilité « pure » (au sens où elle est distinguée des problèmes de persévérance des enquêteurs) qui n'est en général pas étudiée, ni estimée.

## 4. Acceptation des ménages

Dans neuf cas sur dix, l'enquêteur parvient à joindre le ménage. La dernière étape du processus de collecte est alors l'acceptation ou non de l'entretien. Cette acceptation est le résultat d'une interaction entre l'enquêteur et le ménage. Il est donc nécessaire d'examiner l'impact des caractéristiques de l'un et l'autre. L'effet enquêteur observé ici se rapproche davantage de l'effet enquêteur classique<sup>15</sup>, car il est souvent involontaire. Comme précédemment, l'identification des déterminants ménage de l'acceptation améliore la précision de l'enquête doublement. D'une part cela permet de déterminer les probabilités de tirage initiales minimisant la dispersion des pondérations finales. D'autre part le calcul de ces pondérations s'en trouve amélioré.

### 4.1. Statistiques descriptives

Une fois le ménage contacté, l'enquêteur tente de le convaincre de réaliser un entretien. Il y parvient dans 84,7% des cas. Les 15,3% restants ne sont pas tous des refus : le ménage se révèle parfois inapte à répondre au questionnaire, ou indique qu'il sera absent pour une longue durée après ce premier contact. Enfin, certains ménages contactés ne donnant plus de nouvelles sont classés en impossibles à joindre. Ce classement ne semble pas satisfaisant dans la mesure où il s'agit souvent d'un refus déguisé.

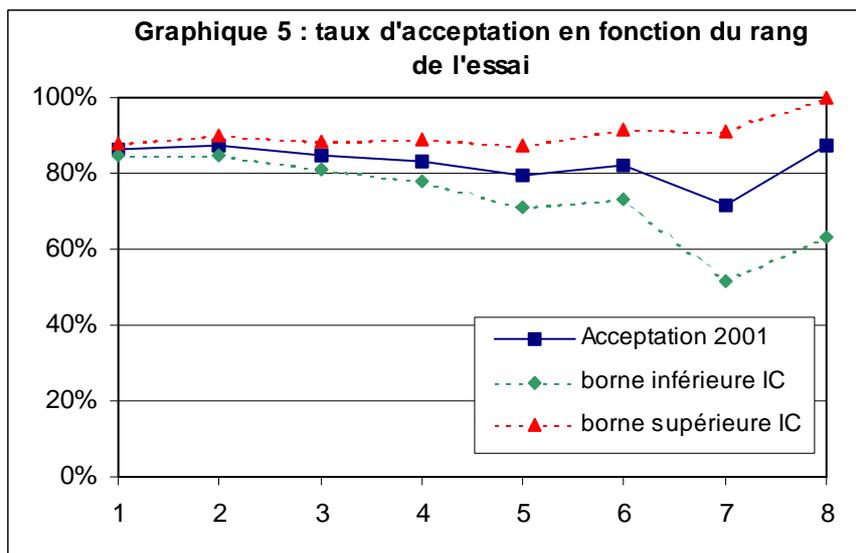
**Tableau 4 : résultats de collecte pour les ménages contactés**

Résultat de collecte	Fréquence
Enquête réalisée	84,7%
Refus	9,7%
Impossible à joindre	2,5%
Absent de longue durée	0,6%
Inapte	2,5%

Source : fiche-adresses de l'enquête PCV d'octobre 2001  
Champ : essais de contact des ménages en 1<sup>ère</sup> interrogation dans le champ de l'enquête

<sup>15</sup> C'est-à-dire son influence sur les réponses au questionnaire.

La persévérance est-elle rentable pour les enquêteurs ? En d'autres termes, les ménages difficiles à contacter sont-ils aussi les plus réticents à répondre ? Le graphique 5 montre l'évolution du taux d'acceptation en fonction du rang de l'essai de contact.



Source : fiche-adresses de l'enquête PCV d'octobre 2001  
 Champ : ménages contactés, dans le champ de l'enquête et en 1<sup>ère</sup> interrogation

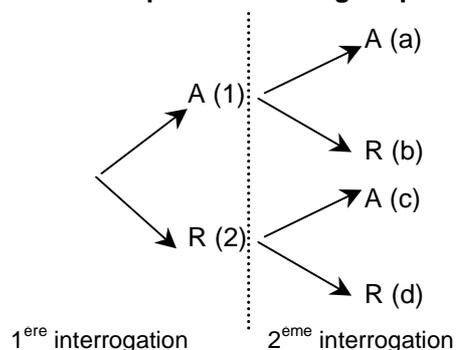
Il n'apparaît pas d'évolution significative du taux d'acceptation en fonction du rang de l'essai, les ménages difficiles à joindre n'étant pas plus réticents à répondre. Ainsi, on supposera par la suite que l'accessibilité et l'acceptation sont indépendantes conditionnellement aux caractéristiques du ménage.

## 4.2. Modélisation

L'acceptation étant le résultat d'une interaction enquêteur-ménage, il est nécessaire d'introduire dans le modèle les caractéristiques de l'un et l'autre. A nouveau, la prise en compte des variables «ménage» pose problème car elles ne sont observées que pour les répondants.

L'idée est alors d'utiliser les deux interrogations de l'enquête PCV pour analyser ces effets. On observe en effet, pour les ménages contactés<sup>16</sup>, quatre cas de figure, comme le montre le schéma ci-dessous (A indiquant une acceptation et R l'absence d'acceptation, un refus par exemple).

**Schéma 1 : les quatre cas de figure pour les ménages contactés**



<sup>16</sup> Le fait de se restreindre aux ménages contactés ne pose pas de problème. En effet, le contact dépend de l'accessibilité du ménage et de la persévérance de l'enquêteur. Or, les caractéristiques du ménage étant fixées, l'accessibilité et l'acceptation sont supposées indépendantes ; de même pour la persévérance et l'acceptation conditionnellement aux caractéristiques de l'enquêteur et aux données inscrites sur la fiche-adresse.

Les variables «ménage» sont inconnues dans le cas (d), mais leur effet peut toutefois s'apprécier en comparant la situation (a) aux situations (b) et (c). Cependant, un problème supplémentaire se pose : le ménage occupant le logement n'est pas nécessairement le même d'une année sur l'autre. Ce phénomène n'est pas négligeable puisque la rotation est d'environ 10%. Pour évaluer correctement l'effet des variables explicatives, il est donc nécessaire de se restreindre aux ménages n'ayant pas changé depuis la première interrogation. Or l'enquêteur n'indique le changement d'occupant que pour les ménages répondants : dans le cas (b), nous ne sommes pas en mesure de savoir si le 2<sup>ème</sup> refus est le fait du ménage qui avait accepté en 1<sup>ère</sup> interrogation ou d'un ménage l'ayant remplacé depuis. A l'inverse, dans le cas (c), l'enquêteur précise en 2<sup>ème</sup> interrogation si le ménage a changé ou non depuis la 1<sup>ère</sup> interrogation. Ne sont donc conservés dans l'analyse que les ménages en (1) et les ménages en (c) déjà présents lors de la première interrogation.

Formellement, on pose :

- $a_{ij1}$  la variable indicatrice d'acceptation à la première interrogation
- $a_{ij2}$  la variable indicatrice d'acceptation à la deuxième interrogation
- $d_i$  la variable indicatrice de changement de logement ( $d_i=1$  si le ménage n'a pas changé)

On a alors :

- $y_{ij} = 1$  si le ménage est en situation (1), c'est-à-dire si  $a_{ij1} = 1$ . 2712 ménages sont dans ce cas de figure.
- $y_{ij} = 0$  si le ménage est en situation (c) et n'a pas changé de logement, c'est-à-dire si  $a_{ij1} = 0$ ,  $a_{ij2} = 1$  et  $d_i = 1$ . 115 ménages sont dans ce cas.

On suppose que  $a_{ij1}$  et  $a_{ij2}$  suivent un modèle logit, et que, conditionnellement à  $X_i$  et  $X_j$ , il y a indépendance entre l'acceptation en première interrogation et en deuxième interrogation<sup>17</sup> :

$$P(a_{ij1} = 1 | X_i, X_j) = q(X_i, X_j) = \frac{1}{1 + e^{-(X_i\delta_1 + X_j\delta_2)}}$$

$$P(a_{ij1} = 0, a_{ij2} = 1 | X_i, X_j) = q(X_i, X_j)(1 - q(X_i, X_j))$$

La variable  $d_i$  est également modélisée par un logit :

$$P(d_i = 1 | X_j) = d(X_j) = \frac{1}{1 + e^{-X_j\gamma}}$$

Enfin, l'acceptation et le changement d'occupant sont supposés indépendants conditionnellement aux caractéristiques du ménage. Par conséquent :

$$P(a_{ij1} = 0, a_{ij2} = 1, d_i = 1 | X_i, X_j) = d(X_j)q(X_i, X_j)(1 - q(X_i, X_j))$$

Le modèle s'écrit finalement :

$$P(y_{ij} = 1 | y_{ij} \in \{0, 1\}, X_i, X_j) = \frac{P(y_{ij} = 1 | X_i, X_j)}{P(y_{ij} \in \{0, 1\} | X_i, X_j)}$$

$$P(y_{ij} = 1 | y_{ij} \in \{0, 1\}, X_i, X_j) = \frac{q(X_i, X_j)}{d(X_j)q(X_i, X_j)(1 - q(X_i, X_j)) + q(X_i, X_j)}$$

C'est-à-dire :

$$P(y_{ij} = 1 | y_{ij} \in \{0, 1\}, X_i, X_j) = \frac{1}{1 + d(X_j)(1 - q(X_i, X_j))} \quad (1)$$

$d(X_j)$ , ici inconnue, est estimée à partir des répondants en 2<sup>ème</sup> interrogation (cas (a) et (c)). Du fait de l'indépendance conditionnelle de  $d_i$  et  $a_{ij2}$ , on a en effet :

$$P(d_i = 1 | X_j, a_{ij2} = 1) = P(d_i = 1 | X_j) = d(X_j) = \frac{1}{1 + e^{-X_j\gamma}}$$

<sup>17</sup> Il serait possible d'améliorer ce modèle en introduisant des effets aléatoires relatifs aux ménages et aux enquêteurs, et ne supposer donc l'indépendance que conditionnellement à ces effets aléatoires.

L'estimation de  $\gamma$  est présentée dans le tableau 5<sup>18</sup> : un coefficient négatif correspond à une plus forte probabilité de changement de l'occupant.

**Tableau 5 : estimation de  $\gamma$**

Variables	Estimation	Ecart-type
<b>Constante</b>	<b>2,44</b>	<b>0,10</b>
<b>Personne de réf. âgée de moins de 30 ans</b>	<b>-1,99</b>	<b>0,16</b>
Personne de réf. âgée de 30 ) 59 ans	Réf.	/
<b>Personne de référence âgée de 60 ans ou plus</b>	<b>0,81</b>	<b>0,19</b>
<b>Revenus inférieurs à 100 000F par an</b>	<b>-0,48</b>	<b>0,14</b>
Autres revenus	Réf.	/

Source : enquêtes PCV d'octobre 2000 et 2001

Champ : répondants à l'enquête 2000, et répondants à l'enquête 2001 ayant refusé en 2000.

Les résultats sont relativement logiques : les jeunes répondants en 2<sup>ème</sup> interrogation avaient une probabilité très forte de ne pas être dans le logement lors de la 1<sup>ère</sup> interrogation, à l'inverse des ménages âgés. Il apparaît également que les ménages à faibles revenus changent plus fréquemment.

Pour estimer  $\delta_1$  et  $\delta_2$ , on remplace ensuite  $d(X_i)$  par  $\hat{d}(X_i)$  dans (1). Le tableau 6 page suivante présente les estimateurs obtenus par maximum de vraisemblance<sup>19</sup>.

**Tableau 6 : estimation de  $\delta_1$  et  $\delta_2$**

Variables	Estimation	Ecart-type
<b>Constante</b>	<b>2,22</b>	<b>0,38</b>
Enquêteur féminin	Réf.	/
Enquêteur masculin	0,02	0,31
Ancien enquêteur (expérience $\geq$ 4 ans)	Réf.	/
Nouvel enquêteur (expérience $<$ 4 ans)	-0,43	0,29
Jeune enquêteur (35 ans ou moins) marié	0,89	1,07
Autre situation	Réf.	/
Pratique d'une autre activité professionnelle	0,18	0,23
Aucune autre activité professionnelle	Réf.	/
<b>Logement en zone rurale</b>	<b>1,02</b>	<b>0,35</b>
Logement en petite zone urbaine	0,46	0,35
Logement en zone urbaine moyenne	0,76	0,39
Logement en grande zone urbaine (hors Paris)	0,59	0,31
Logement en région parisienne	Réf.	/
Ménage d'une personne	Réf.	/
Ménage de 2 personnes	0,41	0,25
Ménage de 3 personnes	0,32	0,33
Ménage de 4 personnes ou plus	0,13	0,41
Personne de réf. âgée de moins de 30 ans	-0,04	0,51
Personne de réf. âgée de 30 à 59 ans	Réf.	/
Personne de réf. âgée de 60 ans ou plus	-0,33	0,25
Revenus inférieurs à 100 000 F	Réf.	/
Revenus compris entre 100 000 F et 180 000 F	0,09	0,28
Revenus supérieurs à 180 000 F	-0,31	0,30
Non-réponse sur la question des revenus	-0,45	0,43
Aucune sortie au cinéma par an	Réf.	/
Entre 1 et 6 sorties au cinéma par an	0,54	0,29
Entre 7 et 12 sorties au cinéma par an	0,36	0,45
<b>Plus de 12 sorties au cinéma par an</b>	<b>0,84</b>	<b>0,36</b>

Source : enquêtes PCV d'octobre 2000 et 2001

Champ : répondants à l'enquête 2000, et répondants à l'enquête 2001 ayant refusé en 2000.

<sup>18</sup> N'ont été conservées dans le modèle que les variables ayant un impact significatif sur d.

<sup>19</sup> L'estimation des écart-types ne prend pas en compte le fait que  $d(X_i)$  soit estimé dans (1).

Ainsi, les caractéristiques de l'enquêteur n'interviennent pas dans l'acceptation. On observe en revanche une hétérogénéité des ménages. En zones rurales, les ménages acceptent plus facilement qu'ailleurs. Et à l'inverse, même si les coefficients correspondant aux petites, moyennes et grandes villes ne sont pas significatifs à 5%<sup>20</sup>, il semble que les Parisiens soient plus réticents à répondre. Par contre, alors que ces variables intervenaient dans l'acceptation, la taille du ménage et l'âge de la personne de référence n'ont pas d'impact sur l'acceptation. De même, et contrairement à une certaine idée reçue, les ménages ayant de hauts revenus ne sont pas plus réticents à répondre. Enfin, la fréquentation assidue du cinéma, qui jouait négativement sur l'accessibilité, est un facteur favorable à l'acceptation : elle pourrait correspondre en ce sens à un indicateur d'extraversion.

Comme dans la partie précédente, l'intérêt de la démarche est, au-delà des résultats obtenus ici, d'expliquer l'acceptation en utilisant des variables obtenues sur les répondants. Cela permet par exemple de corriger la non-réponse « non-ignorable », comme par exemple les cas de refus liés au thème de l'enquête, en introduisant les variables d'intérêt de l'enquête dans les explicatives. La limite de cette modélisation est qu'il est nécessaire de disposer d'enquête avec au moins deux interrogations des ménages. Dans le cas contraire, le recueil, pendant l'enquête, d'informations sur les non-répondants contactés apparaît primordial.

## Conclusion et perspectives

Ce papier propose une modélisation du phénomène de la non-réponse en prenant en compte l'effort que fournit l'enquêteur pour contacter un ménage. La non-réponse peut alors être décomposée en trois phases étudiées successivement. L'étude de la première de ces phases, l'effet enquêteur, nous a permis de montrer que les caractéristiques du ménage n'influencent presque pas sur le comportement des enquêteurs. Ces derniers ne semblent donc pas choisir l'effort qu'ils fournissent par anticipation d'une certaine facilité de contact ou d'une certaine facilité d'acceptation. La seule exception concerne les jeunes, difficilement accessibles, et pour lesquels les enquêteurs sont plus persévérants. En revanche, les effets régionaux semblent assez prononcés et par ailleurs, les enquêteurs peu expérimentés seraient moins persévérants. La deuxième phase étudie plus particulièrement l'accessibilité des ménages. Au-delà des résultats sur leur hétérogénéité, la modélisation choisie a l'avantage de pouvoir inclure comme variables explicatives toutes les variables « ménage » observées dans l'enquête. En effet, grâce à la décomposition du contact en deux parties (persévérance et accessibilité), il est possible d'analyser l'accessibilité des ménages sur les seuls répondants. Enfin, la dernière partie de la non-réponse concerne le refus éventuel des ménages. La modélisation proposée permet là encore d'estimer l'acceptation des ménages à partir de toutes les données individuelles disponibles sur les répondants. Il ressort notamment de cette étude que les ménages habitant en zone rurale acceptent plus facilement. Les deux avantages principaux de la méthode proposée sont les suivants.

Tout d'abord, elle est susceptible d'améliorer les procédures de correction de la non-réponse par repondération. Deux méthodes de repondération sont traditionnellement utilisées. La première consiste à utiliser un calage tel que CALMAR par exemple (méthode de calage sur marges utilisée à l'INSEE). Malheureusement, cela nécessite de connaître les totaux sur toute la population, ce qui limite grandement le choix des variables que l'on peut utiliser pour corriger la non-réponse<sup>21</sup>. La deuxième méthode consiste au contraire à estimer directement la probabilité de réponse et à multiplier les poids de l'enquête par l'inverse de cette probabilité de réponse. Néanmoins, pour estimer ces probabilités, les variables explicatives de la non-réponse doivent, dans les modèles habituels, être disponibles sur les répondants mais aussi les non-répondants : la modélisation est donc en général relativement pauvre. Un des avantages de notre méthode d'estimation qui sépare l'effet enquêteur de l'effet d'accessibilité pure du ménage réside dans le fait que nous sommes capables d'estimer l'accessibilité du ménage en se limitant au sous-échantillon des répondants. On peut alors estimer la probabilité de réponse en utilisant toutes les variables individuelles de l'enquête que l'on désire.

---

<sup>20</sup> Dans les deux derniers cas, ils le sont à 10%.

<sup>21</sup> Ce problème ne se pose pas pour les enquêtes calées sur l'enquête emploi par exemple, puisqu'alors tous les totaux sont calculables. Il n'en reste pas moins vrai que l'enquête sur laquelle on cale, à savoir l'enquête emploi, n'a pas pu subir le même traitement. Par conséquent, il peut rester des biais dans l'enquête emploi qui se propagent aux autres enquêtes.

Le deuxième avantage est de pouvoir améliorer les consignes sur le nombre d'essais de contact que les enquêteurs doivent réaliser. Il sera intéressant d'étudier, en prolongement de ce papier, l'impact d'une règle de l'INSEE différente de la simple règle des 5 essais minimum actuellement en vigueur. Se pose en particulier la question de savoir s'il faut ou non différencier la persévérance en fonction des caractéristiques des ménages, alors que les enquêteurs ne le font pratiquement pas.

## Bibliographie

- [1] Bailar, B., Bailey, L. and Stevens, J. (1977), "Measures of interviewer bias and variance", *Journal of Marketing Research*, 14, 337-343.
- [2] Cheng, S. (1998), "Who are the Reluctant or Rarely-at-home Respondents?", *Survey Methods Centre Newsletter*, 18, pp. 8-11.
- [3] Colombo, R. A. (1995), "An Investigation of Modified Probability Sampling Procedures in Interview Surveys", Master's thesis, The American University, Washington, D. C.
- [4] Deville J.C., et Särndall C.E. (1992), "Calibration estimators in survey sampling", *Journal of the American Statistical Association*, 87, 11, pp. 376-382.
- [5] Drew, J. H. and Fuller, W. A. (1980), "Modelling Nonresponse in Surveys with Callbacks", *Proceedings of Survey Research Method section*, American Statistical Association, pp. 639-642.
- [6] Ellis, R. A., Endo, C. M. and Armer, J. M. (1970), "The Use of Potential Nonrespondents for Studying Nonresponse Bias", *Pacific Sociological Review*, 13, pp. 103-109.
- [7] Fillion, F. L. (1976), "Estimating Bias Due to Nonresponse in Mail Surveys", *Public Opinion Quarterly*, 39, pp. 482-492.
- [8] Fitzgerald, R. and Fuller, L. (1982), "I Hear You Knocking But You Can't Come In : The Effect of Reluctant Respondents and Refusers on Sample Survey Estimates", *Sociological Methods and Research*, 11, pp. 3-32.
- [9] Foster, K. (1997), "The Effect of Call Patterns on Non-Response Bias in Household Surveys", *Survey Methodology Bulletin*, 41, pp. 37-47.
- [10] Fuller, C. H. (1974), "Weighting to Adjust for Survey Nonresponse", *Public Opinion Quarterly*, 38, pp. 239-246.
- [11] Hanson, R.H. and Marks, E. S. (1958), "Influence of the interviewer on the accuracy of survey results", *Journal of American Statistical Association*, 53, pp. 635-655.
- [12] Hox, J. J., de Leeuw, E., and Kreft, G. G. (1991), "The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data : a Multilevel Model", *Measurement Errors in Surveys*, New York : Wiley.
- [13] Lin, I. F. and Schaeffer, N. C. (1995), "Using Survey Participants to Estimate the Impact of Nonparticipation", *Public Opinion Quarterly*, 59, pp. 236-258.
- [14] Lynn, P., Clarke, P., Martin, J. and Sturgis, P. (2002), "The Effects of Extended Interviewer Efforts on Nonresponse Bias", in Groves, M., Dillman, D. A., Eltinge, J. L., Little, R. J. (eds.), *Survey Nonresponse*, New York : Wiley Series, pp. 135-147.
- [15] Pickery, J. and Loosveldt, G. (2000), "Modélisation des effets d'intervieweur dans le cas des enquêtes par panel : une application", *Technique d'Enquête*, 26, pp. 213-224.
- [16] Smith, T. (1984), "Nonattitudes : A Review and Evaluation", in Turner, C. and Martin E. (eds.), *Surveying Subjective Phenomena*, Vol. 2, New York: Russel Sage, pp. 215-255.
- [17] Stinchcombe, A. L., Jones, C. and Sheatsley, P. (1981), "Nonresponse Bias for Attitude Questions", *Public Opinion Quarterly*, 45, pp. 359-375.