



L'impact du mode d'enquête sur la mesure des comportements de mobilité

Caroline Bayart, Patrick Bonnel

► To cite this version:

Caroline Bayart, Patrick Bonnel. L'impact du mode d'enquête sur la mesure des comportements de mobilité. *Economie et Statistique / Economics and Statistics*, 2010, 437, pp.47-50. halshs-01053167

HAL Id: halshs-01053167

<https://shs.hal.science/halshs-01053167>

Submitted on 20 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'impact du mode d'enquête sur la mesure des comportements de mobilité

Caroline Bayart* et Patrick Bonnel**

Les enquêtes de mobilité sont indispensables, tant à la compréhension, à l'analyse et à la modélisation des comportements de déplacement, que pour contribuer à définir et à évaluer les politiques de transport mises en œuvre. Pourtant, les taux de réponse aux *Enquêtes Ménages Déplacements (EMD)* décroissent au fil du temps.

Ces dernières années, les méthodes utilisées pour recueillir les données de mobilité ont évolué, afin d'une part de nourrir des modèles de plus en plus complexes et d'autre part d'intégrer les apports des nouvelles technologies dans les protocoles d'enquête (web, GPS, etc.). En particulier, la combinaison de différents médias s'impose comme un moyen d'améliorer la qualité des données produites à moindre coût, en permettant une augmentation du taux de réponse global. Si le fait de proposer des médias différents permet d'augmenter le taux de réponse global, la comparabilité des données reste un exercice difficile. Les caractéristiques socioéconomiques des répondants varient selon le mode d'enquête, et ces différences peuvent être à l'origine de comportements spécifiques.

Nous montrons dans cet article qu'il est possible de distinguer l'effet dû au mode d'enquête de celui lié aux différences socioéconomiques observées entre les échantillons de répondants, et de quantifier l'impact du mode d'enquête sur la mesure de la mobilité déclarée. Le modèle économétrique envisagé pour cette analyse est emprunté au domaine des variables qualitatives. Plus précisément, il s'agit d'un modèle de sélection de l'échantillon, dont nous estimons les paramètres à l'aide de la procédure en deux étapes, élaborée par Heckman à la fin des années 1970. Les résultats de l'*EMD* menée à Lyon en 2006 en face à face et par internet illustrent nos propos. L'objectif est de quantifier, pour chaque répondant, l'impact du mode d'enquête sur la mesure du comportement de mobilité.

* Université de Lyon, Université de Lyon 1, ISFA, Laboratoire de Sciences Actuarielle et Financière – Article rédigé au Laboratoire d'Économie des Transports, ENTPE, Université Lumière Lyon 2, CNRS.

** Laboratoire d'Économie des Transports, ENTPE, Université Lumière Lyon 2, CNRS.

La difficulté croissante pour obtenir des données d'enquêtes représentatives de la population visée et la complexité des informations nécessaires à l'alimentation de modèles de plus en plus sophistiqués ne permettent généralement plus de recueillir toutes les données au cours d'une même enquête ou selon une méthodologie unique. De plus en plus d'enquêtes reposent sur des protocoles complexes, qui associent plusieurs modes ou méthodologies. L'objectif est d'améliorer la qualité des données produites en augmentant le taux de couverture de la population cible et le taux de réponse global tout en réduisant les coûts dans le cas du média web (Couper, 2000 ; Gunn, 2002 ; Dillman et al., 2001). Mais proposer plusieurs modes ou méthodes de recueil de données n'est pas sans risque, le recueil d'informations via différentes sources pouvant générer des résultats parfois peu comparables. Le danger, lorsque l'on fusionne des bases de données, est de générer un biais de sélection des individus qui compromet la pertinence des modèles explicatifs des comportements de mobilité.

Ce biais de sélection est l'objet d'une importante littérature, théorique et empirique (Winship et Mare, 1992), mais les applications aux enquêtes transport sont rares à ce jour. Des travaux montrent, depuis les années 1950, que l'estimation d'une équation sur un sous-échantillon obtenu de façon sélective dans la population peut conduire à des biais (Roy, 1951). Les premiers développements économétriques des conséquences de cette sélection des individus datent de 1974, avec les travaux d'Heckman (1979). L'exemple souvent cité dans la littérature est celui d'une équation de salaire estimée sur les seules femmes actives, alors même que le comportement d'activité relève d'un arbitrage dans lequel le salaire que la personne peut obtenir sur le marché intervient. Depuis, de nombreux articles ont mis en évidence l'importance du biais de sélection dans les enquêtes réalisées en sciences humaines et sociales (Maddala, 1986). Le laboratoire d'Économie des Transports a proposé de réaliser une enquête par le web auprès des non-répondants à l'*Enquête Ménages Déplacements (EMD)* réalisée à Lyon en face à face en 2006 (Bayart et Bonnel, 2008), c'est-à-dire auprès des individus qui ont refusé de recevoir un enquêteur à leur domicile ou qu'il n'a pas été possible de joindre durant la première vague d'interviews. Les données de cette enquête nous permettent de mettre en évidence le problème d'auto-sélection, les non-répondants à l'enquête standard en face à face

choisissant de remplir ou pas le questionnaire sur le web. L'objet de cet article est de tester l'incidence du mode d'enquête sur le comportement déclaré de mobilité des répondants, et de s'efforcer ensuite de le mesurer.

Une méthodologie innovante

La méthodologie de l'*EMD* (cf. encadré 1) est définie par le Certu (1) (Certu, 2008). À Lyon, l'enquête est traditionnellement menée en face à face.

Les personnes refusant de répondre à l'enquête standard, impossibles à joindre malgré plusieurs tentatives à des horaires variables, ou absentes de longue durée en 2006, ont été contactées par courrier pour remplir le questionnaire en ligne, car nous ne disposions pas de leur adresse électronique et pas toujours de leur numéro de téléphone (cf. graphique et encadré 2). Le périmètre de l'enquête web correspond aux 72 communes du Schéma de COhérence Territorial (Scot) de Lyon.

Au cours de la période d'enquête (de novembre 2005 à avril 2006), 11 666 ménages ont été contactés mais seuls 6 375 ont accepté de recevoir un enquêteur à leur domicile sur le territoire du Scot de Lyon, déclarant 48 143 déplacements réalisés la veille du jour de l'enquête. Le taux de réponse global de l'enquête en face à face est donc égal à 53 % sur ce périmètre. Parmi les non répondants à l'enquête en face à face 4 335 courriers ont été envoyés aux répondants potentiels sur le web, en deux vagues successives, chacune avec deux relances. Au final, 536 individus ont accédé au site internet du Certu pour tenter de répondre à l'enquête, ce qui représente un taux de connexion de 12,4 %. Ce chiffre est satisfaisant, puisque les ménages concernés par l'enquête web sont ceux qu'il n'avait pas été possible d'atteindre durant la première phase de l'enquête, malgré huit tentatives, ou qui avaient refusé de répondre. Cependant, tous les ménages qui se sont connectés n'ont pas terminé la saisie. Étant donné la longueur du questionnaire et son caractère auto-administré, certains individus n'ont pas pu (contraintes techniques : modem bas-débit, navigateur inadéquat, etc.) ou pas voulu (questions parfois très personnelles, demandant un important effort de mémoire, etc.) répondre entièrement à l'en-

1. Certu : Centre d'Études sur les Réseaux, les Transports et l'Urbanisme.

Encadré 1

LES ENQUÊTES MÉNAGES DÉPLACEMENTS

L'objet des *Enquêtes Ménages Déplacements (EMD)* est de recueillir les pratiques de déplacements d'une population. Ces enquêtes, conçues dans un but de comptabilisation et de compréhension des déplacements dans le temps et dans l'espace, constituent un outil privilégié de collecte de données sur la mobilité urbaine. Mises en place depuis les années 1970, les informations qu'elles fournissent permettent d'élaborer des modèles de prévision, indispensables pour évaluer les effets de choix en matière de politique de déplacements ou d'infrastructures de transport. Elles permettent ensuite de mesurer les conséquences de ces choix. Les *EMD* constituent une base de données au service de la collectivité et chaque agglomération peut se situer par rapport aux autres. L'agglomération de Grenoble est celle qui connaît la plus grande périodicité des *EMD* (entre 7 à 10 ans), mais dans la plupart des agglomérations la périodicité de ces enquêtes est de 10 à 12 ans. Et elles sont parfois moins fréquentes.

Le maître d'ouvrage est l'autorité organisatrice, le Syndicat Mixte des Transports pour le Rhône et l'Agglomération Lyonnaise (Sytral) à Lyon. Le Certu a un rôle de contrôle du respect de la méthodologie « standard Certu », qui conditionne le versement de la subvention de l'État. Les Cete (Centre d'Études Techniques de l'Équipement) ont souvent un rôle d'assistant à maître d'ouvrage. Coûteuses, ces enquêtes bénéficient d'un financement mixte (État, Région, Département, autorité organisatrice des transports en commun).

L'enquête permet de collecter tous les déplacements effectués la veille du jour d'enquête par les individus du ménage. Dans les grandes agglomérations, la méthodologie Certu impose un questionnaire en face à face de tous les membres du ménage. Le questionnaire, relativement long, conduit à une durée moyenne d'interrogation de 1 heure 30 minutes.

Encadré 2

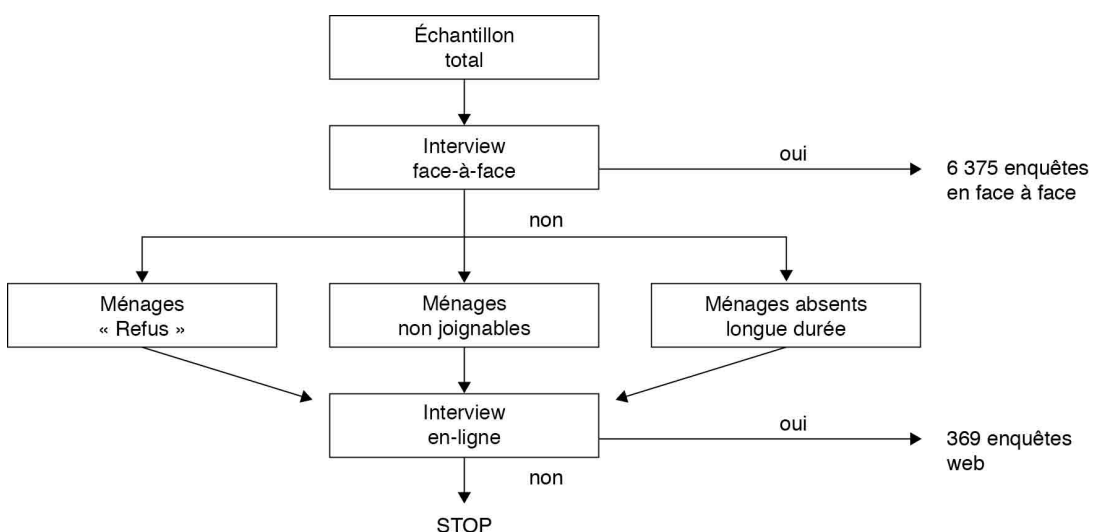
LE RECRUTEMENT DES MÉNAGES DE L'ÉCHANTILLON

Jusqu'en 2008, l'Insee n'était pas autorisé à diffuser à l'extérieur de l'Institut des échantillons issus du recensement (depuis, cette règle a été modifiée avec l'adoption de la loi sur les archives, qui permet, dans des conditions strictes de sécurité, la communication de ces données à des fins de statistique publique, de recherche scientifique ou historique). L'*Enquête Ménages Déplacements* de Lyon n'étant pas réalisée par l'Insee en 2006, le tirage d'un pré-échantillon a été effectué sur la base du fichier des abonnés de France Télécom. À partir du fichier d'adresses obtenu, les enquêteurs repèrent d'autres adresses sur le terrain, en suivant une règle précise, celle des itinéraires. Cette méthode, couramment utilisée pour obtenir des

échantillons de ménages ou de logements, consiste à imposer à l'enquêteur un itinéraire en lui indiquant exactement les points du circuit où il doit procéder à une interview. À une adresse donnée correspond une résidence principale. L'échantillon ainsi constitué est relativement récent et minimise le taux de rebut (logements vacants, logements détruits ou ayant perdu leur usage d'habitation, résidences secondaires, etc.). Cette méthode permet la constitution d'un échantillon aléatoire représentatif de l'ensemble de la population du territoire étudié. Le recours à l'annuaire inversé permet ensuite d'obtenir le numéro de téléphone des logements sélectionnés qui sont listés dans l'annuaire.

Graphique

Schéma de recrutement des ménages



Lecture : les personnes refusant de recevoir un enquêteur à domicile, non joignables ou absentes de longue durée sont incitées à remplir le questionnaire web.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

quête. Seuls 369 individus, déclarant un total de 1 108 déplacements, ont fourni une réponse suffisamment complète et exploitable, soit un taux de réponse à l'enquête web de 8,5 % (cf. tableau 1).

La perception des questions et l'organisation du questionnaire dépendent du potentiel du média utilisé (Dillman *et al.*, 1998). Le questionnaire utilisé pour l'enquête en face à face a donc été simplifié, pour l'adapter à l'enquête web, de nature auto-administrée. Si la définition du terme déplacement est précise : « le mouvement d'une personne, effectué pour un certain motif, sur une voie publique, entre une origine et une destination, selon une heure de départ et une heure d'arrivée, à l'aide d'un ou plusieurs moyens de transport » (Cete de Lyon, 2005), sa traduction au sein du questionnaire et surtout sa compréhension par les enquêtés le sont beaucoup moins et nécessitent une attention particulière, notamment dans un questionnaire auto-administré.

L'utilisation conjointe de plusieurs media offre une situation plus claire du point de vue méthodologique (cf. : *enquête Emploi* de l'Insee ; Givord, 2003). Pour réaliser l'enquête web, nous avons adapté le questionnaire papier, traditionnellement utilisé dans les enquêtes ménages en face à face, en tenant compte de plusieurs contraintes. D'une part, des contraintes méthodologiques, imposées par le Certu (2008) pour toute réalisation d'enquêtes ménages déplacements (thèmes, vocabulaire, ordre et formulation des questions, etc.). D'autre part, des contraintes techniques, imposées par le cahier des charges de l'enquête de Lyon. Le web étant un module additionnel à l'enquête en face à face, nous n'avons pu choisir ni le prestataire en charge de l'administration de l'étude en ligne, ni le logiciel utilisé pour le développement du questionnaire. Par ailleurs, une seule personne du ménage remplit le questionnaire web, alors qu'en face à face tous les membres du ménage sont interrogés.

Tableau 1
Taux de réponse par mode d'enquête

	Enquête Face à face	Enquête Web
Ménages contactés	11 666	4 335
Ménages interviewés	6 375	536
dont interviews complètes	6 375	369

Lecture : 11 666 ménages ont été contactés pour répondre à l'enquête en face à face.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

Les répondants web, un profil particulier

Nous ne connaissons pas *a priori* la cible des répondants à l'enquête web, puisqu'il s'agit de ménages qu'il n'a pas été possible d'enquêter en face à face. Nous allons d'abord qualifier les deux échantillons sur la base de caractéristiques socioéconomiques, avant de nous intéresser aux données de mobilité. Pour rendre la présentation plus lisible, nous raisonnons d'une part sur les données relatives au ménage : localisation, nombre de personnes, revenu, équipement en moyens de communication, etc., et d'autre part sur celles concernant les personnes : sexe, âge, profession et catégorie socioprofessionnelles (PCS), statut, niveau d'études, etc.. Suivant les variables étudiées, nous utilisons le test d'égalité des moyennes ou des proportions. Les conclusions se limitent aux résultats dont le niveau de significativité statistique est supérieur à 5 %.

Des ménages à haut niveau de revenus, adeptes des nouvelles technologies

Au niveau des ménages, trois types de différences sont observées : sociodémographiques, d'accès à la voiture particulière et d'équipement en moyens de communication (cf. tableaux 2 et 3).

Les ménages internautes ont davantage de membres (2,60 contre 2,38 dans l'enquête en face à face). Cette différence peut être liée au statut des répondants, puisque nous verrons que l'échantillon web regroupe majoritairement des actifs. L'étude de la distribution de la taille des ménages montre que leur composition diffère entre les deux modes : pour l'enquête web, les ménages sont davantage des familles et plus rarement des personnes isolées (position dans le cycle de vie qui coïncide avec la présence d'enfants). Les ménages internautes ont également un revenu annuel moyen supérieur. Les différences de revenus peuvent s'expliquer par la taille des ménages. Afin de neutraliser l'effet de composition du ménage, nous avons créé une variable qui tient compte du nombre de personnes, et calculé un revenu moyen par unité de consommation, selon les règles établies par l'Insee (Hourriez et Olier, 1997) (2). Les ménages ayant répondu à l'enquête web ont davantage de revenus annuels nets disponibles par unité de consommation que ceux ayant répondu en

face à face (20 000 euros contre 15 000 euros en moyenne). Cette différence s'explique par le fait que la pénétration d'internet est encore inégale sur le territoire, et concerne davantage les ménages ayant des hauts revenus (coût de l'investissement dans du matériel informatique, abonnement internet, niveau d'étude supérieur, etc.), cf. Visser et Lanzerdorf, 2004 ; Wang et Law, 2007. Par ailleurs, ceux qui disposent d'une connexion sur le lieu de travail ont des professions généralement supérieures, avec un bon niveau de revenu. De plus, la part des ménages refusant de donner leur niveau de revenu annuel est beaucoup moins importante sur le web (1 ménage sur 4, contre 1 ménage sur 3). En général, les individus sont réticents à communiquer des informations personnelles, surtout lorsqu'il s'agit de leurs revenus. Il est plus difficile de déclarer ses revenus à un enquêteur en face à face, que sur le web, qui est par définition auto-administré (Potoglou et Karanoglou, 2008). Il se peut également que les ménages réticents à fournir des données personnelles n'aient pas répondu à l'enquête web. Ceux qui ont accepté de remplir le questionnaire

en ligne sont donc moins réticents à répondre aux questions concernant leurs revenus.

Les ménages internautes sont davantage motorisés. Cette sur-motorisation persiste, si l'on considère le nombre moyen de véhicules par individu en âge de conduire (plus de 18 ans). Elle s'explique par une proportion de ménages dépourvus de voiture beaucoup plus faible parmi les répondants web (9 %, contre 21 %), et par une proportion plus importante de ménages possédant entre 0,5 et 1 voiture par personne de 18 ans et plus (65 %, contre 49 %). L'accès à la voiture est donc plus important au sein de la population web que pour celle de l'enquête face à face. Ceci est en partie lié à la taille du ménage et aux caractéristiques socioéconomiques des répondants, notamment en termes de revenus. Ce résultat perdure même lorsque l'on se limite à la population des ménages comportant au moins un actif.

2. Le chef de ménage compte pour une unité de consommation, les autres personnes du ménage de 14 ans et plus comptent pour 0,5 unité de consommation, et les autres personnes du ménage de moins de 14 ans comptent pour 0,3 unité de consommation.

Tableau 2
Taille et revenu des ménages par mode d'enquête

	Enquête Face à face	Enquête Web
Taille du ménage (nombre de personnes)		
Moyenne	2,38	2,60
Écart-type	1,364	1,313
Revenus annuels nets par unité de consommation (euros)		
Moyenne	15 051	20 107
Écart-type	9 651	11 019

Lecture : les ménages interrogés en face à face contiennent en moyenne 2,38 membres.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

Tableau 3
Caractéristiques sociodémographiques des ménages par mode d'enquête

En %

	Enquête Face à face	Enquête Web
Accepte de déclarer son revenu		
Oui	67	76
Non	33	24
Connexion Internet au domicile		
Oui	51	80
Non	49	20
Téléphone portable		
Oui	69	80
Non	31	20
Inscription annuaire France Telecom		
Oui	78	58,5
Non	22	41,5

Lecture : en moyenne, 67 % des ménages interrogés en face à face déclarent leurs revenus annuels nets.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

Les ménages ayant répondu sur le web sont mieux équipés en connexion internet que ceux de l'enquête standard (80 % ont une connexion internet au domicile, contre 51 %). Cette différence est évidemment directement liée au média utilisé pour l'enquête web. Il est intéressant de noter que les 20 % des ménages ayant répondu sur le web qui ne se sont pas connectés à internet à domicile ont donc dû répondre sur leur lieu de travail. La proportion de personnes possédant un téléphone portable, pour motif personnel ou professionnel, est plus importante chez les répondants sur le web (80 %, contre 69 %). Cet écart est probablement lié à la PCS, certaines catégories possédant davantage de téléphones pour motif professionnel, et au niveau de revenu (Bigot et Crouette, 2007). Enfin, si la proportion de ménages possédant une ligne de téléphone fixe à domicile est proche entre les deux enquêtes (90 %), les ménages internautes sont moins souvent sur l'annuaire (seuls 58,5 % sont inscrits sur l'annuaire France Télécom, contre 78 %). Des packages de téléphonie haut débit, avec des numéros spéciaux, sont accessibles aux internautes, ce qui explique en partie leur absence de l'annuaire France Telecom et la plus grande difficulté à les joindre. Le web est alors une opportunité pour eux de répondre à l'enquête sans avoir à recevoir un enquêteur à domicile.

Des individus actifs qui possèdent un bon niveau de formation

Au niveau des individus, les différences socioéconomiques observées (cf. tableau 4) permettent d'affiner certaines hypothèses. Les femmes sont un peu plus présentes dans l'échantillon web, même si la différence reste assez faible (57 % contre 52 %). Ce constat rejoint les résultats observés couramment lorsqu'il n'y a pas de sélection aléatoire des répondants au sein du ménage. Les actifs sont très fortement surreprésentés (71 % contre 46 % en face à face). Ils disposent de peu de temps pour répondre aux enquêteurs à domicile. En revanche, ils sont davantage familiers avec les technologies de la communication, et disposent souvent d'un accès internet haut-débit sur leur lieu de travail. Les personnes dont l'âge est compris entre 30 et 59 ans constituent l'essentiel de l'échantillon des répondants web (70 % contre 45 % en face à face). Les deux variables âge et occupation sont liées : les jeunes (scolaires et étudiants), tout comme les plus âgés (retraités) sont largement sous représentés. Ce résultat découle très probablement de la non-sélection du répondant

au sein du ménage puisque les destinataires du courrier informant de la possibilité de répondre à l'enquête en ligne sont le chef de ménage et son conjoint. Par ailleurs, les personnes âgées sont généralement moins familières avec internet et préfèrent répondre en face à face.

Deux catégories socio professionnelles sont sur-représentées dans l'enquête web : les cadres ou professions intellectuelles supérieures et les employés (environ 39 % chacun), qui ne totalisent respectivement que 14 % et 23 % des effectifs dans l'enquête en face à face. Les cadres et les employés ont essentiellement des fonctions administratives, qui nécessitent une certaine connaissance d'internet et un accès web au travail, ce qui favorise certainement une connexion web à domicile (Bates, 2001). De plus, leur type d'emploi se situe davantage dans le centre de l'agglomération. Les internautes ont un niveau d'études nettement supérieur (62 % ont fait des études après le baccalauréat contre 30 % en face à face). Les internautes possèdent davantage le permis de conduire. Ceci s'explique par l'âge des répondants. Les personnes jeunes ne disposent pas toujours du permis de conduire (contrainte financière) et les personnes appartenant aux classes d'âge les plus élevées n'ont pas toutes passé leur permis, surtout les femmes (Pochet, 2003). Ces résultats

Tableau 4
Caractéristiques sociodémographiques des personnes par mode d'enquête

	Enquête Face à face	Enquête Web
Sexe		
Homme	48	43
Femme	52	57
Permis de conduire		
Oui	81	93
Non	19	7
Âge		
< 30 ans	30	17
De 30 à 59 ans	46	70,5
> 59 ans	24	12,5
Occupation		
Actifs	46	71
Étudiants-Scolaires	20	7
Chômeurs	5	4
Inactifs	5	5
Retraités	24	13
Niveau d'études		
En cours	20	7
Supérieur	30	62
Non supérieur	50	31

Lecture : les hommes représentent 48 % des personnes interrogées en face à face.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

tats confirment ceux de Srinivasan et Athuru (2004), qui montrent que les individus ayant le permis de conduire sont davantage utilisateurs d'internet.

La comparaison des données de l'EMD face à face avec celles de l'enquête web nous permet d'obtenir quelques précisions sur les critères sociodémographiques des individus ayant répondu à l'enquête web. S'agissant majoritairement d'actifs, cette population est moins disponible pour répondre à un long questionnaire en face à face, même si l'analyse des données de l'enquête en face à face comparativement aux données du recensement ne met pas en évidence une sous-représentation de ces actifs. Probablement plus sensibles aux enjeux de l'étude et familiers avec internet, ils ont accepté de se connecter pour répondre au questionnaire en ligne. Les faibles effectifs recensés dans les classes d'âge extrêmes (moins de 19 ans et plus de 74 ans) sont un biais important en termes de comportements de mobilité pour la population ayant répondu à l'enquête web. La localisation des internautes, en revanche, ne diffère pas sensiblement de celle de la population face à face.

Une mobilité globalement plus faible pour les internautes

Il est probable que les différences socioéconomiques constatées entre les deux populations web et face à face impactent les habitudes de déplacement des répondants (Madre *et al.*, 2007 ; Baudelle *et al.*, 2003). Le nombre de déplacements déclarés par les internautes est plus faible : 3,00 contre 3,63 en face à face (cf. tableau 5). Cette différence s'explique tout d'abord par une immobilité (c'est-à-dire une absence de déplacement pendant la période de

référence) nettement plus importante dans l'enquête web (19 % contre 11 % en face à face). Il est en effet plus facile de se déclarer immobile pour raccourcir l'enquête dans une enquête auto-administrée qu'en présence d'un enquêteur. Personne ne vient relancer l'internaute, et l'ergonomie limitée du logiciel que nous avons utilisé ne favorise pas la saisie des déplacements en ligne (3). Ainsi, si nous reprenons la même analyse uniquement pour les personnes mobiles le jour de l'enquête, l'écart subsiste mais est réduit (3,71 déplacements contre 4,08). Les « petits » déplacements ont été saisis moins fréquemment par les répondants en ligne, soit pour ne pas prendre trop de temps lors de la saisie du questionnaire, soit parce qu'ils considèrent que certains déplacements ne sont pas importants. Il est également possible, comme les études des années 1980 sur les schémas d'activités l'ont montré (Jones *et al.*, 1980), qu'une entrée « activité » et non pas « déplacement » aurait permis de réduire ces omissions. Mais pour assurer la cohérence avec l'enquête en face à face, ce choix n'a pas été possible.

L'analyse en termes de sorties du domicile (enchaînement des déplacements entre une sortie et un retour successif au domicile) fournit des résultats très proches (1,19 sorties contre 1,47 pour l'ensemble de la population, et 1,47 contre 1,65 pour les personnes mobiles). Une analyse plus fine montre en fait que pour l'ensemble de la population, le nombre de déplacements par sortie est similaire pour les deux modes d'enquête (2,5 déplacements contre 2,47 pour l'ensemble de la population, et 2,52 contre 2,47 pour les personnes mobiles).

3. Les contraintes de réalisation n'ont pas permis de choisir le prestataire et ont conduit à transposer un logiciel CATI, utilisé pour les enquêtes téléphoniques assistées par ordinateur.

Tableau 5
Nombre moyen de déplacements le jour de référence de l'enquête par personne et par mode d'enquête

	Nombre de personnes	Minimum	Maximum	Moyenne	Écart-type
Enquête web					
Total	369	0	15	3,00	2,33
Mobiles (1)	299	1	15	3,71	2,02
Enquête face à face					
Total	13 271	0	23	3,63	2,49
Mobiles (1)	11 812	1	23	4,08	2,26

1. Mobiles : personnes s'étant déplacées pendant la période de référence.

Lecture : les personnes interrogées en face à face ont effectué en moyenne 3,63 déplacements le jour de référence de l'enquête.
Champ : ménages résidant dans le Scot de Lyon.
Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

Le redressement de l'échantillon face à face

Compte tenu des différences socio-économiques des deux échantillons, il n'est pas possible d'imputer les différences de mobilité uniquement au choix du média. Pour tenter d'y remédier, nous allons redresser l'échantillon face à face, de façon à le rendre comparable à l'échantillon web (nous préférons redresser l'échantillon en face à face plutôt que celui du web, en raison du faible effectif de ce dernier).

La méthode du calage sur marges

Ce redressement consiste à remplacer les pondérations initiales observées dans un échantillon en utilisant une information auxiliaire disponible sur un certain nombre de variables, appelées variables de calage. Les poids de sondage sont ainsi remplacés par des poids de calage. Après cette opération, l'échantillon peut restituer les totaux de variables quantitatives ou les effectifs de modalités de variables catégorielles connus sur une population. L'intérêt de cette méthode est notamment d'assurer la cohérence entre les résultats de plusieurs enquêtes. La méthode de redressement utilisée pour le calage sur marges est le *raking ratio*. Les variables utilisées pour le redressement étant toutes catégorielles, le redressement consiste à « caler » les marges du tableau croisant toutes les variables de calage sur des effectifs déterminés *a priori*. La macro « Calmar », développée par l'Insee (Sautory, 1993), nous a permis de réaliser ces calculs. La méthode est itérative et consiste à minimiser la somme pondérée par les poids initiaux des distances entre les poids de sondages et les poids de calage, sous les contraintes du calage.

Les variables de calage doivent être corrélées avec les indicateurs à estimer. Nous cherchons donc dans un premier temps les variables (sociodémographiques, de localisation, d'équipement, etc.) qui impactent le plus le choix de

répondre au questionnaire en ligne. À l'aide d'une régression logistique (modèle Logit), nous avons identifié sept variables importantes, qui se répartissent en deux types : des variables sociodémographiques (sexe, âge, possession du permis, niveau d'études, nombre de personnes du ménage) et des variables d'équipement en moyens de communication (possession d'un téléphone portable et possession d'une connexion internet). Une fois sélectionnées les variables principales qui caractérisent les répondants en ligne, nous calculons les effectifs des différentes modalités. La macro « Calmar » redresse ensuite l'échantillon en face à face, de sorte que ces poids soient égaux dans les deux populations, web et face à face.

Un différentiel de mobilité qui subsiste

En comparant les résultats de l'enquête web à ceux de l'enquête face à face redressée, il est intéressant de constater que la part des répondants ne s'étant pas déplacés dans l'aire d'étude durant la période de référence est beaucoup plus élevée dans l'enquête web (19 % contre 7,5 %) alors que celle des hyper mobiles (sept déplacements et plus) est plus faible (6 % contre 13,7 %). Nous recensons en moyenne 4,04 déplacements par individu interrogé en face à face, contre seulement 3,00 par internaute. L'écart s'est creusé par rapport à la population totale (3,00 contre 3,63). Cette population étant plus active et diplômée que celle de référence, les individus se déplacent davantage (cf. tableau 6).

Après redressement de l'échantillon en face à face, le nombre de déplacements en voiture est un peu plus faible pour les internautes (2,08 contre 2,52, cf. tableau 7). Pour la marche à pied, on observe une très faible utilisation (0,45 déplacements pour la population internaute contre 0,98 pour la population face à face). En revanche, pour les autres modes de transport, la mobilité est du même ordre de grandeur. Il est également intéressant de com-

Tableau 6

Nombre moyen de déplacements le jour de référence de l'enquête par personne et par mode d'enquête, après redressement

	Nombre de personnes	Minimum	Maximum	Moyenne	Écart-type
Enquête web	369	0	15	3,00	2,32
Enquête face à face	13 271	0	23	4,04	2,60

Lecture : après redressement, les personnes interrogées en face à face ont effectué en moyenne 4,04 déplacements le jour de référence de l'enquête.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

parer le nombre de déplacements par mode de transport, pour les personnes mobiles seulement, c'est-à-dire celles s'étant déplacées dans l'aire d'étude pendant la période de référence. La proportion de personnes immobiles étant nettement supérieure parmi les internautes, la différence de mobilité se réduit lorsque l'analyse se limite aux personnes mobiles. Elle subsiste toutefois pour la marche à pied, les internautes se déplaçant moins à pied que les individus de l'échantillon face à face. On observe également une utilisation moins soutenue de la voiture en tant que conducteur. Pour les autres modes de transport, les résultats sont très proches, même si les effectifs sont parfois limités.

L'analyse des motifs de déplacement montre un poids très fort du « travail » (21,6 % en face à face et 22,9 % sur le web), puisqu'il s'agit d'une

population en âge de travailler. La mobilité liée au travail est très proche pour les deux populations. En revanche, on note une sous-estimation pour les accompagnements (10,2 % en face à face contre 8,2 % sur le web). Si l'on analyse les données, non plus en répartition par motif, mais en nombre de déplacements par motif, on arrive à une conclusion un peu différente. La mobilité des personnes mobiles est un peu plus faible pour les internautes par rapport aux personnes interrogées en face à face. L'écart concerne tous les motifs, même s'il est plus prononcé pour les motifs non contraints et pour l'accompagnement. Il concerne donc surtout de petits déplacements ou des déplacements considérés comme moins importants par les internautes (cf. tableau 8).

La distance moyenne des déplacements effectués par les individus après le redressement

Tableau 7
Nombre moyen de déplacements le jour de référence de l'enquête par mode de transport et par mode d'enquête après redressement

	Enquête web		Enquête face à face	
	Total	Mobiles (1)	Total	Mobiles (1)
Deux-roues motorisé	0,04	0,05	0,02	0,02
Marche à pied	0,45	0,56	0,98	1,10
Transports en commun	0,38	0,47	0,45	0,50
Vélo	0,06	0,07	0,07	0,07
Voiture conducteur	1,74	2,15	2,21	2,48
Voiture conducteur + transport en commun	0,06	0,08	0,03	0,04
Voiture passager	0,27	0,33	0,27	0,30
Voiture passager + transport en commun	0,01	0,01	0,01	0,01
Total	3,00	3,71	4,04	4,53

1. Mobiles : personnes s'étant déplacées pendant la période de référence.

Lecture : après redressement, les personnes interrogées en face à face ont effectué en moyenne 0,98 déplacement à pied le jour de référence de l'enquête.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

Tableau 8
Nombre moyen de déplacements par motif et par mode d'enquête après redressement

	Enquête web		Enquête face à face	
	Total	Mobiles (1)	Total	Mobiles (1)
Domicile	1,15	1,43	1,53	1,71
Travail	0,69	0,85	0,87	0,98
Secondaire – Université	0,03	0,03	0,09	0,10
Achats – Services	0,49	0,60	0,60	0,68
Accompagnements	0,26	0,32	0,41	0,46
Loisirs	0,38	0,47	0,51	0,57
Autres	0,01	0,01	0,03	0,03
Total	3,00	3,72	4,04	4,53

1. Mobiles : personnes s'étant déplacées pendant la période de référence.

Lecture : après redressement, les personnes interrogées en face à face ont effectué en moyenne 0,87 déplacement vers leur lieu de travail le jour de référence de l'enquête.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

est légèrement supérieure dans l'enquête web (4,74 km, contre 4,08 km), cf. tableau 9. L'écart est toutefois moins important que pour la population globale de l'enquête (4,61 km sur le web, contre 3,78 km en face à face), car les actifs sont à présent sur représentés dans l'enquête en face à face, comme dans l'échantillon web, et les déplacements domicile-travail augmentent la moyenne des déplacements quotidiens.

La durée moyenne des déplacements est supérieure dans l'enquête web (22,96 mn, contre 19,20 mn). Cet écart reste ici du même ordre que celui observé pour l'ensemble de la population (22,96 mn, contre 19,25 mn). Ici encore, l'évaluation des budgets temps et budgets distances fournit des renseignements intéressants (cf. tableau 10). Il semble que la durée quotidienne consacrée aux déplacements soit quasi identique entre les deux échantillons après redressement (environ 86 mn), ce qui n'infirme pas l'hypothèse de Zahavi (1979) selon laquelle le budget de temps de transport quotidien moyen est stable et environ égal à une heure. Toutefois, la distance parcourue semble être plus faible chez les internautes (17,18 km, contre 18,55 km).

La distribution des déplacements dans le temps montre des pointes un peu plus marquées pour les répondants web surtout le matin et une pointe plus tardive le soir alors que celle du matin se

déroule aux mêmes horaires. Les déplacements débutant entre 14h00 et 17h00 sont sous-représentés dans l'enquête web (14,3 %, contre 18,8 %), alors que les déplacements débutant entre 19h00 et 22h00 sont sur-représentés dans cet échantillon (13,5 % contre 10,4 %). L'amplitude de la présence hors domicile est donc plus importante chez les internautes. Ce constat s'explique notamment par les différences observées en termes de PCS, puisque parmi les individus ayant répondu à l'enquête web, les cadres et professions intellectuelles supérieures sont sur représentés.

Des populations proches, mais aux pratiques différentes

La comparaison des données des deux enquêtes montre que les répondants web déclarent en moyenne moins de déplacements quotidiens que les répondants à l'enquête en face à face. Ce constat s'explique notamment par les différences socioéconomiques qui caractérisent les deux échantillons. En redressant l'enquête en face à face, on cherche à isoler une population dont les caractéristiques socio-économiques sont proches de celles de l'enquête web et dont les effectifs sont suffisants pour avoir des résultats statistiquement significatifs. Les différences en termes de nombre de déplacements subsistent et ont même tendance à s'amplifier un peu par rapport à l'ensemble de la population. Cet écart s'explique à la fois par une plus forte immobilité des internautes et par une moindre déclaration de déplacements. Dans l'enquête web, la marche à pied est sous-estimée et l'utilisation de la voiture est un peu plus faible. Les internautes ont une mobilité en termes de déplacements pour le motif « travail » proche de celle des individus de l'enquête standard. L'écart de mobilité se porte donc sur les autres motifs et tout particulièrement sur les « accompagnements et loisirs ». Enfin, les déplacements restent beaucoup plus longs en termes de temps et de distance, très probablement du fait du poids très réduit des motifs « accompagnements et loisirs » et du très faible nombre de déplacements à pied.

Deux interprétations de ces résultats sont possibles. La première conduit à une sous-estimation de la mobilité imputable à l'utilisation du média web, le manque d'ergonomie et la lourdeur de la saisie des déplacements pouvant inciter à des omissions. Il est ainsi possible que la plus forte immobilité soit en grande partie liée au média web. La seconde interprétation

Tableau 9
Distance et durée moyenne des déplacements par mode d'enquête après redressement

	Enquête web	Enquête face à face
Distance (km)	4,61	4,08
Durée (mn)	22,96	19,20
Nombre d'observations	1 066	47 415

Lecture : après redressement, la distance moyenne des déplacements quotidiens effectués par les personnes interrogées en face à face est de 4,08 km
Champ : ménages résidant dans le Scot de Lyon.
Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

Tableau 10
Budgets temps et distances par enquête et par personne après redressement

	Enquête web	Enquête face à face
Budget Distance (km)	17,18	18,45
Budget temps (mn)	85,57	86,88
Nombre d'observations	286	11 621

Lecture : après redressement, le budget distance des personnes interrogées en face à face est de 18,45 km.
Champ : ménages résidant dans le Scot de Lyon.
Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

concerne les caractéristiques socio-économiques des internautes. Il peut en effet exister des raisons pour que les individus qui répondent sur le web ne présentent pas un niveau de mobilité équivalent à ceux qui ont répondu en face à face et que ces raisons ne soient pas liées au mode de réponse, mais à des caractéristiques socio-économiques inobservées ou à d'autres facteurs explicatifs. Dans ce cas, l'enquête web permet d'enquêter des individus qui échappent à l'enquête face à face. On pourrait alors par exemple expliquer une part de la moindre mobilité des internautes par des journées professionnelles plus longues et des déplacements plus longs qui viendraient « saturer » le budget temps quotidien de déplacements en reprenant la conjecture de Zahavi (1979). Nous pouvons scinder la différence de mobilité totale observée entre les répondants web et face à face en deux parties : l'effet de sélection et l'effet du mode d'enquête. La simple comparaison de la mobilité des répondants aux deux médias ne permet pas de discerner ces deux interprétations. Dans la section suivante, nous proposons une méthode économétrique permettant de tester la pertinence de chacune des deux interprétations. Elle consiste dans un premier temps à tester, puis à corriger, l'existence d'un biais de sélection (4) entre les deux échantillons web et face à face. Nous pouvons alors, dans un second temps, identifier l'existence d'un effet lié au mode d'enquête.

Le biais de sélection de l'échantillon

Selon la méthodologie que nous avons utilisée, les individus choisissent de répondre ou non en face à face, et dans la négative de remplir ou pas le questionnaire en ligne. Les groupes de répondants peuvent différer sur des aspects systématiques et lorsque des observations sont exclues d'un échantillon de manière non aléatoire, il y a un risque de biais de sélection. Les réponses ne sont pas comparables, puisque la présence des répondants dans un groupe est déterminée par des facteurs extérieurs, qui peuvent également impacter la variable d'intérêt du modèle étudié. Dit autrement, il est fort probable que des caractéristiques socioéconomiques, pas toujours observables, influencent le choix des individus de recevoir un enquêteur à domicile ou de répondre, le cas échéant, sur le web et impactent leurs comportements de mobilité (Berk, 1983 ; Resource system group, 2002). Ignorer l'existence du biais de sélection peut

donc avoir des conséquences sur la validité du modèle.

Origine du biais de sélection

Considérons l'équation (1) qui permet d'examiner l'effet du mode d'enquête sur le nombre de déplacements quotidiens moyen déclaré par un individu à l'aide d'une régression linéaire :

$$Y_i = \sum_k \beta_k X_{ki} + \alpha I_i + u_i \quad (1)$$

avec Y_i le nombre moyen de déplacements déclarés (variable dépendante), X_i un vecteur de variables explicatives et I_i une variable muette indiquant si l'individu a répondu sur internet. On peut se demander si le coefficient α mesure l'impact réel du mode d'enquête sur la mobilité quotidienne déclarée. La réponse est positive si l'individu qui choisit de répondre sur le web déclare un nombre de déplacements identiques à celui qu'il aurait déclaré en face à face. Cependant, la variable I ne peut pas être considérée comme exogène dans ce modèle. Les individus contactés choisissent de répondre ou pas en face à face, et éventuellement acceptent de saisir leur réponse en ligne, selon certaines caractéristiques qui peuvent également avoir un impact sur leur niveau de mobilité. Ainsi, parmi les ménages qui répondent en ligne il y en a pour lesquels il n'a pas été possible de fixer un rendez-vous avec un enquêteur à domicile. On peut supposer qu'il s'agit de ménages peu disponibles (contraints par leur activité professionnelle par exemple), dont le manque de disponibilité impacte négativement la mobilité, et qui disposent d'un accès à internet pour accéder à l'enquête en ligne. L'échantillon web n'est donc pas représentatif des habitants de l'agglomération, et cette auto-sélection des répondants doit être corrigée afin d'obtenir des estimations non biaisées des coefficients des variables explicatives du nombre de déplacements quotidiens. En excluant systématiquement des observations de la même manière (individus n'ayant pas accès au web par exemple), on introduit le besoin d'un régresseur additionnel que la méthode des moindres carrés ordinaires ignore (Kmenta, 1971). Le problème de l'évaluation de l'effet du mode d'enquête sur la mobilité provient du fait qu'il n'est pas possible d'observer simultanément la mobilité déclarée sur le web et en face à face pour un répondant donné.

4. Un modèle de sélection est un modèle dans lequel la variable dépendante Y n'est pas toujours observée. Le critère de sélection ne porte pas directement sur la valeur de Y , mais est défini par une équation auxiliaire.

Développements théoriques

Dillman (1978) a mis en évidence le biais de sélection qui résulte de protocoles d'enquêtes où plusieurs modes de recueil de données sont disponibles. Les sciences sociales contiennent plusieurs présentations formelles du biais de sélection (Heckman, 1979 ; Goldberger, 1981). Le modèle utilisé pour mettre en évidence le problème d'auto-sélection des répondants est le modèle Tobit II (5), appelé plus couramment « modèle de sélection de l'échantillon ». La méthode traditionnellement utilisée pour mettre en évidence le biais de sélection des répondants est la procédure d'estimation en deux étapes développée par Heckman en 1979 (cf. annexe 1). La première étape consiste à estimer l'équation de sélection traduisant la probabilité pour un individu de refuser l'interview en face à face puis d'accepter de répondre sur le web, à l'aide d'un modèle Probit. Pour chaque observation sélectionnée, le modèle calcule la valeur d'une variable appelée « inverse du ratio de Mills ». Cette variable représente pour chaque observation la probabilité d'être inclus dans l'échantillon web ou face à face et capture la valeur espérée de l'équation de mobilité (Y_i) après l'effet de sélection. La seconde étape consiste à estimer l'équation de mobilité, par une régression des moindres carrés ordinaires.

L'existence d'un biais de sélection est testée par l'hypothèse que le coefficient estimé de la variable appelée inverse du ratio de Mills est nul dans chaque groupe (répondants web et face à face). S'il n'est pas possible de rejeter l'hypothèse d'absence de corrélation entre les termes d'erreur des équations de sélection et d'intérêt, alors il n'y a pas de biais de sélection significatif dans le modèle et on peut appliquer la méthode des moindres carrés ordinaires pour estimer directement les coefficients des variables explicatives du modèle. Dans le cas contraire, l'effet de sélection est significatif et le choix du mode d'enquête apparaît, sous ces hypothèses, endogène au niveau de mobilité. La méthode en deux étapes permet d'obtenir des estimations non biaisées des coefficients des variables explicatives du modèle.

Le paramètre qui fait que le modèle de sélection proposé par Heckman diffère d'un modèle Probit suivi d'un modèle de régression linéaire est l'existence d'un coefficient de corrélation (ou covariance) entre les termes d'erreur des équations de sélection et d'intérêt (Verbeek, 2004). Il est alors nécessaire de supposer que des facteurs inobservables jouent à la fois sur

le niveau de mobilité des individus et sur le média utilisé pour répondre à l'enquête. Il est difficile d'estimer le signe et l'importance de la corrélation entre les deux termes d'erreur (6). Les variables représentant les inverses du ratio de Mills sont par définition positives. En revanche, leurs coefficients peuvent prendre tous les signes. Si ce coefficient est négatif dans le modèle appliqué aux répondants web, alors les termes d'erreurs des équations de sélection et d'intérêt sont négativement corrélés. Il existe des facteurs inobservés qui font que les individus ne peuvent répondre en face à face, mais remplissent le questionnaire en ligne, et qui impactent négativement la mobilité. Cela signifie que le nombre de déplacements déclarés pourrait être en moyenne significativement plus élevé si ces individus avaient répondu en face à face. Par analogie, si ce coefficient est négatif dans le modèle appliqué aux répondants en face à face, le nombre de déplacements déclarés par les individus en face à face aurait été en moyenne significativement inférieur si ces derniers avaient choisi le web comme mode d'enquête. L'interprétation est inversée si ces coefficients sont positifs.

Application à l'Enquête Ménages Déplacements de Lyon

Dans notre échantillon, nous posons l'hypothèse que le mode de collecte des données a un impact sur la mobilité déclarée des individus. Après avoir présenté les variables explicatives disponibles, nous appliquons la méthode en deux étapes détaillée ci-dessus. Il s'agit d'abord d'estimer l'équation de « choix » du média d'enquête à l'aide d'un modèle Probit, puis d'expliquer les différences de mobilité des individus, au moyen d'un modèle spécifique qui permet de s'affranchir des effets du mode d'enquête.

Les variables disponibles pour l'analyse de la mobilité

Face au grand nombre de variables disponibles (sociodémographiques, concernant le niveau d'équipement en télécommunication des ména-

5. Cette classification des biais de sélection est due à Amemiya (1984).

6. Un moyen d'estimer le signe et l'importance de la corrélation entre les deux termes d'erreur est de trouver une variable omise dans les deux équations, qui ne soit pas corrélée avec les régresseurs.

ges, ou caractéristiques de la mobilité individuelle), le choix des variables pertinentes à inclure dans le modèle est complexe. Les analyses exploratoires ont mis en évidence l'importance de dix-sept d'entre-elles (cf. tableaux 11 et 12).

Nous pouvons formuler des hypothèses concernant l'impact de ces variables explicatives sur le nombre de déplacements quotidiens moyen des individus. *A priori*, les femmes se déplacent davantage que les hommes, car elles cumulent souvent des activités professionnelles et domestiques (accompagnement des enfants, achats et démarches administratives du ménage, etc.). L'âge est lié au statut des individus et à leur cycle de vie, donc à la composition du ménage. On observe que le nombre de déplacements augmente avec l'âge jusqu'à un maximum (situé entre 40 et 50 ans), puis décroît par la suite. L'impact du nombre de personnes du ménage sur la mobilité est ambigu. Nous supposons toutefois que les déplacements contraints sont mieux répartis lorsque la taille du ménage augmente. Par ailleurs, les analyses exploratoires laissent penser que le nombre de déplacements augmente avec le nombre d'enfants du ménage. Le niveau d'étude est quant à lui lié au revenu et au statut. Il est corrélé positivement avec la mobilité. Il en est de même pour l'éloignement géographique du domicile (par rapport au centre ville), puisque les activités sont dispersées en périphérie, ce qui occasionne davantage d'accompagnements. Cet éloignement en périphérie dépend d'un ensemble de variables liées au cycle de vie du ménage (effet de l'âge, du revenu, de la taille du ménage, de sa motorisation, etc.).

Concernant les équipements du ménage, la possession d'un téléphone portable est liée positivement à la mobilité (nombreux contacts personnels, ou nécessité de se déplacer pour motif professionnel), cf. Mokhtarian *et al.*, 2006. La

variable « liste de téléphone » est scindée en trois modalités, selon que l'individu ne possède pas de téléphone fixe, est inscrit sur la liste rouge ou

Tableau 12
Effectifs des variables qualitatives (échantillon face à face et web)

Variables	Effectifs
Sexe	10 081
Homme	4 757
Femme	5 324
Possession du permis	10 081
Oui	8 671
Non	1 410
Déplacements le vendredi	10 081
Oui	1 585
Non	8 496
Activité	10 081
Actif : travail au centre	2 474
Actif : travail périphérie	2 923
Actif : travail non précis	274
Non actif	4 410
Accepte de déclarer ses revenus	10 081
Oui	6 649
Non	3 432
Connexion internet	10 067
Oui	5 916
Non	4 151
Liste de téléphone	10 063
Annuaire des abonnés	6 934
Liste (rouge, orange)	2 136
Pas de ligne fixe	993
Téléphone portable	10 064
Oui	7 392
Non	2 672
Niveau d'études	9 980
Supérieur	3 682
Non supérieur	5 432
En cours	866
Diplôme	9 980
Supérieur	4 337
Non supérieur	5 643
Mode d'enquête	10 081
Web	270
Face à face	9 811

Lecture : 4 757 hommes et 5 324 femmes ont été interrogés, tous modes d'enquête confondus.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

Tableau 11
Statistiques descriptives des variables quantitatives (échantillon face à face et web)

Variables	Moyenne	Écart-type	Observations
Âge	46,21	17,51	10 081
Nombre d'enfants du ménage	0,68	1,02	10 081
Nombre de voitures du ménage	0,68	0,39	10 081
Nombre de personnes du ménage	2,79	1,41	10 081
Distance du domicile au centre de l'agglomération (m)	7 607	5 163	10 081
Densité de la zone de résidence (habitants / km ²)	5 899	5 876	10 081

Lecture : en moyenne, l'âge des personnes interrogées est de 46 ans. Cette information a été observée sur 10 081 individus, tous modes d'enquête confondus.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

orange ou est inscrit sur l'annuaire des abonnés. Les packages de téléphonie haut débit, couplés à l'abonnement internet, permettent de ne plus être abonnés à France Télécom, et donc de ne plus figurer sur l'annuaire. Il est probable que les internautes affectionnent les nouvelles technologies, et certains ne disposent plus de ligne fixe, mais d'un téléphone mobile uniquement. Le niveau de motorisation (nombre de voitures du ménage par personne de 18 ans et plus) et le fait de posséder le permis de conduire sont corrélés à la mobilité (les personnes qui n'ont pas le permis de conduire sont plus contraintes dans leurs déplacements), même si d'autres facteurs interviennent. Nos premiers tris ont également montré un accroissement des déplacements le vendredi. L'hypothèse est que la réduction du temps de travail (RTT) profite aux déplacements d'achats et de loisirs. Par ailleurs, on observe que la mobilité est plus importante quand la personne choisit de communiquer le niveau de revenus annuels nets du ménage. Il est possible que ces répondants doivent faire face à moins de freins pour délivrer des réponses précises, bien que le fait de dévoiler le niveau de salaire ne soit pas une variable directement liée au niveau de mobilité. Enfin, nous avons vu que la mobilité est plus faible chez les individus ayant répondu sur le web.

Première étape : équation de sélection

Deux types de variables vont être introduites dans notre modèle en qualité de variables explicatives : des caractéristiques sociodémographiques (âge, occupation, nombre de personnes du ménage, niveau d'études, niveau de revenus annuels du ménage, lieu de travail), puis des variables concernant l'équipement en télécommunication des ménages (connexion internet, téléphone portable, inscription sur l'annuaire des abonnés). Le modèle Probit explicatif du choix du questionnaire web, est appliqué aux variables sélectionnées ci-dessus (cf. tableau 13).

Les coefficients estimés dans le modèle Probit sont significatifs (probabilité inférieure à 5 % de rejeter à tort l'hypothèse « le coefficient est nul »). L'étude de leur signe donne une idée de l'influence des divers facteurs sur le choix de répondre par le web : un coefficient positif augmente la probabilité de répondre en ligne, mais un coefficient négatif diminue la probabilité de répondre sur internet (7). L'ensemble des signes des coefficients sont ceux que nous attendions.

7. Nous travaillons en écart par rapport à une personne de référence, ou à une modalité en ce qui concerne les variables nominales.

Tableau 13
Équation de sélection (modèle Probit)

	Coefficient	Écart-type	Pr(> z)	Significativité
Constante	- 3,712	3,469e-01	< 2e-16	***
Âge	0,049	1,437e-02	0,000663	***
(Âge) ²	- 4,78e-04	1,618e-04	0,003136	**
Connexion internet : oui	0,495	8,648e-02	1,03e-08	***
Téléphone portable : oui	0,257	9,186e-02	0,005156	**
Liste téléphone : oui	0,372	6,598e-02	1,68e-08	***
Liste téléphone : pas de téléphone	0,248	1,180e-01	0,035711	*
Nb de personnes / ménage	- 0,101	2,494e-02	5,10e-05	***
Lieu de travail : non précisé	- 0,993	3,640e-01	0,006401	**
Lieu de travail : périphérie	- 0,24	7,392e-02	0,001184	**
Lieu de travail : inactif	- 0,244	8,802e-02	0,005611	**
Diplôme : supérieur	0,41	6,674e-02	8,16e-10	***
Densité de la zone de résidence	- 2,169e-05	6,684e-06	0,001174	**
Revenu déclaré : oui	0,401	7,543e-02	1,04e-07	***
Déplacements le vendredi : oui	0,469	6,661e-02	1,89e-12	***
Distance domicile / centre	- 1, 9e-05	7,898e-03	0,014361	*
Degrés de liberté = 9 980				
Qualité d'ajustement (pseudo R ²) = 15,2 %				

Lecture : le fait de posséder une connexion internet à domicile augmente la probabilité de répondre au questionnaire web (coefficient = 0,495 et p-value < 0,001 %). *** : significatif au seuil de 0,1 %, ** : significatif au seuil de 1 %, * : significatif au seuil de 5 %, '.' : significatif au seuil de 10 %.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

La possession d'une connexion internet au domicile, d'un téléphone portable, l'inscription sur la liste rouge ou orange ou l'absence de ligne de téléphone fixe au domicile ou le fait de déclarer son niveau de revenus augmentent la probabilité de répondre sur le web. Bien que l'accès au site hébergeant le questionnaire puisse se faire en dehors du domicile, les ménages possédant une connexion internet ont une probabilité plus élevée de répondre sur le web. Les internautes sont également mieux équipés que la moyenne en moyens de communication, ou appartiennent à une catégorie socioprofessionnelle plus élevée (possession d'un téléphone portable personnel ou professionnel). Les packages de téléphonie haut-débit sont accessibles aux personnes possédant une connexion internet. Fortement sollicités par les démarches commerciales, il est probable que les internautes préfèrent s'exclure de l'annuaire des abonnés en s'inscrivant sur la liste rouge ou orange.

La probabilité de répondre en ligne augmente également avec le niveau d'études et l'âge. Inversement, elle diminue avec l'éloignement géographique du lieu de travail des actifs (par rapport au centre de l'agglomération), le fait d'être inactif et le nombre de personnes du ménage. Les personnes diplômées sont plus familières avec internet (Wang et Law, 2007). Elles ont utilisé cet outil durant leurs études et l'utilisent encore au domicile ou sur leur lieu de travail. L'âge est lié au cycle de vie des individus. Des études montrent qu'avec la progression dans la vie professionnelle, les revenus et l'équipement en moyens de communication s'améliorent (Bigot et Crouette, 2007). En revanche, internet est une technologie relativement récente et non maîtrisée par tous, l'usage d'internet diminuant avec l'âge. L'introduction d'un terme quadratique permet de prendre en compte un impact non linéaire de l'âge. Par ailleurs, la localisation des entreprises est liée aux types d'emplois. Les bureaux, où travaillent des personnes qualifiées qui disposent plus facilement d'un accès à internet sur leur lieu de travail, sont davantage présents dans le centre de l'agglomération. Les inactifs ont en revanche du temps à consacrer à un enquêteur à domicile, et sont donc moins tentés par un questionnaire web.

La valeur du coefficient de détermination du modèle (pseudo R^2) indique qu'environ 15,2 % de la variation dans le choix du mode de réponse (web ou face à face) est expliqué par les variables explicatives retenues dans le modèle. Ce résultat peut s'expliquer par l'absence d'une variable explicative importante dans le modèle

de sélection. Si cette variable influe également sur le comportement de mobilité, alors il est impossible de modéliser le nombre de déplacements quotidiens par une simple régression linéaire. Le modèle Probit utilisé ci-dessus apporte une solution, en calculant pour chaque individu un facteur de correction du biais de sélection, appelé inverse du ratio de Mills.

Deuxième étape : équation d'intérêt

La deuxième étape consiste à expliquer les différences de comportements en termes de mobilité, au moyen d'un modèle spécifique qui comprend : une variable dépendante (le nombre moyen de déplacements réalisé par les individus), plusieurs variables indépendantes ou explicatives (les facteurs observés censés avoir un effet sur la mobilité), l'inverse du rapport de Mills (variable obtenue dans la première étape) et un terme d'erreur (pour tenir compte des facteurs non observés qui pourraient influencer sur la mesure des résultats). En réalité, nous estimerons deux modèles, un sur le sous-échantillon des répondants en face à face, et un sur le sous-échantillon des répondants sur le web. L'estimation des coefficients des variables explicatives et de la variable relative au biais de sélection se fait par une régression par moindres carrés ordinaires. Nous ne retenons ici que les variables qui impactent directement la mobilité des individus. Les variables sociodémographiques prises en compte dans l'équation de sélection ne sont réintroduites que si elles semblent jouer un rôle significatif sur la mobilité quotidienne. Le modèle restreint au sous-échantillon des individus ayant répondu en face à face nous donne des résultats intéressants, puisque l'ensemble des variables sont significatives et que les coefficients prennent les signes attendus (cf. tableau 14).

Le fait d'être un homme et d'appartenir à un ménage de taille élevée impacte négativement la propension à se déplacer. À l'inverse, la mobilité semble être une fonction croissante de l'éloignement du domicile par rapport au centre de l'agglomération, du nombre d'enfants et de la volonté de déclarer son niveau de revenus annuels nets. Ces résultats s'expliquent notamment par les déplacements pour motif « accompagnement », plus nombreux pour les habitants de la périphérie (dispersion des activités), les ménages avec enfants (vie scolaire et associative) et les femmes. La moindre réticence des personnes qui déclarent leurs revenus à communiquer l'ensemble des activités effectuées la

veille de l'interview impacte favorablement la mobilité mesurée. Le nombre de déplacements quotidiens moyen augmente également avec la possession du permis de conduire et le nombre moyen de voitures par personne du ménage en âge de conduire. L'effet de la motorisation ou de la possession du permis de conduire est important, puisque le fait de ne pas pouvoir se déplacer en voiture particulière limite les occasions de déplacement. Enfin, l'introduction de termes quadratiques traduit un impact non linéaire de l'âge et du nombre d'enfants sur la mobilité. Le nombre de déplacements croît jusqu'à l'âge de 40 ans, puis décroît par la suite.

Le coefficient associé à l'inverse du ratio de Mills est significatif. Il existe donc un biais de sélection des individus. Le signe de ce coefficient est négatif, ce qui signifie qu'en moyenne le choix du web comme mode d'enquête par ces répondants aurait un impact négatif sur le nombre de déplacements déclarés.

Peu de coefficients sont significatifs, dans le modèle restreint au sous-échantillon des individus ayant répondu en ligne (cf. tableau 15). Ceci s'explique notamment par les différences d'effectifs entre les deux échantillons de répon-

dants. La comparaison des deux modèles estimés ci-dessus montre que l'ordre de grandeur des coefficients est le même, mais que celui de leurs écarts-types varie fortement, puisque 9 811 individus ont été interrogés en face à face contre seulement 270 sur le web (soit un rapport de 1 à 36). L'ordre de grandeur des écart-types des coefficients estimés varie dans un rapport de 1 à 6 entre les deux échantillons web et face à face. Les valeurs de la statistique de test sont donc beaucoup plus faibles dans le cas de l'échantillon web et ne dépassent que rarement le seuil critique de 1,96 permettant de conclure à la significativité statistique des coefficients (au risque $\alpha = 5\%$). Les estimations non corrigées et corrigées des coefficients diffèrent beaucoup. Les coefficients non corrigés du biais de sélection auraient pu inclure des faux positifs ou des faux négatifs, mais ce n'est pas le cas ici puisqu'après correction les coefficients conservent leur signe. L'ajout de la variable « Mills » (biais de sélection) permet d'identifier l'impact réel des facteurs socioéconomiques sur la mobilité des répondants web.

Le fait d'être un homme impacte négativement la propension à se déplacer, les femmes effectuant globalement plus de déplacements. En

Tableau 14
Analyse de la mobilité (échantillon face à face)

	Régression sans correction			Régression avec correction		
	Coefficient	Pr(> z)	Significativité	Coefficient	Pr(> z)	Significativité
Constante	2,013	< 2e-16	***	2,640	< 2e-16	***
Sexe : homme	- 0,24	2,14e-07	***	- 0,24	2,41e-07	***
Âge	0,065	1,38e-15	***	0,057	1,73e-11	***
(Âge) ²	- 7,3e- 04	< 2e-16	***	- 6,4e- 04	2,44e-12	***
Possession permis : oui	0,465	4,61e-10	***	0,444	4,23e-09	***
Nb d'enfants / ménage	0,711	< 2e-16	***	0,710	< 2e-16	***
(Nb d'enfants / ménage) ²	- 0,004	0,0267	*	- 0,003	0,04632	*
Nb de voitures / personne	0,571	< 2e-16	***	0,534	2,00e-14	***
Nb de personnes / ménage	- 0,188	5,13e-12	***	- 0,181	4,78e-11	***
Accepte de déclarer son revenu : oui	0,325	1,33e-11	***	0,268	2,90e-07	***
Activité : non actif	0,288	6,85e-07	***	0,304	2,30e-07	***
Distance domicile / centre	6,7e-06	0,1478		9,5e-06	0,04358	*
Biais de sélection (Mills)	(1)	(1)		- 0,18	0,00146	**
Degrés de liberté = 9 967						
Qualité d'ajustement (R ²) = 9,2 %						
1. Le biais de sélection n'est pas calculé dans ce cas, puisque nous examinons l'effet du mode d'enquête sur le nombre de déplacements quotidiens, sans corriger une éventuelle auto sélection des répondants.						

Lecture : après correction du biais de sélection, le fait de d'être un homme diminue le niveau de mobilité quotidienne pour les répondants en face à face (coefficient = - 0,24 et p-value < 0,001 %). *** : significatif au seuil de 0,1 %, ** : significatif au seuil de 1 %, * : significatif au seuil de 5 %, . : significatif au seuil de 10 %.
Champ : ménages résidant dans le Scot de Lyon.
Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

revanche, contrairement à ce que l'on observe dans l'échantillon en face à face, certains coefficients ne prennent pas le signe attendu. Ainsi, le fait de résider en périphérie diminue la propension à se déplacer. Les actifs cadres sont fortement représentés dans l'échantillon web. Ils habitent souvent en périphérie, disposent d'un haut niveau de formation et déclarent des revenus élevés, mais leur emploi est « chronophage ». Ils ont peu de temps libre en semaine pour effectuer des activités non contraintes, ce qui limite les possibilités de déplacements. Par ailleurs, la mobilité croît ici, avec le nombre de personnes du ménage. Les répondants web ont des composantes socioéconomiques plus élevées, qui leur permettent de réaliser davantage d'activités. Le coefficient de l'inverse du ratio de Mills est significatif. Il y a donc des variables qui influent sur le choix de remplir le questionnaire en ligne et la mobilité des répondants, et un biais d'endogénéité du mode d'enquête sur la mobilité. Le recours à la méthode d'estimation en deux étapes est justifié, puisque le choix du web apparaît, dans ces conditions, endogène au niveau de mobilité déclaré. Par ailleurs, le signe négatif de l'inverse du ratio de Mills signifie que la mobilité pourrait être en moyenne signifi-

cativement plus élevée si ces répondants avaient répondu en face à face.

Les coefficients estimés de la variable « Mills », correspondant au biais de sélection, peuvent prendre tous les signes, en fonction du signe de la corrélation entre les résidus de l'équation de sélection et ceux de l'équation d'intérêt concernant respectivement les répondants web et face à face. Les résidus correspondent à des variables non observées et par conséquent non prises en compte dans le modèle, qui peuvent avoir un effet sur la variable à expliquer. Par exemple, le fait de travailler à temps partiel n'est pas une variable explicative du modèle de mobilité. Pourtant, cette caractéristique implique une moindre présence au travail, ce qui peut avoir un effet sur la disponibilité des individus (capacité à recevoir un enquêteur à domicile) et leur niveau de mobilité.

Dans chaque équation de mobilité, il est possible d'évaluer la corrélation entre l'inverse du ratio de Mills (le biais de sélection) et le niveau de mobilité, noté ρ . Cette valeur permet d'évaluer la force de l'endogénéité du mode d'enquête au nombre de déplacements déclarés. Nous pouvons conclure que le nombre de déplacements,

Tableau 15
Analyse de la mobilité (échantillon web)

	Régression sans correction			Régression avec correction		
	Coefficient	Pr(> z)	Significativité	Coefficient	Pr(> z)	Significativité
Constante	0,894	0,501		3,038	0,076	.
Sexe : homme	- 0,846	7,22e-04	***	- 0,870	0,001	**
Âge	0,077	0,212		0,039	0,542	
(Âge) ²	- 8,57e-04	0,220		- 4,38e-04	0,547	
Possession permis : oui	0,263	0,638		0,454	0,437	
Nb d'enfants / ménage	0,399	0,418		0,465	0,360	
(Nb d'enfants / ménage) ²	- 0,121	0,526		- 0,151	0,433	
Nb de voitures / personne	0,683	0,077	.	0,690	0,082	.
Nb de personnes / ménage	0,266	0,079	.	0,299	0,056	.
Accepte de déclarer son revenu : oui	0,360	0,230		0,166	0,647	
Activité : non actif	0,741	0,030	*	0,823	0,023	*
Distance domicile / centre	- 5,7e-05	0,030	*	- 5,2e-05	0,058	.
Biais de sélection (Mills)	(1)	(1)		- 0,731	1,46e-03	**
Degrés de liberté = 234 Qualité d'ajustement (R ²) = 15,2 %						
1. Le biais de sélection n'est pas calculé dans ce cas, puisque nous examinons l'effet du mode d'enquête sur le nombre de déplacements quotidiens, sans corriger une éventuelle auto sélection des répondants.						

Lecture : après correction du biais de sélection, le fait de d'être un homme diminue le niveau de mobilité quotidienne pour les répondants web (coefficient = - 0,870 et p-value < 0,01 %). *** : significatif au seuil de 0,1 %, ** : significatif au seuil de 1 %, * : significatif au seuil de 5 %, . : significatif au seuil de 10 %.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

pour les répondants web, apparaît fortement corrélié à l'inverse du ratio de Mills ($\rho_1 = -0,383$). Cette corrélation est plus faible dans le groupe des répondants en face à face ($\rho_2 = -0,081$).

Test de stabilité du modèle

Il s'agit maintenant de tester si les estimations des coefficients générés par le modèle en deux étapes sont stables et de quantifier un éventuel impact du mode d'enquête. Nous comparons deux modèles de régression simples, permettant d'expliquer le nombre moyen de déplacements des répondants : un modèle contraint et un modèle non contraint. Le modèle contraint ne considère comme facteurs explicatifs de la mobilité que les variables retenues dans le modèle en deux étapes ci-dessus (cf. tableau 16). Le modèle non contraint inclut également l'ensemble des interactions entre les variables explicatives et le mode d'enquête (web et face à face).

Le test de Wald appliqué aux deux modèles, contraint et non contraint, renvoie une probabilité de rejeter à tort l'hypothèse d'absence d'interaction égale à 0,000321. Nous pouvons donc rejeter l'hypothèse nulle d'équivalence des modèles contraints et non contraints et conclure qu'il existe une interaction significative entre les variables explicatives de la mobilité et le mode d'enquête considéré, qui permet d'expli-

quer le nombre de déplacements effectués par les répondants. Trois variables semblent significativement interagir avec le mode de réponse : le sexe, le nombre de personnes composant le ménage et la distance entre le centre de l'agglomération et le lieu de résidence du ménage. En revanche, le coefficient de la variable « Mills » n'est pas significativement différent entre les deux échantillons (cf. tableau 17).

Impact des interactions entre les variables explicatives de la mobilité et le mode d'enquête après correction du biais de sélection

Nous détaillons l'impact des variables explicatives sur le nombre de déplacements déclarés, après correction du biais de sélection. Notre regard se porte essentiellement sur les variables qui interagissent avec le mode de réponse choisi. L'échantillon web est de taille modeste au regard de l'échantillon face à face. Nous conservons dans l'analyse les interactions significatives au seuil d'erreur de 10 %, exception faite du coefficient de la variable mode d'enquête ($p\text{-value} = 17,6\%$). Le mode de recueil de données, web ou face à face, impacte directement le niveau de mobilité déclaré. Si le questionnaire est rempli en ligne, le nombre de déplacements déclarés décroît de 0,5. Par ailleurs, trois variables semblent interagir avec le mode d'enquête :

Tableau 16
Analyse de la mobilité (modèle contraint)

	Coefficients	Pr(> z)	Significativité
Constante	2,538	< 2e-16	***
Sexe : homme	- 0,251	4,19e-08	***
Âge	0,058	7,72e-12	***
(Âge) ²	- 6,41e-04	1,02e-12	***
Possession permis : oui	0,450	1,84e-09	***
Nb d'enfants / ménage	0,710	< 2e-16	***
(Nb d'enfants / ménage) ²	- 0,035	0,042	*
Nb de voitures / personne	0,529	1,61e-14	***
Nb de personnes / ménage	- 0,170	3,38e-10	***
Accepte de déclarer son revenu : oui	0,263	3,62e-07	***
Activité : non actif	0,320	3,20e-08	***
Distance domicile / centre	8,44e-06	0,069	.
Biais de sélection (Mills)	- 0,155	4,85e-03	**
Degrés de liberté = 9 944			
Qualité d'ajustement (R ²) = 8,99 %			

Lecture : quel que soit le mode de collecte de données, le fait d'être un homme diminue le nombre de déplacements quotidiens (coefficient = - 0,251 et $p\text{-value} < 0,001\%$). *** : significatif au seuil de 0,1 %, ** : significatif au seuil de 1 %, * : significatif au seuil de 5 %, '.' : significatif au seuil de 10 %.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

le sexe, le nombre de personnes du ménage et la distance entre le domicile et le centre de l'agglomération.

Le coefficient de la variable « distance du domicile au centre x mode d'enquête » est légèrement négatif (- 0,000059). Ainsi, la mobilité déclarée des répondants web diminue avec l'éloignement (en mètres) de leur lieu de résidence du centre de l'agglomération : l'impact est de 0,00000954 déplacements pour les répondants en face à face, contre - 0,000049 déplacements pour les répondants web. Les répondants web occupent davantage d'emplois de cadres et d'employés, situés dans le centre de l'agglomération. Avec l'éloignement du lieu de résidence, la distance domicile-travail augmente. Le temps disponible pour les activités moins contraintes est donc réduit, ce qui impacte négativement leur mobilité déclarée. Nous remarquons que ce coefficient est très faible par rapport aux autres. Ceci s'explique par l'unité choisie (m) pour calculer la distance entre le domicile et le centre de l'agglomération. Le fait d'être un homme impacte négativement la mobilité déclarée des répondants web : l'impact est de - 0,240 déplacements par jour pour les répondants en face à

face, contre - 0,787 pour les répondants web. Comme nous l'avons décrit précédemment, les femmes se déplacent davantage que les hommes, car elles doivent gérer des activités familiales en plus de leur activité professionnelle. Cet effet est amplifié en ce qui concerne les répondants web. Nous pouvons penser que leur niveau d'emploi, plus exigeant (beaucoup de cadres), leur laisse encore moins de latitude pour leurs déplacements. Rappelons ici que trois quart des répondants web sont des actifs. *A contrario*, la mobilité déclarée des répondants web augmente avec le nombre de personnes du ménage : l'impact sur le nombre de déplacements quotidiens est de - 0,173 par personne pour les répondants en face à face, contre 0,014 pour les répondants web. Ces derniers disposent d'un revenu annuel supérieur. Il est donc probable que le nombre de personnes accroissent les besoins et donc les déplacements (pour motif achat, loisir, etc.) des membres du ménage.

Les coefficients des variables « mode d'enquête », « (Sexe : homme)*mode d'enquête », « (Nb de personnes / ménage)*mode d'enquête » et « (Distance domicile / centre)*mode d'enquête » permettent ainsi de quantifier, pour cha-

Tableau 17
Analyse de la mobilité (modèle final, interactions significatives)

	Coefficients	Pr(> z)	Significativité
Constante	2,650	< 2e-16	***
Sexe : homme	- 0,240	2,08e-07	***
Âge	0,057	1,03e-11	***
(Âge) ²	- 6,36e- 04	1,53e-12	***
Possession permis : oui	0,445	2,63e-09	***
Nb d'enfants / ménage	0,694	< 2e-16	***
(Nb d'enfants / ménage) ²	- 0,033	0,051	.
Nb de voitures / personne	0,534	8,07e-15	***
Nb de personnes / ménage	- 0,173	1,93e-10	***
Accepte de déclarer son revenu : oui	0,265	2,84e-07	***
Activité : non actif	0,315	5,13e-08	***
Distance domicile / centre	9,54e- 06	0,042	*
Biais de sélection (Mills)	- 0,191	6,15e-04	***
Mode d'enquête	- 0,504	0,176	
(Sexe : homme) * mode d'enquête	- 0,547	0,059	.
(Nb de personnes / ménage) * mode d'enquête	0,187	0,102	
(Distance domicile / centre) * mode d'enquête	- 5,9e-05	0,053	.
Degrés de liberté = 9 940			
Qualité d'ajustement (R ²) = 9,2 %			

Lecture : le fait d'être un homme pèse davantage sur la mobilité des répondants web (- 0,787 = coefficients (Sexe : homme) * mode d'enquête + Sexe : homme) que sur celle des répondants en face à face (- 0,240). *** : significatif au seuil de 0,1 %, ** : significatif au seuil de 1 %, * : significatif au seuil de 5 %, '.' : significatif au seuil de 10 %.

Champ : ménages résidant dans le Scot de Lyon.

Source : Enquêtes Ménages Déplacements web et face à face (Lyon 2006).

que répondant, l'impact du mode d'enquête sur le comportement de mobilité. Soit un homme de 35 ans, actif et possédant le permis de conduire, qui réside à un km du centre ville. On fait l'hypothèse qu'il appartient à un ménage composé de quatre personnes, dont un enfant de moins de 18 ans, avec deux voitures et qu'il n'a pas communiqué ses revenus lors de l'enquête. Ce profil de répondant, déclare sur le web 3,99 déplacements, contre 4,58 en face à face. Si l'on considère à présent un homme actif de 22 ans, sans enfant, vivant en couple à 500 mètres du centre de l'agglomération, qui a son permis, une seule voiture à disposition du ménage et qui a déclaré ses revenus, le nombre de déplacements est égal à 3,97 sur le web et 3,82 en face à face. La différence de mobilité est donc fortement atténuée. *A contrario*, une femme active de 53 ans ayant son permis, appartenant à un ménage de cinq personnes (quatre adultes et un enfant de moins de 18 ans) résidant à cinq km du centre de l'agglomération, qui possède trois voitures et qui n'a pas déclaré ses revenus indique, toutes choses égales par ailleurs, 4,44 déplacements sur le web et 4,30 déplacements en face à face. D'une manière générale, lorsque les valeurs des variables qui interagissent positivement avec le mode d'enquête web sont élevées, la différence de mobilité déclarée s'atténue. *A contrario*, lorsque les valeurs des variables qui interagissent négativement avec le mode d'enquête web sont élevées, la différence de mobilité déclarée s'accroît.

* *
*

Face à la difficulté de recruter des répondants (augmentation du taux de non-réponses, coût des interviews, etc.), aux évolutions de la société (multiplication des enquêtes, désabonnements téléphoniques, etc.) et aux innovations technologiques (internet, GPS), il semble opportun de combiner les média ou les méthodes au sein d'une même enquête pour améliorer la qualité et la représentativité statistique des données tout en réduisant les coûts dans le cas du média web. Des enquêtes de mobilité ont déjà été menées sur le web, mais elles ont toujours été couplées à des enquêtes postales ou téléphoniques et non pas à des enquêtes en face à face. L'enquête que nous avons réalisée est donc la première du genre en France et à notre connaissance dans le monde. Malgré les différences dans la mobilité recueillie qui sont imputables, au moins en partie, au média web, il nous semble que cette enquête a mis en évidence l'intérêt de ce média pour la réalisation d'une enquête déplacements.

Le fait que les internautes soient beaucoup moins présents dans l'annuaire alors qu'ils sont tout autant équipés de téléphone fixe semble indiquer qu'ils souhaitent davantage se protéger des « intrusions » dans le ménage que constituent les enquêtes. Il est donc possible que le web permette d'accéder à une population qui répond peu aux enquêtes. Nous avons vu que les internautes ont des caractéristiques sociodémographiques différentes du reste de la population (niveau d'études et revenus élevés, plutôt actifs, fortement motorisés et équipés en moyens de communication, etc.). En posant l'hypothèse que ces caractéristiques différentes impactent les comportements de mobilité, l'utilisation du web comme média d'enquête apparaît comme un moyen de réduire la non-réponse.

L'état actuel de diffusion d'internet au sein de la population française ne permet pas la réalisation d'une enquête web exclusive auprès de l'ensemble de la population. L'échantillon ainsi constitué ne serait pas représentatif et les résultats biaisés ainsi que l'illustre l'exemple lyonnais. L'utilisation du web dans les protocoles d'enquêtes mixtes est en revanche intéressante, mais impose d'analyser le biais de sélection des individus. Cet article propose une méthode intéressante pour identifier puis s'affranchir de ce biais. La question est évidemment de savoir si ce biais est suffisamment important pour devoir être corrigé. Ceci dépend de la part de la population exclue par un mode de collecte de données et de la précision des données attendue. Dans cette expérience, le relativement faible taux de pénétration d'internet dans la population et les exigences croissantes des modèles de planification ne permettent pas d'occulter le biais de sélection de l'échantillon.

Des développements complémentaires permettraient d'approfondir l'analyse comparative de la mobilité selon le mode d'enquête. D'abord, nous avons montré que la sous-mobilité des répondants web concernait certains types de motifs (non contraints) et de mode de transports (marche à pied). Une analyse des facteurs explicatifs de ces types de déplacements permettrait de corriger plus finement la mobilité quotidienne des internautes. Par ailleurs, la sous-mobilité des répondants web s'explique par une immobilité plus importante. Certaines techniques ont l'avantage de séparer les facteurs explicatifs de la décision de se déplacer de ceux qui influent sur le niveau de la mobilité. Un modèle explicatif de l'immobilité selon le mode d'enquête serait également un moyen de mieux appréhender les facteurs à l'origine de

la sous-mobilité des répondants web. Enfin, la conception d'une enquête plus ergonomique et une attention plus grande aux facteurs pouvant expliquer ces différences de comportements de

réponse aux enquêtes web et face à face permettrait de réduire l'incidence du média. □

BIBLIOGRAPHIE

Amemiya T. (1984), « Tobit Models : A Survey », *Journal of Econometrics*, vol. 24, pp. 3-61.

Bates N. (2001), *Internet versus mail as a data collection methodology from a high coverage population*, Proceedings of the Annual Meeting of the American Statistical Association.

Baudelle G., Darris G., Ollivro J. et Pihan J. (2003), « Les conséquences d'un choix résidentiel périurbain sur la mobilité : pratiques et représentations des ménages », 3^e colloque du Groupe de Travail Mobilités spatiales et fluidité sociale (GT23) : *Offre urbaine et expériences de la mobilité*, Strasbourg, France, Cybergeo.

Bayart C. et Bonnel P. (2008), « Enquête web auprès des non-répondants de l'enquête ménages déplacements de Lyon 2005-2006 », Rapport pour le PREDIT, Laboratoire d'Économie des Transports, Lyon.

Berk R.A. (1983), « An Introduction to Sample Selection Bias in Sociological Data », *American Sociologic Review*, vol. 48, n° 3, pp 386-398.

Bigot R. et Crouette P. (2007), *La diffusion des technologies de l'information dans la société française*, Crédoc.

Bonnel P. et Le Nir M. (1998), « The quality of survey data : Telephone versus face-to-face interviews », *Transportation*, vol. 25, n° 2, pp. 147-167.

Certu (2008), *L'enquête ménages déplacements standard Certu*, éditions du CERTU, Lyon.

Cete de Lyon (2005), *Enquête ménages déplacements et moyens de communication des ménages 'standard CERTU' de l'aire métropolitaine lyonnaise 2005-2006*, Lyon.

Couper M.P. (2000), « Web Surveys : A Review of Issues and Approaches », *Public Opinion Quarterly*, vol. 65, n° 2, pp. 230-253.

Dillman D.A. (1978), *Mail and Telephone Surveys : The Total Design Method*, John Wiley & Sons, New York.

Dillman D.A., Phelps G., Tortora R.D., Swift K., Kohrell J. et Berck, J. (2001), *Response Rate and Measurement Differences in Mixed Mode Surveys Using Mail, Telephone, Interactive Voice Response and the Internet*, The American Association for Public Opinion Research (AAPOR), 56th Annual Conference.

Dillman D.A., Tortora R.D. et Bowker, D. (1998), « Principles for Constructing Web Surveys », working paper, Washington State University.

Givord, P. (2003), « Une nouvelle enquête Emploi », *Économie et Statistique*, n° 362, pp. 60-66.

Goldberger, A.S. (1981), « Linear Regression After Selection », *Journal of Econometrics*, vol. 15, n° 3, pp. 357-366.

Greene, W.H. (2002), « Econometric Analysis », Prentice Hall, Upper Saddle River, New Jersey.

Gunn H. (2002), « Web-based Surveys : Changing the Survey Process », *First Monday*, vol. 7, n° 12.

Heckman J. (1979), « Sample Selection Bias as a Specification Error », *Econometrica*, vol. 47, n° 1, January, pp. 153-161.

Hoffman D. et Link C.R. (1984), « Selectivity Bias in Male Wage Equation : Black-White Comparison », *The Review of Economics and Statistics*, vol. 66, n° 2, pp. 320-324.

Hourriez J.M. et Olier, L. (1997), « Niveau de vie et taille du ménage : estimations d'une échelle d'équivalence », *Économie et statistique*, n° 308-309-310.

Jones P.M., Dix M.C., Clarke M.I. et Heggie I.G. (1980), *Understanding travel behavior*, Oxford studies of Transport, Gower.

- Kmenta J. (1971)**, *Elements of Econometrics*, McMillan, New York.
- Maddala G.S. (1986)**, *Limited Dependant and Qualitative Variables in Econometrics*, Cambridge University Press, London.
- Madre J.-L., Axhausen K. K. et Brög W. (2007)**, « Immobility in Travel Diary Surveys », *Transportation*, vol. 34, n° 1, pp. 107-128.
- Mokhtarian P., Salomon I. et Handy, S (2006)**, « The Impacts of ICT on Leisure Activities and Travel : A conceptual exploration », *Transportation*, vol. 33, n° 3, pp. 263-289.
- Pochet P. (2003)**, « Mobilité et accès à la voiture chez les personnes âgées : évolutions actuelles et enjeux », *Recherche Transport Sécurité*, n° 79, thématique « Conduite automobile et vieillissement », pp. 93-106.
- Potoglou D. et Karanoglou P. (2008)**, *Comparison of phone and web based surveys for collecting household background information*, International Conference on Survey Methods in Transport, Annecy, France.
- Ressource System Group Inc. (2002)**, *Computer-Based Intelligent Travel Survey System*, Documentation for SBIR Phase II Final Report, DTRS57-00-C-10030, Prepared for the FHWA.
- Roy A.D. (1951)**, « Some Thoughts on the Distribution of Earnings », *Oxford Economic Papers*, vol. 3, n° 2, pp. 135-146.
- Sautory O. (1993)**, « Redressement d'un échantillon par calage sur marges », Document de travail de la DSDS, Insee, n° F9310, www.insee.fr.
- Srinivasan K.K. et Athuru S. (2004)**, « Modeling Interaction Between Internet Communication and Travel Activities : Evidence from Bay Area, California, Travel Survey 2000 », *Transportation research record*, n° 1894, pp. 230-240.
- Verbeek M. (2004)**, *A guide to Modern Econometrics*, John Wiley & Sons, England.
- Visser E. et Lanzerdorf M. (2004)**, « Mobility and Accessibility Effects of B2C E-Commerce : a Literature Survey », *Journal of Economic & Social Geography*, vol. 95, n° 2, pp. 189-205.
- Wang D. et Law F.Y.T. (2007)**, « Impacts of Information and Communication Technologies (ICT) on Time Use and Travel Behaviour : a Structural Equations Analysis », *Transportation*, vol. 34, n° 4, pp. 513-527.
- Winship C. et Mare R.D. (1992)**, « Models for Sample Selection Bias », *Annual Review of Sociology*, vol. 18, pp. 27-350.
- Zahavi Y. (1979)**, « The 'UMOT' Project », Report prepared for the U.S. Department of Transportation and the Ministry of Transport of the Federal Republic of Germany.

MODÈLE DE SÉLECTION DE L'ÉCHANTILLON EN DEUX ÉTAPES

Le modèle explicatif de la mobilité peut se formaliser de la façon suivante pour chaque individu i :

$$Y_{1i} = \sum_k \beta_{1k} X_{1ki} + u_{1i}, \text{ pour les répondants web} \quad (2)$$

$$Y_{2i} = \sum_k \beta_{2k} X_{2ki} + u_{2i}, \text{ pour les répondants en face à face} \quad (3)$$

avec Y_{1i} et Y_{2i} le nombre moyen de déplacements réalisé par l'individu i , X_{1i} et X_{2i} deux vecteurs de variables indépendantes ou explicatives de la mobilité et u_{1i} et u_{2i} deux termes d'erreur, qui tiennent compte des facteurs non observés qui pourraient influencer sur la mesure des résultats. Nous estimons deux modèles, un sur le sous-échantillon des répondants à l'enquête en face à face, et un sur le sous-échantillon des répondants sur le web. Soit la fonction de sélection traduisant la probabilité pour un individu i de répondre sur le web :

$$I_i^* = \sum_m \delta_{mi} Z_{mi} + \varepsilon_i, \quad (4)$$

$$I_i = 1, \text{ si } I_i^* > 0,$$

$$I_i = 0, \text{ si } I_i^* \leq 0$$

avec I_i^* la variable de sélection, Z_i un vecteur de variables déterminantes du choix du web et ε les termes d'erreur supposés normaux. La mobilité observée pour tout individu i se définit comme :

$$Y_i = Y_{1i}, \text{ si } I_i = 1$$

$$Y_i = Y_{2i}, \text{ si } I_i = 0$$

Notons que seulement un des paramètres Y_{1i} et Y_{2i} est observé, selon que l'individu réponde en face à face ou sur le web. Contrairement à l'équation (1), le modèle ainsi défini n'impose pas que les coefficients des variables explicatives X_{1ki} et X_{2ki} d'une part et Z_{mi} d'autre part soient identiques pour les répondants web et les répondants en face à face. Par ailleurs, nous supposons que les termes d'erreur des équations (2), (3) et (4) (u_1 , u_2 , et ε) suivent une loi normale bivariable de moyennes nulles et de corrélations ρ_1 et ρ_2 . En effet, la mobilité n'est observable que si les individus répondent en face à face ou sur le web. Des perturbations aléatoires vont affecter simultanément les variables endogènes des équations de sélection et d'intérêt et les termes d'erreur des deux équations peuvent être corrélés :

$$(\varepsilon, u_1) \sim N(0, 0, \sigma_\varepsilon, \sigma_{u1}, \rho_1)$$

$$(\varepsilon, u_2) \sim N(0, 0, \sigma_\varepsilon, \sigma_{u2}, \rho_2)$$

Considérons tous les individus avec (X_i, Z_i) donné. Formellement, la régression de Y_i sur X_i dans l'échantillon tronqué est :

$$E(Y_{1i} | I_i = 1) = E(Y_{1i} | X_{1i}, Z_i, I_i = 1) = \sum_k \beta_{1k} X_{1ki} + E(u_{1i} | Z_i, I_i = 1) \quad (5)$$

$$E(Y_{2i} | I_i = 0) = E(Y_{2i} | X_{2i}, Z_i, I_i = 0) = \sum_k \beta_{2k} X_{2ki} + E(u_{2i} | Z_i, I_i = 0) \quad (6)$$

Nous imposons une normalisation sur la variance de ε . Soit $\sigma_\varepsilon = 1$. Sous l'hypothèse de normalité, nous pouvons écrire :

$$u_{1i} = \rho_1 \sigma_{u1} \varepsilon_i$$

$$u_{2i} = \rho_2 \sigma_{u2} \varepsilon_i$$

En remplaçant dans les expressions (5) et (6), nous obtenons :

$$E(Y_{1i} | I_i = 1) = \sum_k \beta_{1k} X_{1ki} + \rho_1 \sigma_{u1} E(\varepsilon_i | \varepsilon_i > -\sum_m \delta_{mi} Z_{mi}) \quad (7)$$

$$E(Y_{2i} | I_i = 0) = \sum_k \beta_{2k} X_{2ki} + \rho_2 \sigma_{u2} E(\varepsilon_i | \varepsilon_i \leq -\sum_m \delta_{mi} Z_{mi}) \quad (8)$$

La troncature sur ε entraîne donc une troncature sur Y_1 et Y_2 si respectivement u_1 et ε et u_2 et ε sont corrélés (ρ_1 et $\rho_2 \neq 0$). Par définition, nous avons (Greene, 2002) :

$$E(\varepsilon_i | \varepsilon_i > -\sum_m \delta_{mi} Z_{mi}) = \frac{\varphi\left(-\sum_m \delta_{mi} Z_{mi}\right)}{1 - \theta\left(-\sum_m \delta_{mi} Z_{mi}\right)} \quad (9)$$

$$E(\varepsilon_i | \varepsilon_i \leq -\sum_m \delta_{mi} Z_{mi}) = -\frac{\varphi\left(-\sum_m \delta_{mi} Z_{mi}\right)}{\theta\left(-\sum_m \delta_{mi} Z_{mi}\right)} \quad (10)$$

En remplaçant dans les expressions (7) et (8), nous obtenons :

$$E(Y_{1i} | I_i = 1) = \sum_k \beta_{1k} X_{1ki} + (\rho_1 \sigma_{u1}) \frac{\varphi\left(\sum_m \delta_{mi} Z_{mi}\right)}{\theta\left(\sum_m \delta_{mi} Z_{mi}\right)} = \sum_k \beta_{1k} X_{1ki} + \rho_1 \sigma_{u1} \lambda_{1i} \quad (11)$$

$$E(Y_{2i} | I_i = 0) = \sum_k \beta_{2k} X_{2ki} + (\rho_2 \sigma_{u2}) \frac{-\varphi\left(\sum_m \delta_{mi} Z_{mi}\right)}{1 - \theta\left(\sum_m \delta_{mi} Z_{mi}\right)} = \sum_k \beta_{2k} X_{2ki} + \rho_2 \sigma_{u2} \lambda_{2i} \quad (12)$$

Les fonctions ϕ et θ sont respectivement les fonctions de densité et de répartition de la loi normale. Les ratios λ_{1i}

et λ_{2i} sont appelés « inverse du ratio de Mills ». Le biais de sélection peut provenir de variables omises, qui sont corrélées à la probabilité de choisir le web comme média d'enquête et au nombre de déplacements saisis. Il correspond donc à un biais de valeur manquante. En effet, si on estime les expressions (2) et (3) par la méthode des moindres carrés ordinaires, on omet deux variables :

$$\frac{\varphi\left(\sum_m \delta_m Z_{mi}\right)}{\theta\left(\sum_m \delta_m Z_{mi}\right)} = \lambda_{1i} \quad \text{et} \quad \frac{-\varphi\left(\sum_m \delta_m Z_{mi}\right)}{1 - \theta\left(\sum_m \delta_m Z_{mi}\right)} = \lambda_{2i}$$

On peut s'attendre à ce que le modèle soit biaisé, puisque les estimations de β_{1k} et β_{2k} sont non convergentes. Il est d'ailleurs probable que l'ampleur, le signe et la significativité des coefficients diffèrent lorsqu'ils sont estimés par la méthode en deux étapes. Ces différences dépendent des coefficients $\rho_1 \sigma_{u1}$ et $\rho_2 \sigma_{u2}$ et des coefficients des variables concernées estimés dans le modèle de sélection (Hoffman and Link, 1984). Les variables λ_{1i} et λ_{2i} représentent respectivement pour chaque observation l'espérance conditionnelle des résidus ε_i à $I_i = 1$ et $I_i = 0$. Elles capturent également les valeurs espérées des équations d'intérêt, une fois passé l'effet de sélection. Ces variables sont d'une manière générale la principale source de biais des estimations des coefficients du modèle de régression. Il est maintenant possible d'esti-

mer les fonctions de mobilité en s'affranchissant du biais de sélection des individus, en incluant l'inverse du ratio de Mills dans les expressions (2) et (3). Soit le modèle :

$$Y_{1i} = \sum_k \beta_{1k} X_{1ki} + \rho_1 \sigma_{u1} \lambda_{1i} + e_{1i}, \text{ pour les répondants web} \quad (13)$$

$$Y_{2i} = \sum_k \beta_{2k} X_{2ki} + \rho_2 \sigma_{u2} \lambda_{2i} + e_{2i}, \text{ pour les répondants en face à face} \quad (14)$$

Les paramètres du modèle de sélection de l'échantillon peuvent être estimés par la méthode du maximum de vraisemblance. Cependant, la procédure d'estimation en deux étapes d'Heckman (1979) est davantage utilisée. La première étape consiste à estimer l'équation de sélection à l'aide d'un modèle Probit, pour obtenir des estimations des δ_i . Pour chaque observation sélectionnée, le modèle calcule la valeur λ_{1i} ou λ_{2i} (inverse du ratio de Mills). Dans une seconde étape, on estime les paramètres β_{1k} et $\rho_1 \sigma_{u1}$ par une régression des moindres carrés ordinaires de Y_{1i} sur les vecteurs X_{1i} et λ_{1i} et les paramètres β_{2k} et $\rho_2 \sigma_{u2}$ par une régression des moindres carrés ordinaires de Y_{2i} sur les vecteurs X_{2i} et λ_{2i} . Les équations du modèle de mobilité Y_{1i} et Y_{2i} contiennent donc non seulement le vecteur de variables explicatives (respectivement X_{1i} et X_{2i}), mais aussi une nouvelle variable construite ou inverse du ratio de Mills.