

DEUX EXEMPLES D'ESTIMATIONS REGIONALES PAR MODELISATION SUR « PETITS DOMAINES » : LES TRANSPORTS ET DÉPLACEMENTS ET LE HANDICAP EN FRANCE DE MÉTROPOLÉ

Josiane Le Guennec¹

Résumé. Deux enquêtes à périodicité décennale ont été réalisées sur l'ensemble du territoire français, sur des sujets intéressant fortement les institutions régionales ou départementales. L'enquête nationale sur les Transports et Déplacements (ENTD), réalisée en 2006, mesure la fréquence des déplacements et l'usage des divers moyens de transport par les habitants. L'enquête sur le handicap et la santé des ménages (HSM), réalisée en 2008, estime notamment les prévalences de handicap dans la population. Les lois de décentralisation ont conféré aux assemblées territoriales des responsabilités particulières dans ces deux domaines, suscitant des besoins d'information localisée. L'intersection des échantillons nationaux avec chacune des 22 régions ou chacun des 95 départements métropolitains est généralement de taille trop faible pour permettre une estimation directe à ces niveaux. Des modèles d'estimation par régression ont été testés pour fournir des informations régionales dans le premier cas, départementales dans le second.

Dans l'enquête Transport, le comportement de déplacement est directement modélisé au moyen de régressions utilisant une information présente également dans le recensement de population. Dans l'enquête Handicap-Santé, le plan de sondage en deux phases est mis à profit pour imputer des probabilités estimées de handicap dans l'échantillon complet de première phase, de façon à réduire la variance totale de sondage, sans l'éliminer totalement. Dans les deux cas, les modèles sont construits dans le nuage de points des individus, et utilisent la régression mixte à effet aléatoire.

1 Se déplacer : un phénomène déterminé

La collecte de l'enquête Transports s'est faite en deux visites auprès de chacun des ménages sélectionnés. Le questionnaire de première visite décrivait le ménage et son environnement, et recensait les déplacements contraints de chacun de ses membres âgés de 6 ans ou plus. En fin de première visite, un individu Kish était sélectionné au hasard pour répondre, une semaine plus tard, au questionnaire de deuxième visite. Un carnet de déplacements, à remplir pendant la semaine séparant les deux visites, lui était remis, et servait de support au renseignement de ce second questionnaire. Celui-ci consistait à relever tous les déplacements réalisés pendant un jour, pris au hasard, de cette semaine de référence.

Environ 18600 ménages ont répondu aux deux visites sur le territoire métropolitain. L'allocation initiale de l'échantillon, proportionnelle à la population par région, a été renforcée par des extensions locales en Ile de France et en Pays de la Loire. Mesuré sur des échantillons respectivement de 5400 et 2600 répondants, l'estimateur direct Horvitz-Thomson y était publiable. Dans trois autres régions² où des extensions plus modestes d'échantillon ont pu être financées, quelques indicateurs très synthétiques avaient également une précision acceptable. Partout ailleurs en métropole, les échantillons régionaux ne comprenaient que quelques centaines de ménages répondants.

L'objet de l'étude a été restreint à la mobilité locale un jour de semaine, du lundi au vendredi, motivée par les activités de la vie quotidienne. Ce champ est défini par les déplacements effectués

¹ INSEE-Bretagne, Rennes. josiane.le-guennec@insee.fr

² Bretagne, Languedoc-Roussillon, Midi-Pyrénées.

dans un rayon de 80 km autour du domicile de l'individu Kish interrogé, pendant la journée de référence de l'enquête. La population cible est constituée des personnes âgées de 6 ans ou plus au moment de l'enquête.

Les variables estimées ont été les suivantes :

- population mobile le jour de semaine de référence : effectif et pourcentage
- fréquence des déplacements quotidiens un jour de semaine :
 - nombre total de déplacements par personne (population totale au dénominateur)
 - nombre de déplacements par mode de transport : voiture particulière, transport en commun, marche, autre mode
 - nombre de déplacements par motif
- nombre de personnes s'étant rendues au travail (ou au lieu d'étude), par mode de transport
- distance parcourue pendant la journée au cours des déplacements :
 - total journalier par personne
 - distance moyenne d'un déplacement par mode de transport
- temps passé quotidiennement dans les déplacements :
 - total journalier par personne
 - durée moyenne d'un déplacement par mode de transport

Le mode de transport (voiture ou transport en commun) relevé pour les déplacements professionnels est celui effectivement utilisé le jour de référence, et non le mode de transport déclaré comme habituel par la personne interrogée en première visite.

La distance parcourue est la distance à vol d'oiseau entre les centres des communes (ou des arrondissements à Paris, Lyon, Marseille) d'origine et de destination. Par convention, un déplacement interne à une commune (à un arrondissement) a donc une distance nulle.

Les indicateurs calculés sont des effectifs, des proportions, des moyennes par personne basées sur la population totale, des ratios par déplacement. Le champ de la population mobile inclut tous les individus de 6 ans ou plus ayant effectué au moins un déplacement dans la journée de référence, quelle que soit la distance parcourue.

1.1 Les méthodes d'estimation testées

Notations :

U est la population de référence, de taille N

$U_d \subset U$ est une sous population de U , de taille N_d

s est l'échantillon sélectionné dans U , de taille n

$s_d = s \cap U_d$ est l'intersection entre le domaine U_d et l'échantillon s

Y est une variable d'intérêt de l'enquête

X_j est une variable auxiliaire, connue par ailleurs

$Y_U = \sum_{i \in U} Y_i$ est le total de Y dans la population, et $\bar{Y}_U = \frac{1}{N} \sum_{i \in U} Y_i$ sa moyenne

$Y_d = \sum_{i \in U_d} Y_i$ est le total de Y dans le domaine U_d , et $\bar{Y}_d = \frac{1}{N_d} \sum_{i \in U_d} Y_i$ sa moyenne dans U_d

$X_j = \sum_{i \in U} X_{j,i}$ est le total de X_j dans la population, $\bar{X}_j = \frac{1}{N} \sum_{i \in U} X_{j,i}$ sa moyenne

$X_{jd} = \sum_{i \in U_d} X_{j,i}$ est le total de X_j dans le domaine U_d , $\bar{X}_{j,d} = \frac{1}{N_d} \sum_{i \in U_d} X_{j,i}$ sa moyenne

$\mathbf{X} = \begin{bmatrix} X_1 \\ \dots \\ X_j \\ \dots \\ X_J \end{bmatrix}$ est le vecteur des totaux dans U , de J variables auxiliaires X_j

\mathbf{X}_d est le vecteur des totaux dans U_d des J variables auxiliaires X_j

$\pi_i = \text{Prob}\{i \in s\}$ est la probabilité d'inclusion de l'unité i dans l'échantillon s

Une méthode d'estimation directe et deux méthodes d'estimation indirectes ont été testées.

L'estimateur direct par calage sur des totaux régionaux

L'estimateur calé sur des données régionales est asymptotiquement égal à l'estimateur redressé par régression. Il a pour expression :

$$\hat{Y}_{d,cal} \approx \hat{Y}_{d,reg} = \hat{Y}_{d,HT} + \hat{\beta}_d' (\mathbf{X}_d - \hat{\mathbf{X}}_{d,HT}) \quad (1)$$

où \mathbf{X}_d est le vecteur des totaux de J variables auxiliaires, dans U_d , $\hat{\mathbf{X}}_{d,HT}$ son estimateur Horvitz-Thomson, et $\hat{\beta}_d$ le vecteur des coefficients de la régression linéaire de Y sur X à l'intérieur du domaine, estimé dans le sous-échantillon s_d .

$$\hat{\beta}_d = (\mathbf{X}_d' \mathbf{D} \mathbf{X}_d)^{-1} \mathbf{X}_d' \mathbf{D} \mathbf{Y}_d \text{ avec : } \mathbf{D} = \text{diag} \left(\frac{1}{\pi_i} \right)$$

\mathbf{X}_d est la matrice (n_d, J) des observations des variables X_j sur les unités de l'échantillon appartenant au domaine U_d .

La variance d'un tel estimateur est une expression en $1/n_d$: sa précision reste tributaire de la taille de l'échantillon dans le domaine.

L'estimateur synthétique de type régression

Il est fondé sur un modèle de super-population. La fréquence des déplacements dans un ménage est conditionnée par des facteurs observables, et ces déterminants sont les mêmes sur l'ensemble du territoire.

Hypothèse : $Y_i = b' X_i$ pour toute unité i de l'univers

Relation estimée $Y_i = \hat{b}' X_i + \hat{u}_i$ dans l'échantillon national s

Estimation par domaine $\hat{Y}_D = \hat{b}' X_D$ où $X_D = \sum_{i \in U_D} X_i$

Dans le cas de l'analyse des déplacements, cette hypothèse ne peut être retenue sans précaution, en raison de la spécificité de l'Ile de France. Les modèles ont donc été estimés de façon séparée en Ile de France et dans l'ensemble des autres régions.

$$\hat{Y}_{d,regsyn} = \hat{\mathbf{b}}' \mathbf{X}_d \quad (2)$$

où : $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{D}\mathbf{X})^{-1} \mathbf{X}'\mathbf{D}\mathbf{Y}$ avec : $\mathbf{D} = \text{diag}\left(\frac{1}{\pi_i}\right)$

\mathbf{X} est ici la matrice (n,J) des observations des variables X_j sur toutes les unités de l'échantillon s . Contrairement à l'estimateur direct (1), le coefficient de régression $\hat{\mathbf{b}}$ est estimé dans l'échantillon national s , ce qui le rend plus robuste. En contrepartie, on crée un biais si l'hypothèse sous-jacente d'identité de comportement entre le domaine et l'ensemble de la population n'est pas vérifiée.

Un estimateur indirect avec modélisation explicite

Le modèle a pour forme générale :

$$Y_{d,i} = \boldsymbol{\beta}' \mathbf{X}_i + v_d + e_{d,i}$$

et l'estimateur régional :

$$\hat{Y}_d = \sum_{i \in U_d} (\hat{\boldsymbol{\beta}}' \mathbf{x}_i + \hat{v}_d) = \hat{\boldsymbol{\beta}}' \mathbf{X}_d + \hat{v}_d N_d \quad (3)$$

De façon classique, \mathbf{X}_i est un vecteur de variables auxiliaires observées sur l'unité i appartenant au domaine d , corrélées à la variable numérique continue à estimer Y , dont elles « expliquent » les variations.

Le coefficient régional v_d résulte d'une décomposition du résidu ε_i du modèle de régression classique : $Y_{d,i} = \boldsymbol{\beta}' \mathbf{x}_i + \varepsilon_i$, en une somme de deux éléments : $\varepsilon_{d,i} = \boldsymbol{\gamma}' \mathbf{t}_i + e_i = v_d + e_i$, où $\mathbf{t}_i = [t_{i \in U_d}]$, $d \in \{1, \dots, D\}$, est un vecteur dont les coordonnées sont les indicatrices d'appartenance de l'individu i à chacun des domaines U_d , et $\boldsymbol{\gamma}$ un vecteur de coefficients inconnus à estimer. Cette formulation permet de prendre en compte une éventuelle hétéroscédasticité du modèle selon la région. La variance totale du résidu ε_i est dissociée en une variance associée au domaine d et une variance résiduelle de e_i . Le modèle repose sur les hypothèses :

$$E(\varepsilon_i) = 0$$

$$\text{Var}(\varepsilon_i) = \sigma_\gamma^2 + \sigma_e^2$$

$$\text{Cov}(\varepsilon_{d,i}, \varepsilon_{d,j}) = \sigma_\gamma^2$$

$$\text{Cov}(\varepsilon_{d_1,i}, \varepsilon_{d_2,j}) = 0$$

Il s'agit d'un modèle mixte, combinant un effet aléatoire, l'appartenance au domaine, ici à la région, et des effets déterministes, les variables explicatives X_j .

Le modèle (3) a été estimé par la méthode du maximum de vraisemblance, avec la procédure MIXED du logiciel SAS, dans laquelle le code région est spécifié en effet aléatoire, et les variables de calage en effets déterministes.

Modélisation des ménages

| Variable expliquée | Régions de province | | Ile de France | |
|--|---|------|---|------|
| | Effets significatifs | R2 | Effets significatifs | R2 |
| Nombre total de déplacements par ménage le jour de référence | Zone de résidence *Age*CS des individus Équipement des ménages en voitures Sexe des individus Type de ménage CS du chef de ménage | 0,46 | Sexe*Age*CS des individus Équipement des ménages en voitures Type de ménage CS du chef de ménage | 0,48 |
| Nombre de déplacements pour le travail ou les études | CS des individus*Age Sexe Équipement des ménages en voitures Zone de résidence | 0,40 | CS des individus Équipement des ménages en voitures Sexe Type de ménage | 0,42 |
| Nombre de déplacements en transport en commun par ménage | Sexe*Age Zone de résidence *CS des individus*type de ménage Équipement des ménages en voitures CS du chef de ménage | 0,21 | Sexe*Age Équipement des ménages en voitures *CS du chef de ménage*Type de ménage Zone de résidence*CS des individus | 0,21 |
| Nombre de déplacements en voiture par ménage | Équipement des ménages en voitures Age Zone de résidence *CS des individus*Type de ménage Sexe CS du chef de ménage | 0,37 | Équipement des ménages en voitures Age Zone de résidence *CS des individus*Type de ménage CS du chef de ménage | 0,29 |
| Distance totale parcourue par ménage le jour de référence | Zone de résidence *Age*CS des individus Équipement des ménages en voitures Sexe CS du chef de ménage | 0,23 | Sexe CS du chef de ménage*CS des individus Zone de résidence *Age*Nbre de voitures du ménage | 0,32 |
| Temps total passé dans les déplacements par ménage | Zone de résidence *Age*CS des individus Sexe CS du chef de ménage | 0,33 | Age*CS du chef de ménage*Nbre de voitures du ménage Zone de résidence*CS des individus | 0,39 |

1.2 Le modèle de comportement

Les 22 régions de métropole constituent les domaines sur lesquels portent l'étude. Mettre en oeuvre des régressions sur des données agrégées à ce niveau, par une méthode de type Fay-Herriot, n'était donc pas envisageable. Afin de prendre en compte simultanément les caractéristiques de l'individu et du ménage, c'est le comportement des ménages qui a été modélisé.

Y_k = nombre de déplacements par ménage → variable expliquée.

L'information sur des niveaux géographiques fins nous est apportée par les recensements de population. Cette propriété contraint le choix des variables explicatives, qui doivent être identiquement définies dans l'enquête servant de base au modèle, et dans le recensement, source des totaux des variables auxiliaires dans la population complète. Cette condition a conduit à ne retenir dans le modèle que des variables présentes dans le RP avec une définition homogène à celle de l'ENTD.

L'âge, le sexe, le type d'activité (en emploi, étudiant, retraité), le milieu social, le type de ménage, l'équipement du ménage en moyens de transport personnels, combinés au lieu de résidence, sont les facteurs qui conditionnent le plus la mobilité des individus. Les variables explicatives retenues ont donc été celles-là. Le lieu de résidence est décrit au moyen d'une typologie croisant la taille de l'unité urbaine englobant la commune et un indicateur de distance entre le domicile et le centre de l'agglomération.

$$X_k = \begin{bmatrix} \text{zone de résidence} \\ \text{type de ménage} \\ \text{nombre de voitures du ménage} \\ \text{CS du chef de ménage} \\ \text{nombre de personnes du ménage par âge * sexe} \\ \text{nombre de personnes du ménage par CS} \end{bmatrix}$$

La taille des sous-échantillons régionaux obligeait à limiter le nombre de variables auxiliaires utilisables dans des calages par région, alors que la modélisation directe dans l'échantillon national permet de s'affranchir de telles contraintes. Les variables explicatives ont donc été ventilées dans une nomenclature plus détaillée pour les modèles (2) et (3) que dans le modèle (1). Afin de distinguer l'effet du modèle et celui de la méthode d'estimation sur le résultat, les estimateurs (2) et (3) ont été calculés successivement avec la nomenclature du modèle (1) et la nomenclature la plus détaillée disponible (annexes 1 et 2).

La variable expliquée est numérique, le vecteur des variables explicatives mêle variables numériques et variables catégorielles. La technique de régression utilisée est donc l'analyse de la variance, avec la procédure GLM du logiciel SAS pour le modèle (2), avec la procédure MIXED pour le modèle (3). L'échantillon ayant été préalablement redressé pour non-réponse, les observations ont été pondérées par les poids de sondage. Les relations obtenues sont résumées dans le tableau ci-contre.

1.2 Les résultats

En l'absence d'estimateur analytique sans biais de l'erreur quadratique moyenne dans les modèles (1) et (2), apprécier la validité des résultats obtenus relève de moyens empiriques.

La variance de l'estimateur direct fondé sur les pondérations nationales a été estimée, ce qui permet de comparer les estimateurs modélisés aux bornes des intervalles de confiance à 95 % des estimations directes. Un estimateur « petit domaine » non couvert par l'intervalle de confiance de l'estimateur direct, par construction important, pourrait difficilement être retenu. Une telle situation est l'indice d'un biais important créé par le modèle, non compensé par la réduction de la variance de sondage.

Les 5 régions ayant réalisé des extensions d'échantillon ont constitué des régions-tests. L'écart entre l'estimateur direct, à variance faible dans ces régions, et l'estimateur modélisé, approxime le biais éventuellement généré par le modèle.

Les coefficients de régression des estimateurs (2) et (3) ont été estimés successivement sur l'ensemble de l'échantillon national, et sur l'échantillon hors Ile de France. Ces deux versions donnent des résultats différents avec l'estimateur synthétique (2) et les nomenclatures agrégées du modèle, principalement pour estimer l'usage des transports en commun ou de la voiture, et le temps consacré aux déplacements un jour de semaine. C'est là que l'Ile de France se singularise le plus. Ils se confondent pour les autres indicateurs : nombre de déplacements par motif, distance parcourue, déplacements à pied ou en vélo.

Les nomenclatures les plus détaillées des variables auxiliaires comprennent une modalité spécifique à l'Ile de France. Avec ce modèle, les deux versions de l'estimateur synthétique sont très convergentes. L'effet régional aléatoire du modèle mixte (3) prend en charge la spécificité francilienne, de sorte que régression nationale et régression hors Ile de France donnent ici des résultats presque identiques.

Les estimateurs modélisés ont pour effet de lisser les moyennes régionales, en les rapprochant de la moyenne nationale. L'estimateur synthétique de type régression (2) apparaît plus souvent biaisé, et de façon plus importante, que l'estimateur par modèle mixte (3). Cette dernière méthode donne des résultats en général plus proches de l'estimateur direct que l'estimateur synthétique de type régression. Dans les régions avec extension, l'estimateur (3) reste très généralement à l'intérieur des bornes de l'intervalle de confiance à 95 % de l'estimateur direct. Seule exception : la fréquence des déplacements en transport en commun, variable qui résiste le plus à la modélisation testée. En Midi-Pyrénées et en Languedoc-Roussillon, l'estimateur par modèle mixte reste en-dehors de l'intervalle de confiance de l'estimateur direct. Il en est de même dans 4 autres régions sans extension.

Nombre total de déplacements par jour et par personne

| Régions avec extension d'échantillon | estimateur direct borne inf. de l'IC | estimateur direct (poids nationaux) | estimateur direct borne sup. de l'IC | estimateur synthétique par régression : $Y_r = b * X_r$ b France | estimateur synthétique par régression : $Y_r = a * X_r$ a province | estimateur par modèle mixte effet régional aléatoire |
|--------------------------------------|--------------------------------------|-------------------------------------|--------------------------------------|---|---|--|
| Ile de France | 2.93 | 3.00 | 3.07 | 3.03 | . | 3.00 |
| Pays de la Loire | 3.39 | 3.53 | 3.66 | 3.17 | 3.18 | 3.38 |
| Bretagne | 3.36 | 3.51 | 3.67 | 3.13 | 3.14 | 3.38 |
| Midi-Pyrénées | 3.06 | 3.17 | 3.28 | 3.15 | 3.14 | 3.08 |
| Languedoc-Roussillon | 3.05 | 3.20 | 3.35 | 3.15 | 3.15 | 3.20 |

Déplacements professionnels : fréquence journalière par habitant

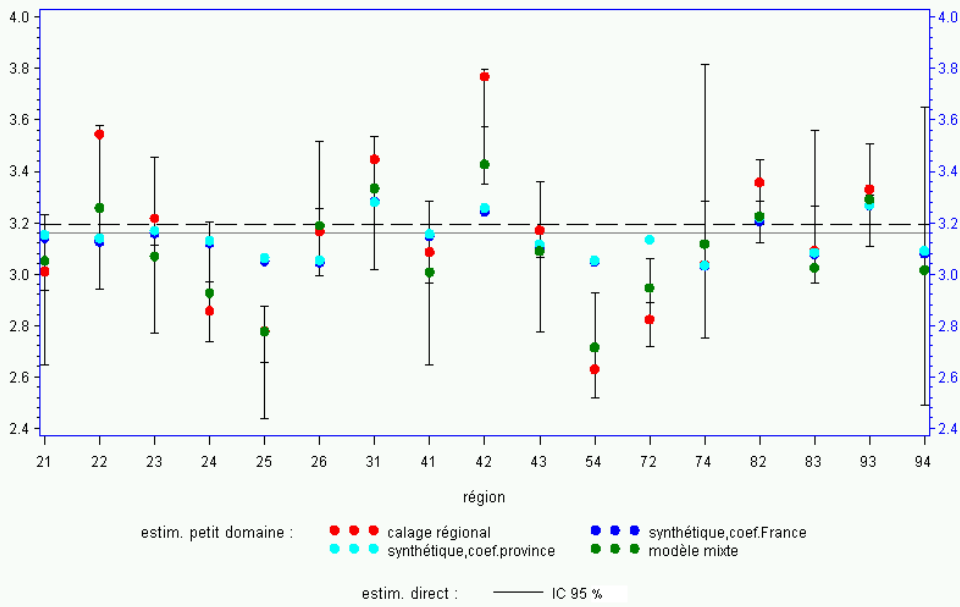
| Régions avec extension d'échantillon | estimateur direct borne inf. de l'IC | estimateur direct (poids nationaux) | estimateur direct borne sup. de l'IC | estimateur synthétique par régression : $Y_r = b * X_r$ b France | estimateur synthétique par régression : $Y_r = a * X_r$ a province | estimateur par modèle mixte effet régional aléatoire |
|--------------------------------------|--------------------------------------|-------------------------------------|--------------------------------------|---|---|--|
| Ile de France | 0.81 | 0.84 | 0.86 | 0.86 | . | 0.86 |
| Pays de la Loire | 0.95 | 1.02 | 1.10 | 0.89 | 0.90 | 0.94 |
| Bretagne | 0.77 | 0.87 | 0.97 | 0.86 | 0.87 | 0.87 |
| Midi-Pyrénées | 0.89 | 0.95 | 1.01 | 0.87 | 0.87 | 0.87 |
| Languedoc-Roussillon | 0.70 | 0.76 | 0.81 | 0.78 | 0.78 | 0.78 |

Temps passé dans les déplacements dans la journée par habitant

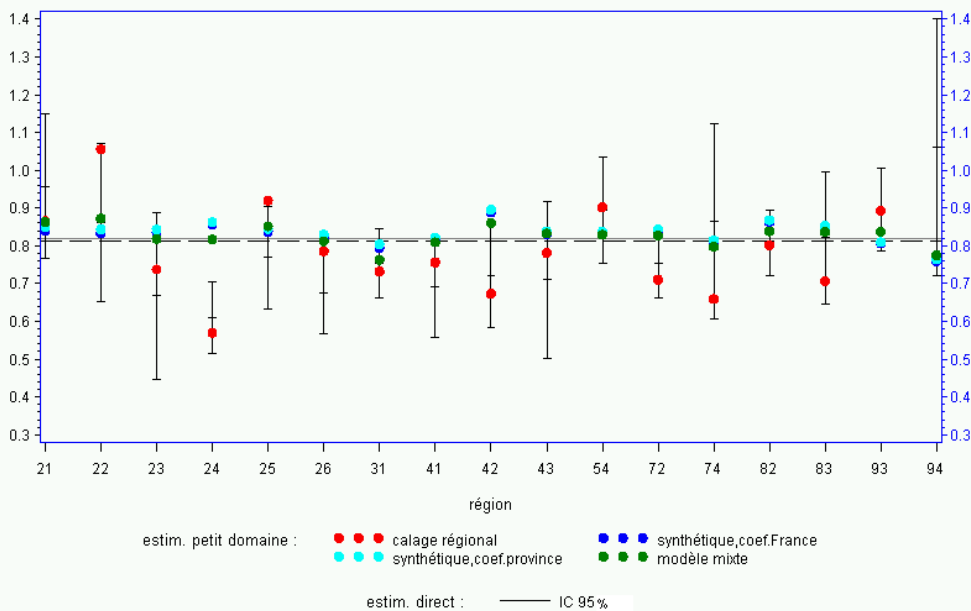
minutes

| Régions avec extension d'échantillon | estimateur direct borne inf. de l'IC | estimateur direct (poids nationaux) | estimateur direct borne sup. de l'IC | estimateur synthétique par régression : $Y_r = b * X_r$ b France | estimateur synthétique par régression : $Y_r = a * X_r$ a province | estimateur par modèle mixte effet régional aléatoire |
|--------------------------------------|--------------------------------------|-------------------------------------|--------------------------------------|---|---|--|
| Ile de France | 70 | 72 | 74 | 71 | . | 71 |
| Pays de la Loire | 52 | 55 | 57 | 54 | 53 | 54 |
| Bretagne | 51 | 55 | 58 | 54 | 53 | 55 |
| Midi-Pyrénées | 51 | 53 | 55 | 54 | 54 | 54 |
| Languedoc-Roussillon | 51 | 53 | 56 | 53 | 53 | 54 |

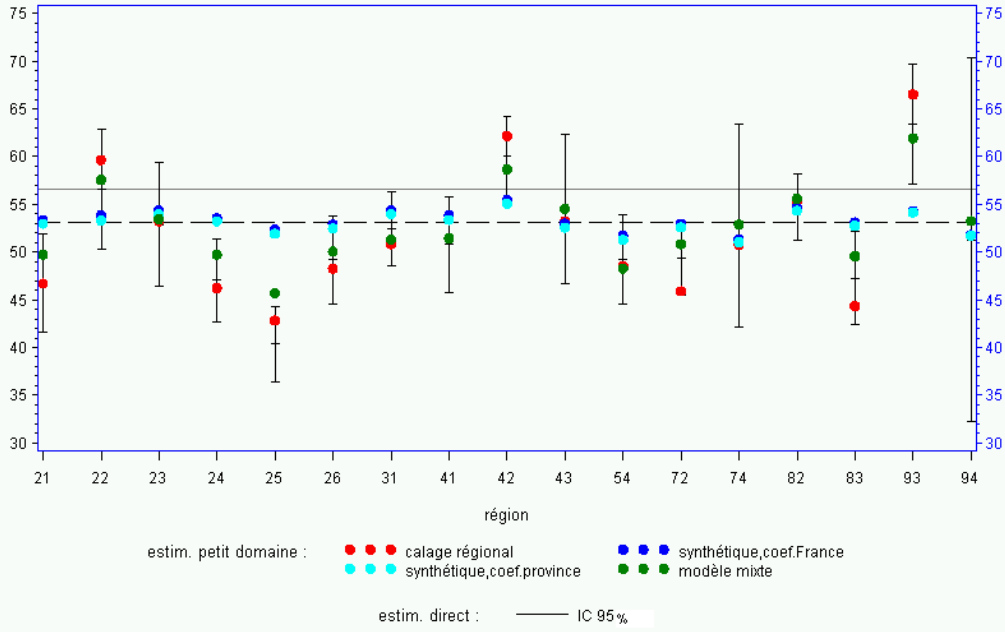
Nombre total de déplacements par jour et par personne



Déplacements professionnels : fréquence journalière par habitant



Temps passé dans les déplacements dans la journée par habitant



Annexe 1 : modèle simplifié du calage régional, nomenclature des variables de calage

Variables de calage de niveau ménage :

☒ zone de résidence

- 1=communes rurales en espace rural,
- 2=communes rurales en espace urbain,
- 3=communes péri-urbaines d'unités urbaines de moins de 100000 habitants,
- 4=villes-centres d'agglomérations de moins de 100000 habitants,
- 5=communes péri-urbaines d'agglomérations de 100000 habitants ou plus,
- 6=villes-centres d'unités urbaines de 100000 habitants ou plus

Mise en forme : Pucés et numéros

☒ type de ménage

- personne seule ou ménage sans famille,
- famille monoparentale,
- couple sans enfants, couple avec enfants

Mise en forme : Pucés et numéros

☒ nombre de voitures du ménage

- aucune,
- 1 voiture,
- 2 voitures ou plus

Mise en forme : Pucés et numéros

☒ catégorie socioprofessionnelle de la personne de référence

- professions indépendantes incluant les agriculteurs,
- cadres et professions intermédiaires,
- ouvriers et employés,
- inactifs

Mise en forme : Pucés et numéros

Variables de calage de niveau individu :

☒ Sexe

☒ Âge

- 6-24 ans,
- 25-34 ans, 35-49 ans,
- 50-64 ans,
- 65 ans et plus

Mise en forme : Pucés et numéros

☒ catégorie socioprofessionnelle de l'individu, les retraités étant reclassés dans leur profession d'origine

- agriculteurs et professions indépendantes en activité ou retraités,
- cadres et professions intermédiaires actifs,
- ouvriers et employés actifs,
- anciens cadres et professions intermédiaires,
- anciens ouvriers et employés,
- inactifs non retraités

Mise en forme : Pucés et numéros

☒ Nombre de voitures du ménage auquel appartient l'individu

- aucune,
- 1 voiture,
- 2 voitures ou plus

Mise en forme : Pucés et numéros

Annexe 2 : modèle détaillé, nomenclature des variables explicatives

Variables explicatives de niveau ménage :

☐ Classe de commune

- 01 rurale dans un espace urbain
- 02 rurale dans un espace à dominante rurale
- 11 agglomération de moins de 20000 habitants - ville centre
- 12 agglomération de moins de 20000 habitants - banlieue
- 41 agglomération de 20000 à 50000 habitants - ville centre
- 42 agglomération de 20000 à 50000 habitants - banlieue
- 51 agglomération de 50000 à 100000 habitants - ville centre
- 52 agglomération de 50000 à 100000 habitants - banlieue
- 61 agglomération de 100000 à 200000 habitants - ville centre
- 62 agglomération de 100000 à 200000 habitants - banlieue
- 71 agglomération de 200000 à 2 M d'habitants - ville centre
- 72 agglomération de 200000 à 2 M d'habitants - banlieue
- 81 Paris
- 82 Banlieue parisienne

Mise en forme : Puces et numéros

☐ type de ménage

- personne seule,
- famille monoparentale,
- couple sans enfants,
- couple avec enfants,
- ménage sans famille

Mise en forme : Puces et numéros

☐ nombre de voitures du ménage

- aucune,
- 1 voiture,
- 2 voitures ou plus

Mise en forme : Puces et numéros

☐ catégorie socioprofessionnelle de la personne de référence

- agriculteurs, artisans, commerçants, chefs d'entreprise
- cadres
- professions intermédiaires
- ouvriers
- employés
- retraités
- autres inactifs

Mise en forme : Puces et numéros

Variables explicatives de niveau individu :

☐ hommes par tranche d'âge :

- 6-13 ans,
- 14-18 ans
- 19-25 ans,
- 26-29 ans,
- 30-39 ans,
- 40-49 ans,
- 50-60 ans,

Mise en forme : Puces et numéros

- 61-69 ans,
- 70-79 ans
- 80 ans et plus

☐ femmes selon les mêmes tranches d'âge

☐ catégorie socioprofessionnelle de l'individu, les retraités étant reclassés dans leur profession d'origine :

- agriculteurs, actifs et retraités
- artisans, commerçants, chefs d'entreprise en activité ou retraités,
- cadres et professions intermédiaires actifs,
- ouvriers et employés actifs,
- anciens cadres et professions intermédiaires,
- anciens ouvriers et employés,
- élèves, étudiants
- inactifs autres

☐ Nombre de voitures du ménage auquel appartient l'individu

- aucune,
- 1 voiture,
- 2 voitures ou plus

Mise en forme : Puces et numéros

Mise en forme : Puces et numéros

2 Handicap : du ressenti à l'estimation du certain

L'INSEE a réalisé en 2007-2008 une enquête auprès des ménages sur la santé et le handicap, visant notamment à mesurer les prévalences des maladies et des situations d'incapacité dans la population. Cette enquête comprenait plusieurs volets. Le premier, dit « enquête handicap-santé des ménages » (HSM), s'intéressait à la population résidant à son domicile, c'est-à-dire aux « ménages ordinaires ». Le second volet concernait les personnes hébergées en institutions, telles les maisons de retraite et établissements hospitaliers spécialisés. L'objet de cette étude est l'estimation des situations de handicap dans la population des ménages ordinaires, par région et département. Elle repose exclusivement sur le volet HSM de l'enquête.

2.1 Le dispositif de l'enquête HSM 2008

L'enquête HSM s'est faite en deux phases.

La première phase, appelée « Vie quotidienne et santé » (VQS), a pour but de repérer les personnes susceptibles de souffrir d'un handicap, afin de construire une base de sondage de la population cible. Celle-ci constituant une population rare, il fallait en première phase interroger un très grand nombre de personnes pour obtenir un échantillon de taille suffisante en deuxième phase. L'échantillon VQS est donc très important : plus de 218000 personnes ont répondu à cette enquête en métropole.

La collecte a été principalement réalisée par voie postale, puis par téléphone auprès des non-répondants, et enfin par enquêteur auprès des ménages sans numéro connu de téléphone. Il s'agit d'un questionnaire auto-administré, constitué de questions simples sur les difficultés physiques ressenties dans les actes de la vie quotidienne. Les réponses peuvent être binaires, oui/non, ou comprendre trois modalités de difficulté ressentie : aucune, un peu, beaucoup. Le ménage était l'unité échantillonnée, mais tous les individus du ménage devaient répondre au questionnaire, qui se présente sous forme d'une feuille recto-verso, avec une colonne à renseigner par membre du ménage. Le plan de sondage est celui habituel aux enquêtes ménages de l'INSEE, à plusieurs degrés, approximativement auto-pondéré, la base de sondage étant recherchée dans la dernière enquête annuelle de recensement disponible, c'est-à-dire celle de 2006.

La deuxième phase de l'enquête a été réalisée en face à face, auprès d'un échantillon d'individus sélectionnés parmi les répondants en première phase. Les réponses recueillies au questionnaire VQS ont servi à stratifier cet échantillon en 4 groupes de sévérité croissante des difficultés de santé déclarées. Le tirage est réalisé par sondage aléatoire simple dans les strates, avec des taux de sondage croissants du groupe 1 (personnes présumées en bonne santé) au groupe 4 (personnes présumées les plus en difficulté). Le sondage était exhaustif dans le groupe 4. Environ 29900 personnes ont répondu, dont 26400 en métropole.

Quatre départements métropolitains et deux départements d'outre-mer ont réalisé une extension d'échantillon par rapport à l'allocation nationale initiale. En métropole, il s'agit du Nord, du Pas-de-Calais, du Rhône et des Hauts de Seine, où l'estimateur direct Horvitz-Thomson constituait une valeur de référence.

2.2 Les variables estimées

Les domaines concernés par l'étude sont les 95 départements de métropole.

Les estimations localisées ont porté sur les prévalences, dans la population adulte de 20 ans ou plus, des handicaps par type : sensoriels, moteurs, cognitifs, ainsi que sur celle des personnes âgées dépendantes. Dans le premier cas, les estimations ont été menées successivement dans la population de 20 à 59 ans et dans celle âgée de 60 ans ou plus. La dépendance n'est analysée que dans la population de 60 ans ou plus.

L'échantillon national HSM comprenait 12602 personnes âgées de 20 à 59 ans en 2008 et 9165 personnes de 60 ans ou plus.

Selon la définition retenue par l'Organisation mondiale de la Santé, **le handicap** résulte de l'interaction entre un problème de santé, des facteurs environnementaux et des facteurs personnels, entraînant pour la personne des limitations dans ses activités et des restrictions dans sa participation à la vie collective. La définition du handicap adoptée dans cette étude n'en retient que les dimensions physiologiques et fonctionnelles, non sa dimension sociale. Le handicap est ici défini par la présence simultanée d'une déficience d'un organe et d'une limitation fonctionnelle connexe sévère. Une personne est potentiellement handicapée lorsqu'elle souffre d'une altération physique et ne peut réaliser la fonction normalement assurée par l'organe déficient. Les formes de handicap sont regroupées en trois grands types génériques : sensoriels, moteurs, cognitifs.

La perte d'autonomie due au vieillissement est évaluée au moyen d'un outil, la grille AGGIR, qui classe les personnes âgées dans six groupes iso-ressources (GIR) selon leurs facultés physiques et mentale. La classification de 1 à 6 constitue une échelle d'autonomie croissante. Dans l'enquête HSM, un algorithme utilisant l'ensemble des réponses au questionnaire a permis d'évaluer à quel groupe GIR pouvait être rattachée chaque personne interrogée âgée de 60 ans ou plus au moment de l'enquête. Le classement individuel obtenu résulte des réponses à un questionnaire statistique. Il pourrait donc s'écarter du diagnostic qu'aurait effectué une équipe de professionnels avec le protocole adapté d'un examen médical. Les effectifs de personnes âgées appartenant à chacun des groupes de la grille AGGIR sont donc ici doublement estimatifs : en raison de l'aléa propre au sondage, mais aussi d'un aléa sur l'observation individuelle. Les estimations départementales et régionales ont porté sur la prévalence des personnes âgées de 60 ans ou plus appartenant à l'un des **GIR 1 à 4**, de façon indifférenciée.

2.3 La méthode d'estimation sur « petits domaines »

Le nombre limité de questions et la simplicité de leur formulation dans le questionnaire VQS ne permettent pas d'identifier de façon certaine, en première phase, les personnes considérées comme « handicapées » ou dépendantes selon les grilles d'analyse de l'Organisation mondiale de la Santé. Cette mesure constitue l'objet de la deuxième phase HSM de l'enquête, grâce à un questionnaire très détaillé.

Néanmoins, il semble légitime d'attendre une forte corrélation entre la situation ressentie par la personne et déclarée dans l'enquête VQS, et celle finalement décrite et « diagnostiquée » en deuxième phase. La méthode d'estimation dans des « petits domaines » adoptée ici repose sur cette corrélation.

Le principe

La méthode retenue utilise les informations recueillies en première phase VQS de l'enquête Handicap-Santé pour imputer, à l'aide d'un modèle, des valeurs du paramètre aux individus non réinterrogés en deuxième phase. L'extrapolation à la population totale du domaine se fait alors à l'aide des pondérations de l'échantillon de première phase. L'estimateur final supporte une variance de sondage réduite par rapport à l'estimateur direct, la taille de l'échantillon de première phase étant beaucoup plus importante, mais la possibilité d'un biais dû au modèle, si l'hypothèse sous-jacente d'homogénéité de comportement sur le territoire n'est pas suffisamment pertinente.

L'effectif total dans la population U des personnes affectées d'une maladie ou d'un handicap peut se décomposer entre l'échantillon S_2 enquêté en deuxième phase et la population non enquêtée :

$$Y_U = \sum_{i \in U} Y_i = \sum_{k \in S_2} Y_k + \sum_{j \notin S_2} Y_j$$

Si tout l'échantillon de première phase était réenquêté en deuxième phase, Y pourrait être estimé par l'estimateur Horvitz Thomson muni des pondérations de l'échantillon VQS S_1 :

$$\hat{Y}_{VQS,HT} = \sum_{i \in S_1} w_{1i} y_i = \sum_{k \in S_2} w_{1k} y_k + \sum_{j \in S_1 \cap \bar{S}_2} w_{1j} y_j$$

Le total départemental Y_d de la variable Y est estimé en substituant, dans cette expression, aux valeurs réelles inconnues, les probabilités prédites par un modèle construit dans l'échantillon HSM de deuxième phase. Deux formes alternatives ont été calculées, dont la convergence constitue un test de bonne adéquation du modèle.

La première utilise les prédictions dans tout l'échantillon de première phase :

$$\hat{Y}_{1d} = \frac{1}{\hat{N}_d} \sum_{i \in U_d \cap S_1} w_{1i} \tilde{y}_i \quad (A)$$

La seconde conserve les valeurs observées dans l'échantillon de deuxième phase auxquelles elle ajoute les valeurs imputées par le modèle dans la partie non réenquêtée de l'échantillon de première phase :

$$\hat{Y}_{2d} = \frac{1}{\hat{N}_d} \left(\sum_{k \in U_d \cap S_2} w_{1k} y_k + \sum_{j \in U_d \cap S_1 \cap \bar{S}_2} w_{1j} \tilde{y}_j \right) \quad (B)$$

où : U_d est la population du département d

$Y_k = 1 \Leftrightarrow$ l'individu k présente le handicap étudié

$Y_k = 0$ sinon.

Pour tout individu i répondant à VQS :

$\tilde{y}_i = \text{Pr} \circ b\{Y_i = 1\}$ est la probabilité d'être handicapé prédite par le modèle

w_{1i} est le poids redressé de l'individu i dans l'enquête de 1^{ère} phase VQS

$\hat{N}_d = \sum_{i \in U_d \cap S_1} w_{1i}$ est la taille estimée de la population dans le département d .

De nouvelles pondérations de l'échantillon VQS ont été calculées par calage sur les populations départementales au 1^{er} janvier 2008, ventilées par sexe et âge croisés en quatre tranches : moins de 19 ans, 20 à 59 ans, 60 à 74 ans, 75 ans et plus.

Le modèle adopté est un modèle de régression logistique mixte, dans lequel le département de résidence est spécifié en effet aléatoire.

Pour tout individu i résidant dans le département d , on postule une relation de la forme :

$$\text{Prob}\{Y_{d,i} = 1\} = p_{d,i} = \frac{1}{1 + \exp[-(\mathbf{b}'\mathbf{x}_i + v_d)]}.$$

\mathbf{x}_i est un vecteur de variables disponibles dans les deux échantillons \mathbf{s}_1 et \mathbf{s}_2 , observées dans l'enquête VQS ou présentes dans la base de sondage

\mathbf{b} est le vecteur des coefficients de régression entre Y et les facteurs fixes X

v_d est un coefficient qui permet de prendre en compte une composante explicative du handicap spécifique au département d où réside l'individu i .

Le coefficient v_d résulte d'une décomposition du résidu ε_i du modèle de régression logistique classique :

$$z_{d,i} = \mathbf{b}'\mathbf{x}_i + \varepsilon_i,$$

où
$$z_{d,i} = \text{Log}\left(\frac{p_{d,i}}{1 - p_{d,i}}\right),$$

en une somme de deux éléments :

$$\varepsilon_{d,i} = v_d + u_i$$

où v_d est un coefficient inconnu à estimer.

Cette formulation permet d'introduire, dans l'ensemble des effets inobservés non pris en compte par les effets fixes du modèle, l'influence éventuelle des différences de politiques locales, en la traitant comme un effet d'espérance nulle et de variance σ_v^2 , conditionnellement aux cofacteurs X . La variance totale du résidu ε_i est dissociée en une variance associée au domaine d et une variance résiduelle u_i .

Les coefficients \mathbf{b} et v_d du modèle sont estimés dans l'ensemble de l'échantillon national de deuxième phase HSM, hors départements d'outre-mer.

En construisant l'estimateur « petit domaine », on substitue à la phase de sondage HSM, donnant des résultats très instables dans de petites sous populations, l'imputation de valeurs individuelles à variance réduite, grâce à des coefficients de régression robustes, puisque estimés dans un très grand échantillon.

Rappelons par ailleurs que la variance totale de l'estimateur direct Horvitz-Thomson dans l'enquête HSM est celle d'un sondage en deux phases, par conséquent de la forme :

$$\text{Var}(\hat{Y}_{HT}) = \text{Var}[E(\hat{Y}_{HT} / \mathbf{s}_1)] + E[\text{Var}(\hat{Y}_{HT} / \mathbf{s}_1)].$$

C'est la somme de deux éléments : le premier exprime la variance due à l'échantillonnage de première phase, tandis que le second représente la variance due à l'échantillonnage de deuxième phase, conditionnellement à l'échantillon de phase 1.

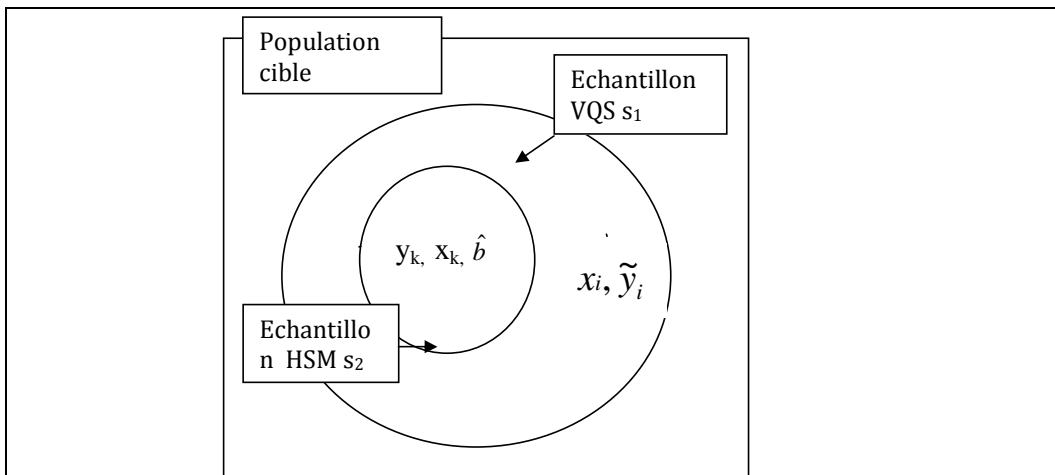
En imputant une valeur de la variable d'intérêt à chaque individu de l'échantillon de phase 1, on supprime toute variance de sondage propre à l'échantillonnage HSM conditionnel à l'échantillon VQS. Une grande partie de l'amélioration de la précision d'un estimateur départemental ou régional repose ici sur cette propriété.

Avec ses deux expressions alternatives (A) et (B), l'estimateur modélisé reste néanmoins dépendant de l'échantillonnage de première phase. Outre l'aléa créé par le modèle, il conserve une variance de

sondage incompressible, égale à celle de l'estimateur Horvitz-Thomson d'un total ou d'une moyenne dans l'échantillon VQS.

La réduction attendue de la variance, grâce à la modélisation, sera donc d'autant plus importante, que la contribution de la deuxième phase HSM à la variance totale de l'estimateur direct est plus élevée. Celle-ci n'est pas uniforme sur le territoire, et varie selon le paramètre estimé. Le gain de précision ne peut, par conséquent, être de même ampleur dans toutes les zones pour tous les indicateurs analysés.

Schéma d'estimation



*Lecture : y_k est la valeur observée pour l'individu k enquêté en deuxième phase dans HSM
 x_i est un vecteur de variables auxiliaires connues pour chaque individu i de l'échantillon
 VQS s_1 , par conséquent aussi pour chaque individu k de l'échantillon HSM s_2
 \hat{b} est le vecteur des coefficients de régression d'une relation entre y_k et x_k estimée dans
 l'échantillon de deuxième phase s_2
 \tilde{y}_i est la valeur imputée par le modèle à l'individu i répondant en première phase
 VQS mais non enquêté en deuxième phase.*

Les facteurs fixes du modèle

La recherche des facteurs explicatifs a été effectuée avec la procédure GLMSELECT du logiciel SAS (version 9.2), au moyen d'une procédure itérative de type « stepwise », et du critère de sélection bayésien de Schwartz (SBC). Les observations sont pondérées par leur poids de sondage redressé dans l'enquête HSM, la procédure se chargeant de normer automatiquement les pondérations.

Les coefficients de régression sont ensuite estimés par modèle logistique mixte, dans lequel le niveau géographique d'estimation est déclaré en effet aléatoire. Un effet local résiduel, non pris en compte par les facteurs explicatifs retenus, est donc pris en charge par le modèle. La régression est réalisée au moyen de la procédure GLIMMIX du logiciel SAS, pondérée par le poids de sondage

redressé de première phase.

Des départements ont été regroupés pour obtenir des domaines comprenant au moins 1000 personnes répondantes dans l'échantillon de première phase, sous la contrainte du respect des limites administratives régionales. La typologie départementale établie par un groupe de travail de la DREES [8] a été utilisée pour réunir les départements les plus proches sur le plan sanitaire, à l'intérieur de chaque région.

Les réponses à l'enquête de première phase ont constitué les premiers facteurs explicatifs des variables à estimer. Le questionnaire VQS comprenait quatre groupes de questions :

- difficultés sensorielles (voir de près, de loin, entendre)
- difficultés de mobilité (marcher, se baisser, attraper en hauteur, se servir de ses mains)
- difficultés intellectuelles (mémoriser, comprendre, communiquer, prendre des initiatives, être autonome)
- besoins d'aide technique (prothèses..) ou humaine (aidant régulier) dans la vie quotidienne.

Parmi les critères sociodémographiques disponibles dans la base de sondage et dans l'enquête de première phase :

- l'âge au-delà de 60 ans
- le sexe au-delà de 60 ans

sont les seules caractéristiques personnelles qui sont apparues influentes.

2.4 Les résultats

Les deux modes de calcul de l'estimateur "petit domaine" donnent des résultats quasiment identiques. Le nuage de points formé des départements ayant en coordonnées les estimations (A) et (B) se confond avec la bissectrice des axes, pour tous les indicateurs.

France entière, la somme des valeurs prédites ne s'écarte pas de plus de 4 % de l'estimateur direct national, quel que soit le paramètre mesuré, et est recouverte par l'intervalle de confiance à 95 % de l'estimateur direct.

L'écart relatif entre estimation directe et prédiction du modèle reste compris entre -10 % et +10 % dans la moitié au moins des régions, entre -20 % et +20 % dans 17 des 22 régions. La différence absolue entre la proportion estimée par l'enquête et celle prédite par le modèle reste inférieure à 1 dans au moins la moitié des régions pour tous les indicateurs calculés, à 2 dans 80 % des régions.

Outre la Corse, les régions dans lesquelles le modèle réduit ou accroît l'estimation directe de plus de 20 % sont les petites régions, ayant par conséquent les plus faibles tailles d'échantillon, qui obèrent fortement la précision de l'estimateur direct.

Dans trois des quatre départements témoins ayant réalisé une extension d'échantillon, l'écart relatif entre l'estimation directe et la valeur prédite ne dépasse pas 11 %. L'adéquation du modèle est plus problématique dans le département des Hauts de Seine, où l'on constate les écarts relatifs les plus importants par rapport à l'enquête. On peut imputer cette singularité au plan de sondage de deuxième phase de l'extension des Hauts de Seine, qui s'est différenciée du plan national par un sous échantillonnage dans la strate 1 au profit d'une saturation de la strate 3, conditionnellement à l'échantillon VQS de première phase. Il en résulte une surpondération des individus de la strate 1, source de perturbation lorsque certains d'entre eux s'avèrent être handicapés d'après l'enquête HSM, malgré une probabilité nulle selon leurs réponses en première phase.

La prédiction du modèle diffère de l'estimation directe, en hausse ou en baisse, de moins de 10 % dans la moitié des départements. Dans un quart des départements, cet écart relatif dépasse 20 % en

hausse ou en baisse.

Le nuage des départements ayant en coordonnées la prévalence estimée directement dans l'enquête HSM et celle prédite par le modèle s'étire linéairement le long de la bissectrice des axes. La hiérarchie des départements selon l'indicateur étudié est à peu près respectée par l'estimateur modélisé. La pente de la droite de régression de la valeur prédite sur l'estimation directe se situe entre 0,79 et 0,87 selon la variable estimée.

Comme souvent dans les estimations de type « petits domaines », on observe un phénomène de « shrinkage ». La méthode tend à réduire la dispersion entre « domaines », par rapport aux estimateurs Horvitz-Thomson, en rapprochant de la moyenne nationale une partie des régions ou départements situés aux extrêmes de la distribution initiale issue de l'exploitation directe de l'enquête HSM.

Comme il a été évoqué plus haut (section 2.3), l'estimateur « petit domaine » conserve une variance de sondage due à l'échantillonnage de première phase, à laquelle s'ajoute également un aléa de sondage du coefficient de régression, lui-même estimé dans un échantillon. Le gain de précision de ce nouvel estimateur repose sur l'importance de la variance de seconde phase dans la variance totale de l'estimateur direct. Ce critère a guidé le choix de diffusion finale des résultats. Ont été publiées sans restriction les estimations obtenues dans les départements où le coefficient de variation de l'estimateur direct, dû à la phase VQS, ne dépassait pas 20 %. Lorsque ce coefficient est compris entre 20 % et 30 %, l'estimateur modélisé est publié avec un avertissement. Au-delà de 30 %, le résultat n'est pas diffusé.

Prévalences en France de métropole : estimations directes et valeurs prédites

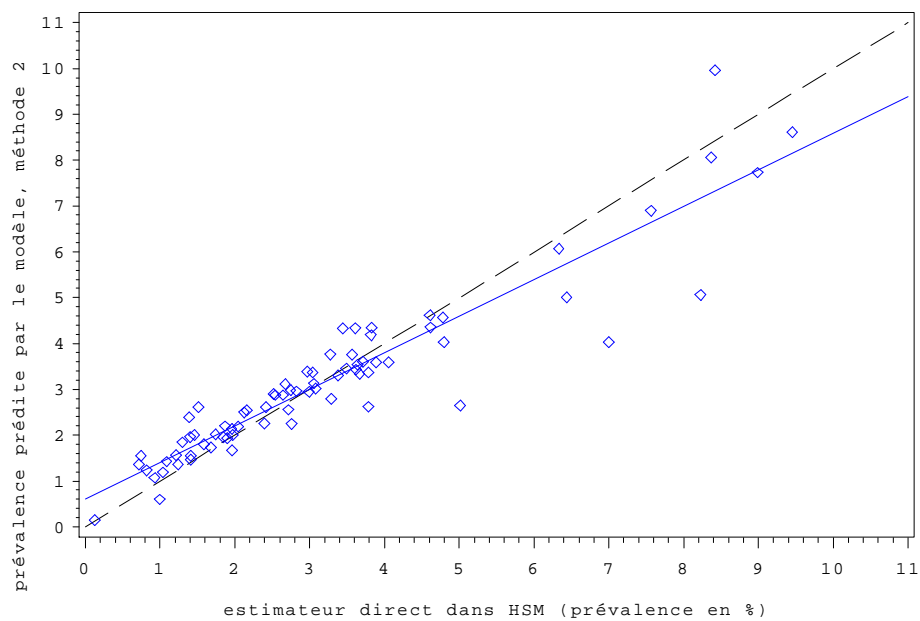
| variable | Estimateur direct | | | Estimateur prédit par le modèle | | | |
|-------------------------------------|---------------------------------|----------------|---------------------------------|--|--|----------------------|---------------------------|
| | borne inf. interv. conf. à 95 % | Prévalence (0) | borne sup. interv. conf. à 95 % | Prévalence prédite par le modèle méthode 1 (1) | Prévalence prédite par le modèle méthode 2 (2) | Ecart absolu (2)-(0) | Ecart relatif (%) (2)/(0) |
| Population de 20 à 59 ans | | | | | | | |
| limitations motrices | 4.1 | 4.4 | 4.8 | 4.3 | 4.3 | -0.1 | -3.2 |
| limitations sensorielles | 2.7 | 3.1 | 3.5 | 3.0 | 3.1 | 0.0 | -0.2 |
| troubles cognitifs | 4.0 | 4.5 | 5.0 | 4.4 | 4.5 | 0.0 | 0.2 |
| Population de 60 ans ou plus | | | | | | | |
| limitations motrices | 23.2 | 24.4 | 25.6 | 24.7 | 24.8 | 0.4 | 1.6 |
| limitations sensorielles | 12.5 | 13.5 | 14.4 | 13.3 | 13.3 | -0.1 | -1.1 |
| troubles cognitifs | 8.3 | 9.1 | 9.9 | 9.2 | 9.4 | 0.3 | 3.1 |
| GIR 1 à 4 | 6.0 | 6.5 | 7.1 | 6.8 | 6.8 | 0.3 | 4.2 |

Estimation dans les départements avec extension d'échantillon

| | Nord | | Pas de Calais | | Rhône | | Hauts de Seine | |
|-------------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| | Estimateur direct | Estimateur prédit | Estimateur direct | Estimateur prédit | Estimateur direct | Estimateur prédit | Estimateur direct | Estimateur prédit |
| Population de 20 à 59 ans | | | | | | | | |
| limitations sensorielles | 3,6 | 3,4 | 2,8 | 3,0 | 2,7 | 3,0 | 1,6 | 1,8 |
| limitations motrices | 4,8 | 4,9 | 6,3 | 5,9 | 4,3 | 4,3 | 2,3 | 2,6 |
| limitations cognitives | 6,6 | 6,6 | 4,8 | 4,9 | 4,8 | 4,5 | 2,9 | 2,8 |
| Population de 60 ans et plus | | | | | | | | |
| limitations sensorielles | 15,5 | 15,7 | 14,4 | 15,0 | 14,2 | 13,6 | 10,2 | 10,8 |
| limitations motrices | 32,2 | 32,7 | 29,9 | 31,1 | 23,8 | 23,1 | 21,0 | 21,8 |
| limitations cognitives | 9,9 | 11,0 | 8,1 | 8,6 | 9,2 | 9,1 | 7,0 | 7,8 |
| GIR 1 à 4 | 6,8 | 7,5 | 8,5 | 8,8 | 5,6 | 5,6 | 4,4 | 5,3 |

Limitations sensorielles dans la population de 20 à 59 ans

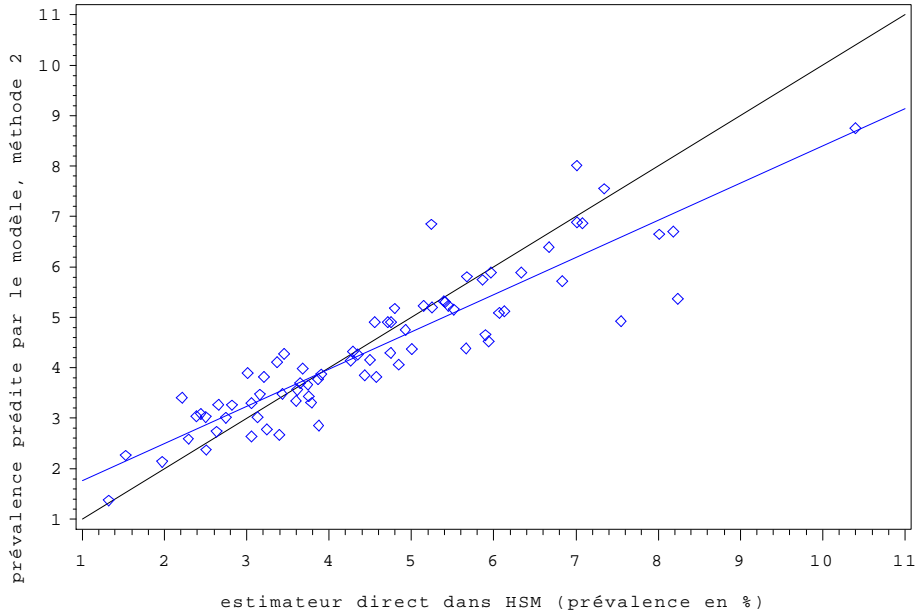
Prévalences départementales prédites (méthode 2) et directes



Equation de régression : prevalence prédite = 0.60 + 0.80 * estimation directe

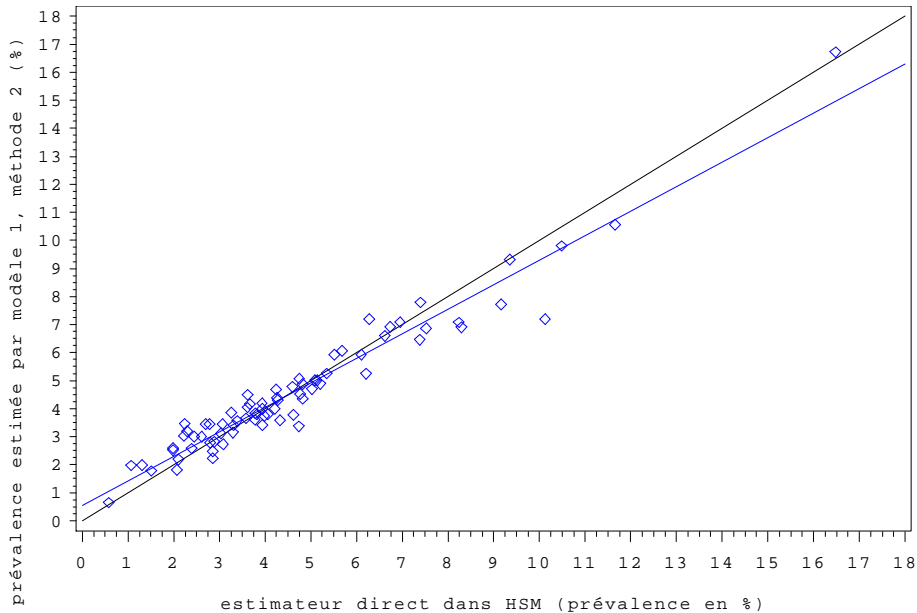
Limitations motrices dans la population de 20 à 59 ans

Prévalences départementales prédites (méthode 2) et directes

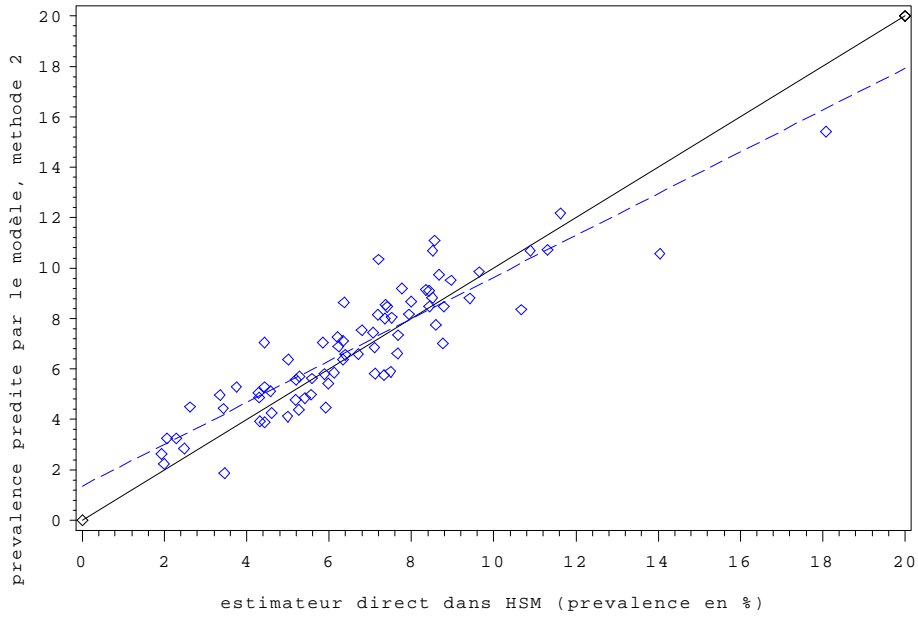


Limitations cognitives dans la population de 20 à 59 ans

Prévalences départementales prédites (méthode 2) et directes



Population de 60 ans ou plus appartenant aux GIR 1 à 4
Prévalences départementales prédites (méthode 2) et directes



Equation de régression : prévalence prédite = 1.36 + 0.83 * estimation directe

Annexe 1 - Les modèles de régression logistique

| Paramètre à estimer | Champ d'observation | Effectif tel que Y=1 dans l'échantillon HSM | Effets fixes explicatifs |
|----------------------------|-------------------------------------|---|---|
| Limitations sensorielles | 20 à 59 ans | 1006 | v_auditif v_vueloin v_vuepres v_impossible v_reconnai v_age |
| Limitations sensorielles | 60 ans et plus | 2003 | v_auditif v_vueloin v_vuepres v_impossible v_aide technique sexe * âge ³ |
| Limitations motrices | 20 à 59 ans | 2247 | v_marcher * ramasser ⁴ v_main v_attraper v_limitation d'activité * aide technique ⁵ v_handicap ressenti * reconnaissance de handicap ⁶ sexe * âge ⁷ |
| Limitations motrices | 60 ans et plus | 4158 | v_marcher * ramasser v_attraper v_limitation d'activité * aide technique v_handicap ressenti sexe * âge |
| Troubles cognitifs | 20 à 59 ans | 1717 | v_initiative v_reconnaissance de handicap v_concentration v_aidepers- un peu v_aide régulière par une personne extérieure v_memoire v_limitation d'activité v_problèmes_quotidiens v_comprendre |
| Troubles cognitifs | 60 ans et plus | 1653 | v_concentration v_impossible v_initiative v_memoire v_problèmes_quotidiens v_aide régulière par une personne extérieure sexe * âge ⁸ |
| Appartenance aux GIR 1 à 4 | 60 ans et plus Strates VQS 1 à 3 | 519 | v_sorties v_aide régulière par une personne extérieure v_limitation d'activité v_initiative v_amenagements du logement v_aides techniques v_main âge |
| Appartenance aux GIR 1 à 4 | 60 ans et plus Strate VQS 4 | 1029 | v_sorties v_aide régulière par une personne extérieure v_main v_concentration v_reconnaissance de handicap âge ⁹ |

³ Fonction affine sur les segments : 60-69 ans, 70 ans et plus.

⁴ interaction : marche * ramasse. 1=pas de difficulté, 2=quelques difficultés, 3=beaucoup de difficultés

⁵ Interaction : limitation d'activité*aide technique (appareillage, prothèse...). Modalités aide technique : 1=non, 2=oui.

⁶ Interaction : se sentir handicapé * avoir une reconnaissance administrative de handicap

⁷ Fonction affine sur les segments : 20-39 ans, 40 ans et plus.

⁸ Fonction affine sur les segments : 60-69 ans, 70 ans et plus.

⁹ Fonction affine sur les segments : 60-84 ans, 85 ans et plus.

Liste des variables

| | |
|------------|---|
| Limitat | limitations d'activité ressenties |
| Aidepers | aide régulière d'un tiers pour les actes de la vie quotidienne |
| Aidetech | utilisation d'un appareillage, de prothèses ou autre aide technique |
| Amenalog | Aménagements du logement pour des raisons de santé |
| Marche | difficultés à marcher plus de 500 mètres ou à monter un escalier |
| Ramasse | difficulté à ramasser un objet par terre |
| Attrape | difficulté à attraper un objet en hauteur |
| Main | difficulté à se servir de ses mains |
| Auditif | difficulté à entendre ce qui se dit dans une conversation entre plusieurs personnes |
| Vueloin | difficulté à voir de loin |
| Vuepres | difficulté à voir de près |
| Comprendre | difficulté à comprendre les autres ou à se faire comprendre des autres |
| Initiative | difficulté à prendre des initiatives |
| Concentra | difficulté à se concentrer |
| Prob_quot | difficulté à résoudre les problèmes de la vie quotidienne |
| Memoire | difficulté à se souvenir |
| Impossible | impossibilité totale à réaliser l'une des activités essentielles citées (voir, entendre, parler, bouger) |
| Sorties | difficulté à sortir seul de chez soi |
| Handicap | handicap ressenti (non/oui) |
| Reconnai | reconnaissance d'un handicap par l'administration |

Annexe 2 - Le questionnaire de l'enquête de première phase VQS

Pour chaque personne du foyer on pose la série de questions suivante.

Prénom

1 : sexe de la personne

2 : année de naissance

3 : *Comment est l'état de santé général de la personne?*

Très bon / bon / moyen / mauvais / très mauvais

4 : *La personne a-t-elle actuellement une ou plusieurs maladie(s) chronique(s) ou problème(s) de santé durable(s) ?*

Non / Oui

5 : *La personne est-elle limitée dans les activités qu'elle peut faire en raison d'un problème de santé ou d'un handicap(à la maison, au travail, à l'école...)?*

Non / Oui, un peu / Oui, beaucoup

6 : *La personne a-t-elle des difficultés pour voir les caractères ordinaires d'un article de journal (avec ses lunettes ou ses lentilles si elle en porte habituellement) ?*

Non / Oui, un peu / Oui, beaucoup

7 : *A-t-elle des difficultés pour voir nettement le visage de quelqu'un à l'autre bout de la pièce (avec ses lunettes ou ses lentilles si elle en porte habituellement) ?*

Non / Oui, un peu / Oui, beaucoup

8 : *A-t-elle des difficultés pour parler ?*

Non / Oui, un peu / Oui, beaucoup

9 : *A-t-elle des difficultés pour entendre ce qui se dit au cours d'une conversation avec plusieurs personnes ?*

Non / Oui, un peu / Oui, beaucoup

10 : *A-t-elle des difficultés pour monter un étage d'escalier ou marcher 500 mètres ?*

Non / Oui, un peu / Oui, beaucoup

11 : *A-t-elle des difficultés pour lever le bras (par exemple pour attraper un objet en hauteur)?*

Non / Oui, un peu / Oui, beaucoup

12 : *A-t-elle des difficultés pour se servir de ses mains ou de ses doigts (par exemple pour ouvrir une bouteille d'eau) ?*

Non / Oui, un peu / Oui, beaucoup

13 : *Lorsqu'elle est debout, a-t-elle des difficultés pour se pencher et ramasser un objet?*

Non / Oui, un peu / Oui, beaucoup

14 : *A-t-elle des difficultés pour se concentrer plus de 10 minutes ?*

Non / Oui, un peu / Oui, beaucoup

15 : *A-t-elle des difficultés pour se souvenir de choses importantes ?*

Non / Oui, un peu / Oui, beaucoup

16 : *A-t-elle des difficultés pour prendre des initiatives dans la vie quotidienne?*

Non / Oui, un peu / Oui, beaucoup

17 : *A-t-elle des difficultés pour résoudre les problèmes de la vie quotidienne (comme se repérer sur un itinéraire ou compter l'argent) ?*

Non / Oui, un peu / Oui, beaucoup

18 : *A-t-elle des difficultés pour sortir du domicile ?*

Non / Oui, un peu / Oui, beaucoup

19 : *A-t-elle des difficultés pour comprendre les autres ou se faire comprendre des autres (en dehors des difficultés liées aux différences de langue)?*

Non / Oui, un peu / Oui, beaucoup

20 : *La personne est-elle dans l'impossibilité totale d'accomplir une ou plusieurs des activités citées précédemment (questions 6 à 19). Par exemple une personne aveugle ne peut absolument pas voir les caractères ordinaires d'un article de journal.*

Non / Oui

21 : *En raison d'un problème de santé ou d'un handicap, reçoit-elle de l'aide d'une autre personne dans la vie quotidienne ?*

Non / Oui, un peu / Oui, beaucoup

22 : *En raison d'un problème de santé ou d'un handicap, des aménagements du logement ont-ils été réalisés pour cette personne?*

Non / Oui

23 : *En raison d'un problème de santé ou d'un handicap, utilise-t-elle régulièrement une prothèse, un appareillage ou une aide technique ?*

Non / Oui

24 : *La personne considère-t-elle qu'elle a un handicap ?*

Non / Oui

25 : *La personne a-t-elle une reconnaissance officielle d'un handicap ou d'une perte d'autonomie (allocation, pension ou carte d'invalidité, admission dans un établissement spécialisé...)?*

Non / Oui

26 : *Si c'est un enfant d'âge scolaire: Est-il inscrit dans une classe ou un établissement spécialisé en raison de problèmes de santé ou de difficultés d'apprentissage ?*

Non / Oui

Bibliographie

- [1] Ardilly, P. (2006). Panorama des principales méthodes d'estimation sur petits domaines : aspects théoriques.- Rennes-ENSAI, photocopié.
- [2] Bouvier,G. (2011). Le volet Ménages de l'enquête Handicap-Santé, présentation, calcul des poids. Paris, INSEE, Document de travail n°F1101.
- [3] Bouvier,G. (2011). L'enquête Handicap-Santé, présentation générale. Paris, INSEE, Document de travail n°F1109.
- [4] Bouvier,G. (2009). L'approche du handicap par les limitations fonctionnelles et la restriction globale d'activité chez les adultes de 20 à 59 ans. France, portrait social, INSEE.
- [5] Clark A.E., Vicard, A. (2007). Conditions de collecte et santé auto-déclarée : une analyse sur données européennes. Economie et Statistique n°403-404, INSEE.
- [6] Couet, C. (2002). Estimations locales sur les personnes handicapées vivant en domicile ordinaire, enquête HID 1999, résultats détaillés. Paris, INSEE-Résultats Société n° 12-2002.
- [7] Davezies, L. (2011). Modèles à effets fixes, à effets aléatoires, modèles mixtes ou multi-niveaux : propriétés et mises en œuvre des modélisations de l'hétérogénéité dans le cas de données groupées. Paris, INSEE, Document de travail n°G2011/03.
- [8] Gauthier, P., Mansuy M. (2009). Rapport du groupe de travail « indicateurs sociaux départementaux », Paris, CNIS.
- [9] Hidirolou, M.A. et Patak, Z. (2004). Estimation par domaine par la régression linéaire, Toronto, Techniques d'enquête, volume 30 n°1, pp.73-85.
- [10] Hubert, J.P. (2009). Dans les grandes agglomérations, la mobilité quotidienne des habitants diminue, et elle augmente ailleurs, Paris, INSEE-Première n°1252.
- [11] Lavergne, C. et Trottier, C. (2000). Sur l'estimation dans les modèles linéaires généralisés à effets aléatoires, Paris, Revue de statistique appliquée, XLVIII, pp.49-67.
- [12] Le Guennec, J. (2012). Estimations locales du handicap dans l'enquête Handicap-Santé 2008, Paris, INSEE, Document de travail n°H2012/02.
- [13] Le Guennec, J. (2012). Estimations régionales dans l'enquête nationale Transports et Déplacements 2007-2008, INSEE, Document de travail n°H2012/03.
- [14] Midy, L. (2010). Présentation de l'enquête « Vie quotidienne et Santé » de 2007. Paris, INSEE, Document de travail n°F1001.
- [15] Midy, L. (2009). Enquête « Vie quotidienne et Santé » : limitations dans les activités et sentiment de handicap ne vont pas forcément de pair. Paris, INSEE-Première n°1254.
- [16] Moussalam, K. (2010). Approximation de la variance de l'enquête HSM 2008. Note interne non publiée, Paris, INSEE.

[17] Panorama issu de l'enquête nationale transports et déplacements 2008, revue du CGDD, décembre 2010.

[18] Techniques d'enquête, volume 29, n°1, juin 2003 : numéro spécial de la revue de Statistique Canada consacré à l'estimation sur petits domaines.