

Persévérance, accessibilité et acceptation :  
Une nouvelle modélisation de la non-réponse

*Xavier d'Haultfoeuille et Philippe Février*

*JMS*

*17 décembre 2002*

# Introduction

## Décomposition de la non-réponse

Non-réponse : IAJ, ALD, refus,...

Estimer la non-réponse sans décomposer les effets

1ère décomposition (Lynn et al., 2000)

- Contact du ménage
- Acceptation du ménage

2ème décomposition

- Persévérance de l'enquêteur : nombre d'essais qu'il est prêt à réaliser
- Accessibilité du ménage
- Acceptation du ménage

# Les données

Fichiers des Enquêtes PCV d'octobre 2000 et 2001  
(Permanente sur les Conditions de Vie des ménages)

Fichier de données sur les enquêteurs (sexe, expérience,...)

Un fichier particulier : les données de la fiche adresse

# Les données

## La fiche adresse

Document de quatre pages remis à l'enquêteur

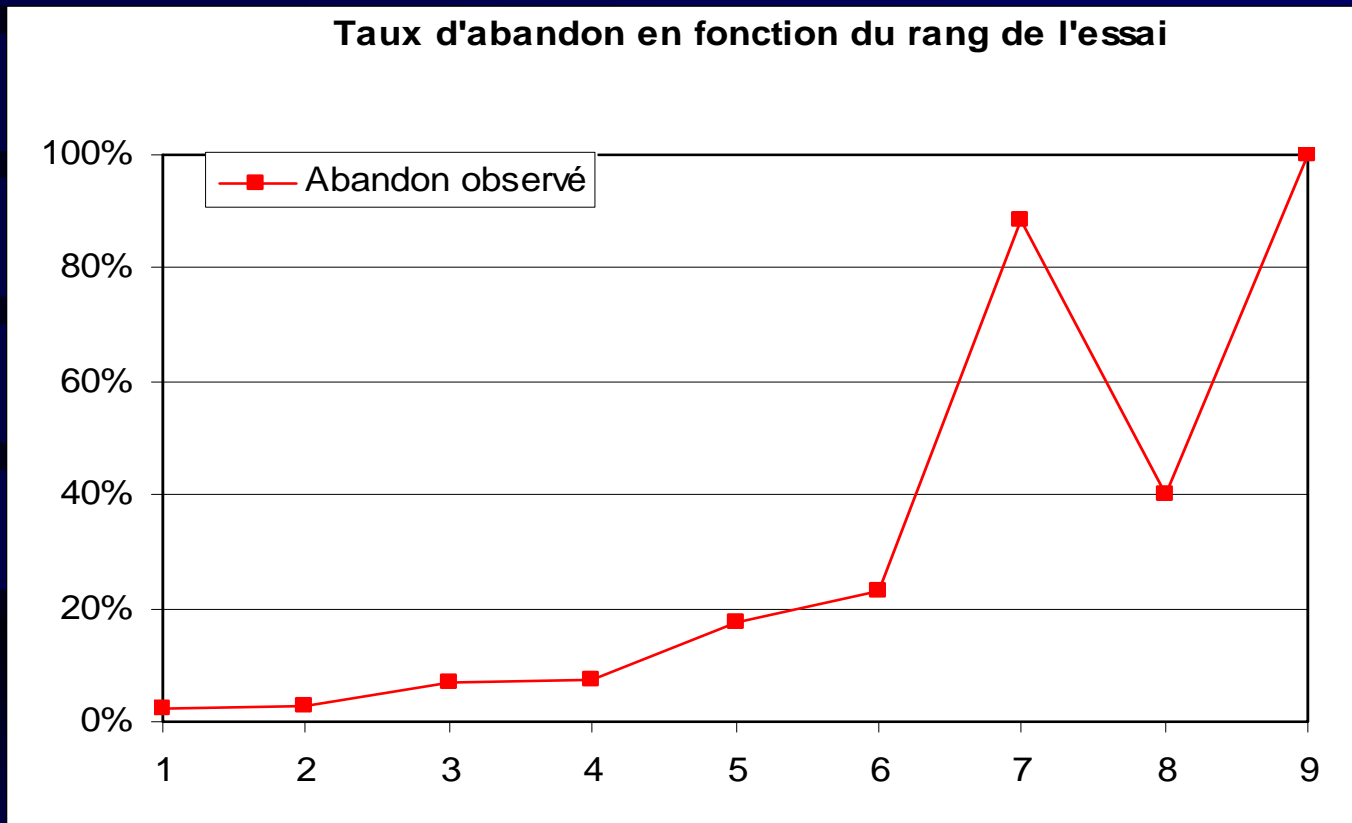
- Variables logement (adresse, collectif, pièces)
- Variables ménage occupant le logement au RP (taille, âge, profession (de mauvaise qualité))

Sur la quatrième page, l'enquêteur doit remplir la chronologie des essais :

- Jour et heure de l'essai
- Moyen utilisé (téléphone ou visite)
- Résultat de l'essai (contact ou pas, acceptation ou refus....)

# La persévérance

## Statistiques descriptives



Les enquêteurs sont prêts à réaliser 5.9 essais en moyenne, étalés au cours de la journée

# La persévérance

## Modélisation de la persévérance

Soit  $\tilde{n}_{ij}$  le nombre d'essais réalisé par un enquêteur  $i$  pour contacter un ménage  $j$  observé dans les données.

$\tilde{n}_{ij}$  correspond à la persévérance  $n_{ij}$  s'il n'y a pas contact:

$$n_{ij} = \tilde{n}_{ij} \text{ si } c_{ij} = 0$$

$\tilde{n}_{ij}$  correspond à une borne inférieure de la persévérance si le ménage est contacté:

$$n_{ij} \geq \tilde{n}_{ij} \text{ si } c_{ij} = 1$$

# La persévérance

## Modélisation de la persévérance

La persévérance de l'enquêteur dépend

- De ses caractéristiques
- Des caractéristiques observées du ménage au RP

Nous modélisons la persévérance par le modèle log-logistique suivant :

$$\ln(n_{ij}^*) = X_i \beta + X_j \gamma + \sigma \varepsilon_{ij} \quad \text{où } \varepsilon_{ij} \text{ suit une loi logistique}$$

$$n_{ij} = E[n_{ij}^*] + 1 \quad \text{si } n_{ij}^* < 8$$

$$n_{ij} = 9 \quad \text{si } n_{ij}^* \geq 8$$

# La persévérance

## Modélisation de la persévérance

On peut alors écrire la vraisemblance du modèle que nous cherchons à maximiser.

Par exemple, si  $c_{ij} = 0$  et  $\tilde{n}_{ij} < 9$ , nous avons :

$$l(\tilde{n}_{ij}, c_{ij}) = P(n_{ij} \geq \tilde{n}_{ij}) = P(n_{ij}^* \geq \tilde{n}_{ij} - 1)$$

soit

$$l(\tilde{n}_{ij}, c_{ij}) = \frac{e^{-\frac{1}{\sigma}[\ln(\tilde{n}_{ij}-1) - X_i\beta - X_j^*\gamma]}}{1 + e^{-\frac{1}{\sigma}[\ln(\tilde{n}_{ij}-1) - X_i\beta - X_j^*\gamma]}}$$



# La persévérance

## Résultats

Tableau 1 : estimation des facteurs de la persévérance des enquêteurs

Variable	Estimation	Ecart-type
<b>Constante</b>	<b>1,89</b>	<b>0,13</b>
Logement neuf	-0,01	0,11
<b>Personne de réf. âgée de moins de 30 ans</b>	<b>0,19</b>	<b>0,07</b>
Personne de référence âgée de 30 à 65 ans	0,00	0,06
Pas d'indication sur l'âge de la personne de réf.	-0,06	0,08
Logement de 2 pièces	-0,10	0,07
Logement de 3 pièces	-0,04	0,07
Logement de 4 pièces	0,01	0,08
Logement de 5 pièces ou plus	-0,01	0,09
Pas d'indication sur le nombre de pièces	-0,16	0,12
Logement en habitat collectif	-0,09	0,05
Logement en grande zone urbaine	-0,03	0,06
Logement en petite zone urbaine	0,00	0,08
Logement en zone rurale	-0,07	0,08
<b>Région Ile de France</b>	<b>-0,21</b>	<b>0,09</b>
Régions du centre	-0,08	0,09
Régions de l'est	-0,03	0,10
Régions de l'ouest	0,06	0,10
Région du sud-ouest	0,01	0,10
Régions Rhône Alpes et Auvergne	0,03	0,10
<b>Régions Paca et Languedoc-Roussillon</b>	<b>-0,24</b>	<b>0,09</b>
Enquêteur masculin	-0,01	0,05
<b>Nouvel enquêteur (expérience &lt; 4 ans)</b>	<b>-0,11</b>	<b>0,05</b>
Pratique d'une autre activité professionnelle	0,00	0,04
<b>Sigma</b>	<b>0,21</b>	<b>0,01</b>



Pas d'effet marquant des Variables ménages sauf Pour les jeunes



Effet régional fort avec Persévérance plus faible Pour Paris et le Sud



Effet de l'expérience : Les jeunes persévèrent moins

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

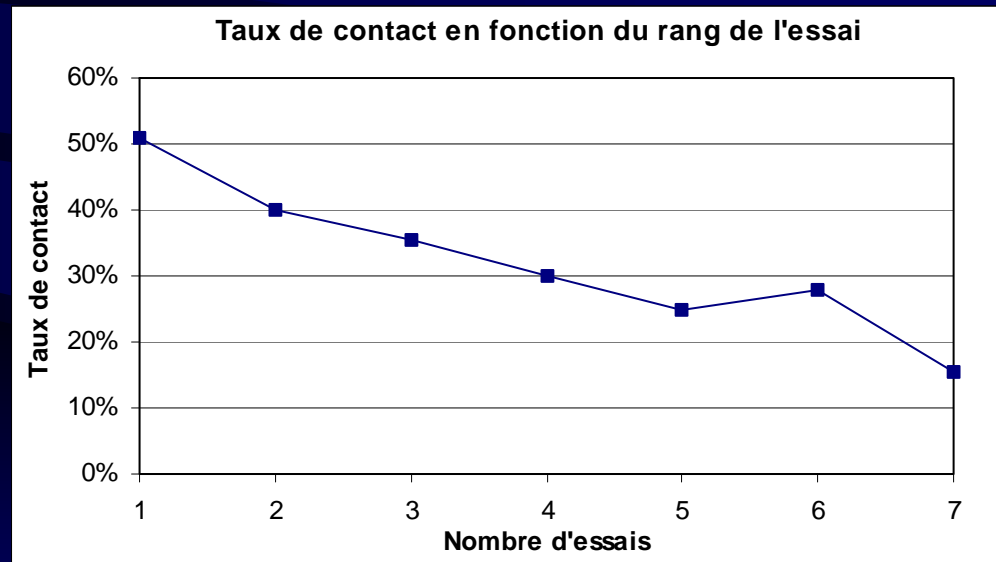
Champ : ménages en 1<sup>ère</sup> interrogation dans le champ de l'enquête

# Analyse de l'accessibilité

Dépend :

- des difficultés d'accès au ménage
- de la présence du ménage au logement

Statistiques descriptives :



Baisse due à l'hétérogénéité des ménages

# Analyse de l'accessibilité modélisation

Déterminants de cette hétérogénéité ?

$z_{ij}$ , accessibilité, est analysée à partir de  $\tilde{n}_{ij}$  nombre d'essais réalisés :

$$\begin{aligned} z_{ij} &= \tilde{n}_{ij} \text{ si contact} \\ z_{ij} &> \tilde{n}_{ij} \text{ si non contact} \end{aligned}$$

Problème : observation des caractéristiques uniquement sur les répondants  $\Rightarrow$  on conditionne par le contact et l'acceptation.

Vraisemblance conditionnelle :

$$P(z_{ij} = k | X_j, X_i, X_j^*, z_{ij} \leq n_{ij}, a_{ij} = 1)$$

# Analyse de l'accessibilité modélisation

Hypothèses :

- indépendance conditionnelle entre l'accessibilité et la persévérance et entre l'accessibilité et l'acceptation
- $z_{ij}$  suit une loi géométrique de paramètre  $p(X_j)$

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

Les  $P(k \leq n_{ij} | X_i, X_j^*)$  ne sont pas connues

=> On les remplace par leurs estimations obtenues à partir du modèle sur la persévérance

# Analyse de l'accessibilité modélisation

## Estimation des déterminants de l'accessibilité des ménages :

Variables	Estimation	Ecart-type
<b>Constante</b>	<b>-0,24</b>	<b>0,12</b>
Ménage de 2 personnes	0,09	0,08
<b>Ménage de 3 personnes</b>	<b>0,30</b>	<b>0,10</b>
<b>Ménage de 4 personnes ou plus</b>	<b>0,72</b>	<b>0,13</b>
<b>Logement en zone rurale</b>	<b>0,31</b>	<b>0,11</b>
<b>Logement en petite zone urbaine</b>	<b>0,29</b>	<b>0,12</b>
Logement individuel en zone urbaine moyenne	0,07	0,15
Logement collectif en zone urbaine moyenne	0,08	0,16
Logement en grande zone urbaine (sauf Paris)	0,10	0,11
<b>Personne de réf. âgée de moins de 30 ans</b>	<b>-0,40</b>	<b>0,12</b>
<b>Personne de réf. âgée de 60 ans ou plus</b>	<b>0,40</b>	<b>0,08</b>
Entre 1 et 6 sorties au cinéma par an	-0,09	0,08
Entre 7 et 12 sorties au cinéma par an	-0,15	0,14
<b>Plus de 12 sorties au cinéma par an</b>	<b>-0,26</b>	<b>0,10</b>



Effet fort de la  
taille du ménage

Facilité d'accès aux  
logements en zone  
rurale et petite urbaine

Les personnes âgées  
sont plus présentes ;  
les jeunes moins

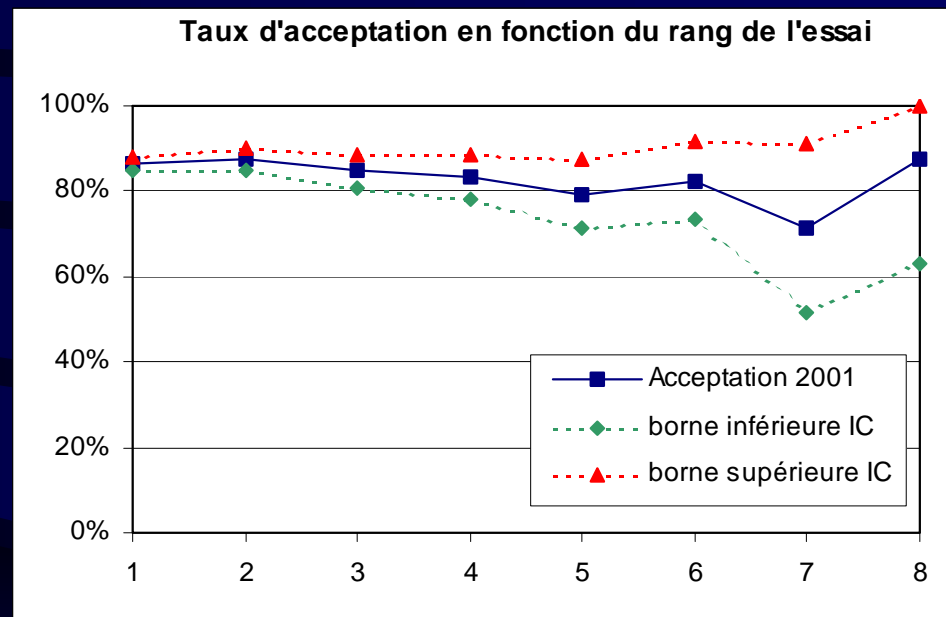
Effet de l'extraversion  
(présence plus rare au  
logement)

Source : enquête PCV d'octobre 2001

Champ : ménages répondants en 1<sup>ère</sup> interrogation

# Analyse de l'acceptation statistiques descriptives

Ménages contactés 9 fois sur 10, mais encore 15% de non-réponse après le contact (refus, mais aussi inaptes, IAJ)



Pas d'évolution significative du taux d'acceptation en fonction du rang de l'essai

=> Hypothèse d'indépendance conditionnelle entre accessibilité et acceptation

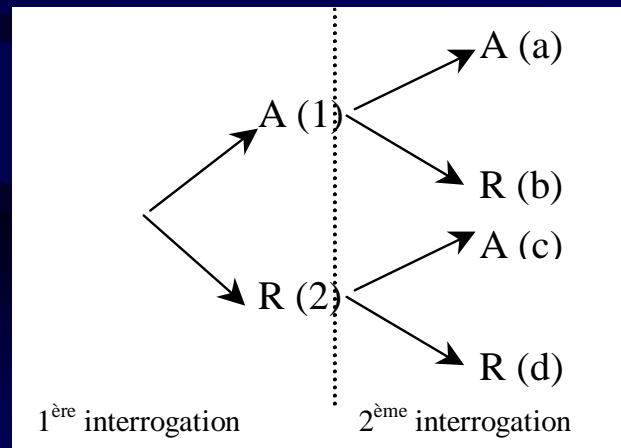
# Analyse de l'acceptation

## Modélisation

Effet des caractéristiques du ménage et de l'enquêteur ?

Même difficulté que précédemment : les caractéristiques ne sont observées qu'en cas d'acceptation

Utilisation des deux interrogations de PCV :



L'effet des caractéristiques peut être estimé en comparant (a) à (b) et (c)

# Analyse de l'acceptation

## Modélisation

Deuxième difficulté : les ménages peuvent changer de logement d'une année sur l'autre (on le sait uniquement pour les répondants en 2<sup>ème</sup> interrogation)

⇒ On ne peut pas savoir si le refus en (b) provient du même ménage ou d'un ménage l'ayant remplacé

Finalement, on compare les situations (1) ( $y_{ij}=1$ ) et (c) ( $y_{ij}=0$ ).

Vraisemblance conditionnelle :

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



# Analyse de l'acceptation

## Modélisation

Hypothèses d'indépendance conditionnelle :

- entre la première et la deuxième décision d'acceptation
- Entre l'acceptation et le changement de logement

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

$d(X_j)$  : probabilité de ne pas avoir changé depuis la 1<sup>ère</sup> interrogation

$q(X_i, X_j)$  : probabilité d'accepter l'entretien

$d(X_j)$ , inconnue, est estimée puis remplacée dans (1) par son estimation

# Analyse de l'acceptation

## Modélisation

### Résultats :

Estimation des déterminants du refus

Variables	Estimation	Ecart-type
<b>Constante</b>	<b>2,44</b>	<b>0,31</b>
Enquêteur homme	-0,01	0,30
Nouvel enquêteur (expérience < 4 ans)	-0,29	0,27
Pratique d'une autre activité professionnelle	0,24	0,23
<b>Logement en zone rurale</b>	<b>1,01</b>	<b>0,34</b>
Logement en petite zone urbaine	0,38	0,33
<b>Logement en zone urbaine moyenne</b>	<b>0,80</b>	<b>0,37</b>
<b>Logement en grande zone urbaine (sauf Paris)</b>	<b>0,69</b>	<b>0,31</b>
Personne de réf. âgée de moins de 30 ans	0,18	0,50
Personne de référence âgée de 60 ans ou plus	-0,31	0,23
<b>Ménage d'une personne</b>	<b>-0,49</b>	<b>0,23</b>
Entre 1 et 6 sorties au cinéma par an	0,40	0,28
Entre 7 et 12 sorties au cinéma par an	0,30	0,44
<b>Plus de 12 sorties au cinéma par an</b>	<b>0,79</b>	<b>0,36</b>

Pas d'effet des caractéristiques de l'enquêteur

Refus plus fort à Paris et dans les petite villes

Personnes seules plus réticentes

« Extraversion » facteur positif de l'acceptation ?

Source : enquêtes PCV d'octobre 2000 et 2001

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

# Conclusion et perspectives

La modélisation permet d'utiliser l'ensemble des données disponibles sur les répondants. D'où :

- Une analyse fine de la non-réponse
- Une meilleure repondération, permettant de tenir compte de la non-réponse « non-ignorable »

La persévérance des enquêteurs n'est pas différenciée suivant le type de ménages :

=> détermination d'une consigne « optimale » sur le nombre d'essais à réaliser pour accentuer les efforts sur les ménages difficiles à joindre

Limites :

- Qualité des données (exhaustivité des essais ?)
- Spécification des modèles (absence d'effets fixes)