

TRAITEMENT DE LA NON-REPONSE NON-IGNORABLE PAR CALAGE GENERALISE: UNE SIMULATION A PARTIR DE L'ENQUETE BUDGET DES MENAGES AU LUXEMBOURG

Guillaume OSIER (*)

(*) *Institut National de la Statistique et des Etudes Economiques du Grand-duché de Luxembourg (STATEC), Division des Statistiques Sociales*

Introduction

La non-réponse totale¹ est devenue aujourd'hui un sujet de préoccupation majeure pour tous les gestionnaires d'enquête, dans un climat où la participation des ménages aux enquêtes tend à se dégrader depuis deux décennies [1]. Si l'on cherche à augmenter la participation des ménages par des mesures préventives comme, par exemple, l'envoi de lettres avant la collecte, l'utilisation de modes de collecte adaptés à l'enquête, une meilleure formation des enquêteurs ou des rappels systématiques des ménages non répondants, il faut bien se rendre à l'évidence qu'en dépit de tous les efforts déployés pour tenter de réduire la non-réponse, celle-ci ne pourra jamais être totalement éradiquée au stade de la collecte des données. Une des raisons à cela est que les contraintes de temps et de coût qui pèsent généralement sur une collecte ne donnent pas les moyens suffisants pour atteindre et de convaincre tous les ménages sélectionnés dans l'échantillon de participer à l'enquête.

Si on ne peut pas empêcher la non-réponse, on va alors chercher à la traiter *ex post* en haussant les poids d'échantillonnage des unités répondantes de manière à éliminer le biais de non-réponse, causé par des différences entre les caractéristiques des unités répondantes et non répondantes. L'approche va consister à estimer explicitement les probabilités de réponse des unités répondantes (sur la base par exemple d'un modèle logistique) pour ensuite ajuster le poids de ces unités en le multipliant par l'inverse de leur probabilité de réponse [1, 2, 13]. Le poids ainsi obtenu conduit à un estimateur qui, sous l'hypothèse où les probabilités de réponse ont été estimées de manière exacte, est sans biais.

Cette approche nécessite pour l'estimation des probabilités de réponse de pouvoir disposer de variables explicatives dont les valeurs sont connues à la fois sur les unités répondantes et non répondantes. Cette contrainte est particulièrement restrictive dans la mesure où la disponibilité d'une telle information auxiliaire est souvent réduite à quelques variables sociodémographiques simples comme l'âge, le genre ou la taille du ménage. Lorsque la non-réponse dépend d'autres variables pas forcément disponibles sur les non répondants, voire lorsque la non-réponse est causée par le sujet même de l'enquête (mécanisme *non-ignorable*), on ne sera plus en mesure de construire un modèle permettant d'estimer convenablement les probabilités de réponse. D'autres approches seront alors nécessaires afin de traiter ces situations. Ce sera précisément l'objet de ce papier.

Dans ce papier on va envisager le recours aux méthodes de calage [1, 4, 13] pour corriger la non-réponse. On verra que la théorie généralisée du calage [1, 3, 7, 8, 11] permet de traiter les cas de non-réponse non-ignorable, ce que l'approche basée sur l'estimation explicite des probabilités de réponse ne permet pas. Après quelques rappels théoriques sur le calage, on présentera les résultats d'une simulation réalisée à partir des données de l'enquête Budget des Ménages au Luxembourg, et dans laquelle le calage généralisé a été testé comme moyen de traitement de la non-réponse non-ignorable. Les calculs ont été rendus possibles grâce à la nouvelle version de la macro Calmar (Calmar 2 – [8]) dans laquelle la technique du calage généralisé a été implantée. On verra que les résultats de cette simulation sont prometteurs, même si des problèmes persistent.

¹ Dans ce papier on se limitera à évoquer la non-réponse totale, c'est-à-dire l'absence totale de réponse au questionnaire. En outre on ne parlera pas de non-réponse partielle, dans laquelle l'information est manquante seulement pour certaines parties du questionnaire. Dans la suite on abrégera non-réponse totale en non-réponse.

1. Quelques rappels sur la théorie du calage

La philosophie de toutes les méthodes de calage est la même: il s'agit d'intégrer une information auxiliaire dans l'expression de l'estimateur afin d'en améliorer la précision. Historiquement, des formes particulières d'estimateurs calés ont été introduites selon la nature de l'information auxiliaire disponible : estimateur post-stratifié, estimateur par ratissage (*raking ratio*), estimateur par le quotient ou estimateur par la régression [1]. Un cadre général pour toutes les méthodes de calage, englobant ces cas particuliers, a ensuite été introduit [4].

1.1. Calage simple

Le cadre théorique développé dans [4] est le suivant : on dispose d'un échantillon s d'unités tirées dans une population U selon un mécanisme probabiliste donné. Chaque unité k de s est affectée d'un poids d'échantillonnage d_k . Cette pondération reflète le plan de sondage ainsi que le mécanisme de non-réponse. Elle est souvent calculée à partir du poids de sondage de l'unité, c'est-à-dire l'inverse de la probabilité de sélection de l'unité dans l'échantillon, que l'on ajuste pour la non-réponse en le divisant par la probabilité de réponse estimée de l'unité.

L'information auxiliaire se présente sous la forme d'un vecteur de p variables auxiliaires $x = (x^{(1)}, x^{(2)} \dots x^{(p)})$ tel que :

- (i) la valeur $x_k = (x_k^{(1)}, x_k^{(2)} \dots x_k^{(p)})$ de x est connue pour chaque unité k de l'échantillon s ,
- (ii) le total $X = (X^{(1)}, X^{(2)} \dots X^{(p)})$ sur la population U , où $X^{(l)} = \sum_{k \in U} x_k^{(l)}$, est connu avec une très bonne précision. Cette connaissance provient généralement de sources administratives (par exemple un registre de la population ou des données fiscales) ou d'enquêtes réalisées sur des échantillons de grande taille².

L'idée du calage simple consiste à modifier « à la marge » les pondérations initiales $d = \{d_k, k \in s\}$ de sorte que le vecteur de totaux X sur lequel on cale notre échantillon soit estimé de manière exacte à partir des nouvelles pondérations. Si D désigne une fonction « de distance », on cherche à calculer de nouvelles pondérations $\omega = \{\omega_k, k \in s\}$ qui soient le plus proche possible des pondérations initiales d (au sens de la « distance » D):

$$\omega = \text{Arg min}_{\rho = \{\rho_k, k \in s\}} \sum_{k \in s} D(\rho_k, d_k) \quad (1)$$

sous la contrainte que les totaux de calage sont estimés de manière exacte à partir des nouvelles pondérations :

$$\sum_{k \in s} \omega_k x_k = X \quad (2)$$

On peut montrer [13] que les poids calés s'écrivent pour tout k :

$$\omega_k = d_k F(x_k, \lambda) \quad (3)$$

où λ est un vecteur colonne (inconnu) de longueur p et F une fonction dont l'expression est déduite de la distance D (voir Tableau 1). En remplaçant dans (2) ω_k par (3) on obtient un système (généralement non linéaire) de p équations à p inconnues λ .

$$\sum_{k \in s} d_k F(x_k, \lambda) x_k = X \quad (4)$$

² L'INSEE utilise les structures de population provenant de l'enquête Emploi pour caler les échantillons d'autres enquêtes. L'échantillon trimestriel de l'enquête Emploi comprenant environ 100,000 individus âgés de 15 ans ou plus, le niveau de précision attendu sera très élevé.

On résout (4) par des algorithmes itératifs (Newton-Raphson) qui fournissent généralement une solution approchée pour λ . On déduit ensuite les poids calés à partir de (3)

Le calcul des poids calés est implémenté dans de nombreux logiciels spécialisés comme la macro SAS CALMAR (développée par l'INSEE – [14]), G-Calib (développé par Statistique Belgique) ou CLAN (développé par Statistique Suède)

Tableau 1 : les fonctions de distance dans CALMAR

$D(\rho_k, d_k)$	$F(x)$	NOM DE LA METHODE
$\frac{(\rho_k - d_k)^2}{d_k}$	$1 + x$	Linéaire
$\rho_k \text{Log}\left(\frac{\rho_k}{d_k}\right) + d_k - \rho_k$	$\exp(x)$	Raking Ratio
$(\rho_k - Ld_k) \text{Log} \frac{\frac{\rho_k}{d_k} - L}{1 - L}$ $+ (Ud_k - \rho_k) \text{Log} \frac{U - \frac{\rho_k}{d_k}}{U - 1}$ <p>Si $L \leq \frac{\rho_k}{d_k} \leq U$ $+\infty$ SINON</p>	$\frac{L(U-1) + U(1-L)\exp(Ax)}{U-1 + (1-L)\exp(Ax)}$ <p>avec : $A = \frac{U-L}{(1-L)(U-1)}$</p>	Logistique
$\frac{(\rho_k - d_k)^2}{d_k}$ <p>Si $L \leq \frac{\rho_k}{d_k} \leq U$ $+\infty$ SINON</p>	$1 + x$ si $L-1 \leq x \leq U-1$ L si $x < L-1$ U si $x > U-1$	Linéaire tronquée

1.2. Calage et correction de la non-réponse

A l'origine, les méthodes de calage ont été introduites comme un moyen de réduire la variance d'échantillonnage. Le résultat fondamental (voir [4]) sur la variance d'un estimateur calé dit que celle-ci dépend de la variance des résidus de la régression sur les variables de calage. Autrement dit, plus les variables de calage seront corrélées à la variable d'intérêt de l'enquête, plus la variance d'échantillonnage sera réduite.

Néanmoins on montre [1, 5, 11, 12, 13] que si les variables de calage sont également corrélées avec la probabilité de réponse alors le calage permettra de réduire aussi le biais de non-réponse. Comme on l'a évoqué, ce biais est causé par des différences entre les caractéristiques des unités répondantes et non répondantes. Concrètement on effectue un calage en une étape pour corriger à la fois le biais de non-réponse et réduire la variance d'échantillonnage. Si s désigne l'échantillon brut, de taille n , tiré dans la base de sondage et r l'échantillon des unités répondantes, de taille m ($m < n$), l'équation du calage en une étape s'écrit :

$$\sum_{k \in r} \frac{d_k}{\theta} F(x_k, \lambda) x_k = X \quad (5)$$

d_k est le poids de sondage de l'unité k , égal à l'inverse de sa probabilité de sélection, et $\bar{\theta} = \frac{m}{n}$ est le taux de réponse global calculé sur l'échantillon brut s . La division des poids de sondage par le taux $\bar{\theta}$ est une opération de simple normalisation. On peut montrer [1, 5] que le calage en une étape (5) est équivalent à la pondération en deux étapes suivantes :

1. division du poids de sondage d_k par : $r_k = \frac{m}{n} \frac{1}{F(x_k, \theta)}$ (6), où F désigne la fonction de calage utilisée dans (5) et θ un paramètre déterministe. Toute la validité du calage en une étape repose sur l'hypothèse d'une modélisation exacte des probabilités de réponse par (6),
2. calage des poids ajustés $\frac{d_k}{r_k}$ en utilisant les variables de calage x .

Pour corriger la non-réponse, le calage en une étape ne s'appuie que sur les valeurs x_i des variables sur les unités i répondantes. En outre, il n'y a pas besoin de connaître les valeurs sur les unités non répondantes, ce qui constitue un avantage significatif par rapport à l'approche basée sur l'estimation des probabilités de réponse (sur la base de variables auxiliaires dont les valeurs devaient être connues à la fois sur les unités répondantes et non répondantes). On pourra par exemple utiliser des variables actualisées, à condition que l'on connaisse le total de ces variables sur la population. Ce total pourra provenir d'enquêtes reposant sur des échantillons de grande taille (par exemple, l'enquête Emploi de l'INSEE). Par contre, si le total sur la population est inconnu alors le calage ne sera d'aucun secours. La connaissance du total sur la population constitue encore une contrainte assez restrictive dans la pratique. Ce sera notamment le cas lorsque la non-réponse est causée par le sujet même de l'enquête. On parlera alors de mécanisme *non-ignorable* [9, 10]. Pour traiter ces cas, le calage simple sera impuissant et il faudra faire appel à la théorie généralisée du calage.

1.3. Théorie généralisée du calage

Comme précédemment, on cherche par un seul et unique calage à réduire à la fois le biais de non-réponse et la variance d'échantillonnage. On cherche dans le même temps à s'affranchir de la contrainte qui pèse sur le calage simple selon laquelle les variables utilisées pour corriger la non-réponse doivent être aussi des variables de calage, c'est-à-dire que leur total sur la population doit être connu. On pose l'équation de calage suivante :

$$\sum_{k \in r} \frac{d_k}{\theta} F(z_k, \lambda) x_k = X \quad (7)$$

Les notations sont identiques à celle utilisées dans (5). La nouveauté cependant par rapport à (5) réside dans l'introduction de p nouvelles variables $z = (z^{(1)}, z^{(2)} \dots z^{(p)})$. Ces variables vont servir à réduire le biais de non-réponse et seront dorénavant qualifiées de variables de non-réponse. Pour être efficaces, elles devront être corrélées avec la probabilité de réponse. Quant aux p variables de calage $x = (x^{(1)}, x^{(2)} \dots x^{(p)})^3$, ce sont des variables de calage au sens original du terme, c'est-à-dire que leur objectif est de réduire la variance d'échantillonnage. Elles devront donc être corrélées avec la variable d'intérêt de l'enquête.

A l'instar du calage simple, On montre [1, 3] que le calage généralisé (7) est équivalent à la pondération en deux étapes suivantes :

³ Dans l'équation du calage généralisé, le nombre de variables de non-réponse doit être égal aux nombre de variables de calage. Cela assure l'identifiabilité des paramètres de l'équation de calage.

1. division du poids de sondage d_k par : $r_k = \frac{m}{n} \frac{1}{F(z_k \theta)}$ (8), où F est la fonction de calage utilisée dans (7) et θ un paramètre déterministe. Là encore, toute la validité du calage repose sur l'hypothèse d'une modélisation exacte des probabilités de réponse par (8),
2. calage des poids $\frac{d_k}{r_k}$ ajustés pour la non-réponse en utilisant les variables de calage x .

On montre aussi [1, 3] que la variance de l'estimateur du calage généralisé dépend de la variance des résidus de la régression instrumentale [6] sur les variables de calage, avec les variables de non-réponse comme instruments. La formule de décomposition de la variance n'étant plus valable dans le cas d'une régression instrumentale, la variance de l'estimateur du calage généralisé pourra être dégradée par rapport à celle de l'estimateur avant calage. Nous reviendrons sur ce point plus loin dans le document.

A la différence du calage simple, le calage généralisé fait donc la distinction entre les variables qui vont contribuer à réduire le biais de non-réponse (les variables de non-réponse) et celles qui vont réduire la variance d'échantillonnage (les variables de calage). La seule condition sur les variables de non-réponse est de connaître leurs valeurs sur les unités répondantes. En outre, il n'y a plus besoin de connaître le total sur la population. La contrainte du calage simple est donc levée, ce qui ouvre des perspectives beaucoup plus étendues pour le traitement de la non-réponse. On pourra par exemple utiliser comme variables de non-réponse des variables qui ont été collectées au moment de l'enquête. Ces variables actualisées sont souvent explicatives de la non-réponse (mais leur total reste souvent inconnu). Si la non-réponse est non-ignorable, c'est-à-dire qu'elle est causée par le sujet de l'enquête, on pourra même insérer dans l'équation de calage (7) les variables d'intérêt de l'enquête.

2. Description de la simulation

L'objectif de la simulation est de tester le calage généralisé pour corriger la non-réponse dans le cas non-ignorable. On va s'appuyer sur les données de l'enquête Budget des Ménages (EBM) pour le Luxembourg. Cette enquête est réalisée de façon permanente par le Statec depuis 2003 sur un échantillon annuel compris entre 1000 et 1500 ménages, avec comme objectif de mesurer la dépense de consommation moyenne des ménages (pour l'ensemble des dépenses et par catégorie COICOP⁴) ainsi que la structure de consommation.

Les calculs ont été rendus possibles grâce à la nouvelle version de la macro Calmar (Calmar 2 – [8]) dans laquelle la technique du calage généralisé a été implantée.

2.1. Les différentes étapes de la simulation

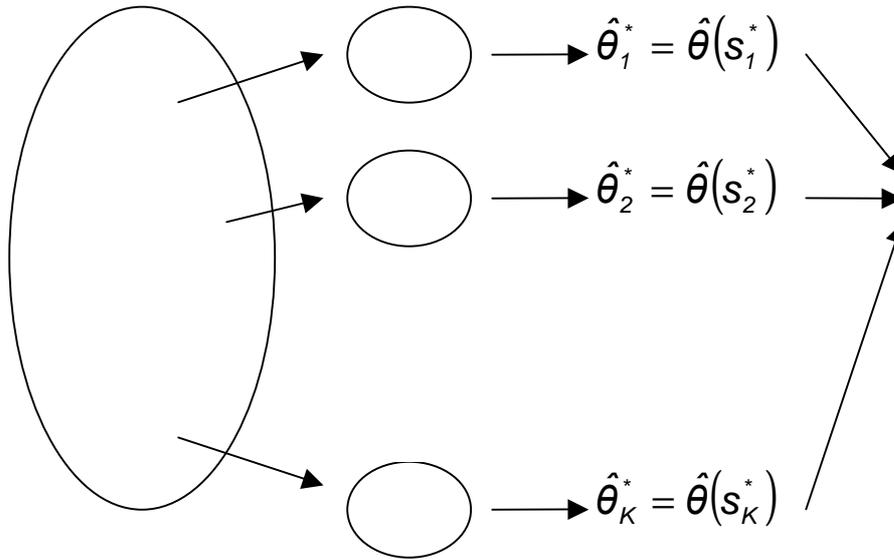
La population de référence comprend l'ensemble des ménages qui ont participé à l'EBM au moins une année au cours de la période 2005-2010. Cela représente 8000 ménages, soit 4% du nombre total de ménages qui résident au Luxembourg. A l'intérieur de cette population, on va tirer $K=2000$ répliques selon un mécanisme non-ignorable. Ce mécanisme va représenter le mécanisme de réponse des ménages à l'EBM. Chaque réplique peut donc être vue comme un échantillon de ménages répondants à partir duquel on va estimer un paramètre-cible θ . Dans le cas présent, on s'intéressera à la dépense de consommation moyenne par ménage. On calculera finalement le biais, la variance et l'erreur quadratique moyenne (EQM) sur l'ensemble des répliques.

L'estimation du paramètre θ implique de pondérer les ménages de chaque réplique. A ce niveau, on va tester trois stratégies de pondération :

- Utilisation du taux de réponse global
- Calage simple en une étape
- Calage généralisé

⁴ COICOP=Classification Of Individual COnsumption by Purpose

Figure 1 : Les différentes étapes de la simulation



$$\text{Biais} = \frac{1}{K} \sum_{i=1}^K \hat{\theta}_i^* - \theta$$

$$\text{Variance} = \frac{1}{K} \sum_{i=1}^K \left(\hat{\theta}_i^* - \frac{1}{K} \sum_{i=1}^K \hat{\theta}_i^* \right)^2 = \frac{1}{K} \sum_{i=1}^K (\hat{\theta}_i^*)^2 - \left(\frac{1}{K} \sum_{i=1}^K \hat{\theta}_i^* \right)^2$$

$$\text{EQM} = \frac{1}{K} \sum_{i=1}^K (\hat{\theta}_i^* - \theta)^2 = \text{Variance} + \text{Biais}^2$$

2.2. Construction du mécanisme non-ignorable

Chacune des $K=2000$ répliques est tirée selon un sondage de Poisson. Chaque ménage k de la population (les 8000 ménages qui ont participé à l'enquête au moins une fois entre 2005 et 2010) se voit attribuer une probabilité r_k d'être tiré. Cette probabilité correspond à la probabilité de réponse du ménage à l'EBM. On génère ensuite pour chaque ménage un nombre réel aléatoire u_k compris entre 0 et 1. Si $u_k \leq r_k$ alors on retient le ménage. Dans le cas contraire, celui-ci n'est pas sélectionné. Ces nombres aléatoires sont générés indépendamment d'un ménage à l'autre. Il en résulte que les tirages successifs sont indépendants, et que chaque ménage k dans la population a bien une probabilité r_k d'apparaître dans l'échantillon.

Il faut maintenant calculer la probabilité r_k . On postule que la probabilité de réponse r_k d'un ménage k suit une forme logistique :

$$r_k = \frac{\exp(a' x_k)}{1 + \exp(a' x_k)} \quad (9)$$

où x_k désigne la valeur sur k d'un vecteur de variables explicatives de la probabilité de réponse, dont les valeurs sont connues pour les ménages répondants et non répondants, et a un paramètre constant. On estime a à partir du fichier de l'enquête en utilisant une procédure d'estimation standard (par exemple, maximum de vraisemblance)

Comme variables du modèle (9), on a utilisé :

- l'âge (**AGE**) de la personne de référence du ménage⁵ : entre 15 et 29 ans, entre 30 et 49 ans, entre 50 et 64 ans ou 65 ans ou plus,
- le genre (**GEN**) de la personne de référence du ménage : homme/femme,
- la nationalité (**NAT**) de la personne de référence du ménage : luxembourgeois/étranger,
- la taille du ménage (en nombre de personnes) (**TAI**) : 1, 2, 3 ou plus de 3 personnes.

En croisant les données de l'enquête avec celles du Répertoire Général des Personnes Physique (RGPP), on a pu observer que le taux de réponse dépendait bien de ces variables. Pour construire un mécanisme non-ignorable, on a également rajouté à la liste des variables le logarithme de la dépense totale de consommation du ménage. Cette variable n'étant connue que pour les ménages répondants, il a fallu préalablement imputer une valeur aux ménages non répondants. On a fait l'hypothèse que la dépense de consommation des ménages non répondants suivait une loi log-normale, dont la moyenne a été choisie de manière à être supérieure à celle observée sur les ménages répondants. On a construit ainsi un mécanisme de réponse dans lequel la probabilité de réponse d'un ménage dépend du niveau de sa dépense de consommation : plus celle-ci est élevée, plus la probabilité de réponse du ménage diminue. Ce postulat est assez intuitif : on peut facilement imaginer que les ménages qui dépensent beaucoup auront tendance à moins participer à l'enquête, en raison notamment de la charge de travail liée au remplissage des carnets auto-administrés.

Il faut bien comprendre qu'il s'agit là malgré tout d'un véritable « pari » sur le mécanisme de réponse des ménages dans l'EBM : on a supposé ici que la probabilité de réponse d'un ménage dépendait de son niveau de dépense. Même si cette idée peut sembler plausible (cf. paragraphe précédent), il faut bien se rendre compte que nous n'avons pas arguments tangibles permettant d'étayer cette hypothèse. Il faut donc y croire. C'est sur cette hypothèse que nous nous sommes appuyés par la suite afin de tester la force du calage généralisé pour corriger la non-réponse.

En fonction des paramètres choisis, on a obtenu les trois modèles suivants. Ce sont ces modèles que nous allons utiliser dans la simulation :

Modèle 1 (non-ignorable)

$$\begin{aligned} \text{Logit}(r_k) = & 11.983 \\ & + (-0.466) \cdot (\text{NAT} = \text{"Etranger"})_k \\ & + (0.451) \cdot (\text{AGE} = \text{"15 - 29"})_k + (0.454) \cdot (\text{AGE} = \text{"30 - 49"})_k + (0.370) \cdot (\text{AGE} = \text{"50 - 64"})_k \\ & + (-0.630) \cdot (\text{TAI} = \text{"1"})_k + (-0.266) \cdot (\text{TAI} = \text{"2"})_k + (-0.077) \cdot (\text{TAI} = \text{"3"})_k \\ & + (0.166) \cdot (\text{GEN} = \text{"Femme"})_k \\ & + (-1.211) \cdot \text{Log}(t_k) \end{aligned}$$

Modèle 2 (non-ignorable)

⁵ D'après les recommandations d'Eurostat, la personne de référence du ménage est définie comme le plus gros contributeur au budget du ménage.

$$\begin{aligned}
\text{Logit}(r_k) &= 8.541 \\
&+ (-0.456) \cdot (\text{NAT} = \text{"Etranger"})_k \\
&+ (0.422) \cdot (\text{AGE} = \text{"15 - 29"})_k + (0.443) \cdot (\text{AGE} = \text{"30 - 49"})_k + (0.342) \cdot (\text{AGE} = \text{"50 - 64"})_k \\
&+ (-0.582) \cdot (\text{TAI} = \text{"1"})_k + (-0.250) \cdot (\text{TAI} = \text{"2"})_k + (-0.062) \cdot (\text{TAI} = \text{"3"})_k \\
&+ (0.168) \cdot (\text{GEN} = \text{"Femme"})_k \\
&+ (-0.905) \cdot \text{Log}(t_k)
\end{aligned}$$

Modèle 3 (ignorable)

$$\begin{aligned}
\text{Logit}(r_k) &= -1.435 \\
&+ (-0.424) \cdot (\text{NAT} = \text{"Etranger"})_k \\
&+ (0.390) \cdot (\text{AGE} = \text{"15 - 29"})_k + (0.406) \cdot (\text{AGE} = \text{"30 - 49"})_k + (0.305) \cdot (\text{AGE} = \text{"50 - 64"})_k \\
&+ (-0.479) \cdot (\text{TAI} = \text{"1"})_k + (-0.195) \cdot (\text{TAI} = \text{"2"})_k + (-0.031) \cdot (\text{TAI} = \text{"3"})_k \\
&+ (0.180) \cdot (\text{GEN} = \text{"Femme"})_k
\end{aligned}$$

Les noms de variables indiqués entre parenthèses et suivis d'un signe = correspondent à l'indicatrice 0/1 de la modalité qui suit le =. Par exemple, $(\text{NAT} = \text{"Etranger"})_k$ vaut 1 si la personne de référence du ménage k est de nationalité étrangère, 0 sinon. t_k est la dépense totale du ménage k .

Les deux premiers modèles sont non-ignoraux puisqu'ils intègrent dans leur expression le logarithme de la dépense de consommation du ménage. A l'opposé, le troisième modèle n'intégrant pas la dépense de consommation, celui-ci sera qualifié d'ignorable. On va pouvoir ainsi regarder ce que le calage généralisé donnerait si l'on supposait à tort que la probabilité de réponse d'un ménage dépendait de sa dépense de consommation.

2.3. Les différents scénarios de pondération

Une fois une réplique tirée, il s'agit d'estimer le paramètre (la dépense moyenne de consommation par ménage) à partir de l'information contenue dans l'échantillon. Cette étape requiert de pondérer les ménages. Différents scénarios de pondération ont été testés au cours de la simulation :

- **Utilisation du taux de réponse global**

Les ménages de l'échantillon sont pondérés par l'inverse du taux de réponse global $\bar{\theta} = \frac{m}{n}$, où m désigne le nombre de ménages dans l'échantillon et n le nombre de ménages dans la population. Ce modèle naïf considère que tous les ménages de la population ont en fait la même probabilité de répondre. Ce modèle revient à "ne rien faire pour traiter les non-réponses" c'est-à-dire à utiliser les formes habituelles d'estimateurs en les limitant aux seules unités répondantes [2]

- **Calage simple en une étape**

Le calage a porté sur les variables suivantes :

- l'âge, le genre et la nationalité de la personne de référence du ménage,
- la taille du ménage (en nombre de personnes)

- **Calage généralisé**

On a utilisé comme variables de non-réponse :

- l'âge, le genre et la nationalité de la personne de référence du ménage,
- la taille du ménage (en nombre de personnes)
- la dépense totale de consommation du ménage

A l'exception de la dernière variable, toutes les autres ont été également utilisées comme variables de calage dans (7). Comme il faut assurer le même nombre de variables de non-réponse que de

variables de calage, il a fallu cependant ajouter une variable de calage supplémentaire à la liste. Celle-ci a été générée de façon à avoir un niveau de corrélation donné ρ avec la dépense totale de consommation du ménage. Si t_k désigne la dépense totale du ménage k et ρ un paramètre fixé compris entre 0 et 1, on pose :

$$t_k^{(\rho)} = \tilde{K} \times \left(t_k + S_t \frac{\sqrt{1-\rho^2}}{\rho} \varepsilon \right) \quad (10)$$

- S_t désigne la racine-carrée de la dispersion de la dépense totale t sur l'ensemble des n ménages de la population :

$$S_t^2 = \frac{1}{n-1} \sum_{i=1}^n (t_k - \bar{t})^2 \quad (11)$$

- ε est une variable aléatoire de loi normale centrée réduite qui a été générée indépendamment de t
- \tilde{K} est un facteur de normalisation qui assure que le total de la variable $t^{(\rho)}$ sur la population est toujours égal au nombre total n de ménages ($n \approx 8000$)

On vérifie très facilement que $Corr(t^{(\rho)}, t) \approx \rho$.

2.4. Le calage généralisé dans Calmar 2

Les calculs ont été réalisés à partir de la nouvelle version de la macro Calmar (Calmar 2 – [8]) dans laquelle la technique du calage généralisée a été programmée. La marche à suivre pour lancer un calage généralisé avec Calmar 2 est très simple :

1. Dans la table qui contient les marges du calage, il faut insérer l'ensemble des variables de calage ainsi que l'ensemble des variables de non-réponse. Si une variable sert à la fois à la non-réponse et au calage, elle sera enregistrée deux fois. La table devra également contenir une colonne supplémentaire, appelée R , qui va servir à distinguer les variables de non-réponse des variables de calage : R vaudra 0 pour les variables de calage et 1 pour les variables de non-réponse. Pour ces dernières, les marges ne seront pas renseignées.

Tableau 2 : un exemple de table des marges dans CALMAR 2

VAR	N	R	MAR1
age1	0	0	568
age2	0	0	3872
age3	0	0	2153
sex1	0	0	1927
nat2	0	0	5183
D60	0	0	7948
age1	0	1	.
age2	0	1	.
age3	0	1	.
sex1	0	1	.
nat2	0	1	.
t_00	0	1	.

2. Lors de l'appel à la macro, pour indiquer que l'on veut un calage généralisé, le paramètre NONREP doit être égal à OUI (Sinon, un message d'erreur apparaît)

```
%CALMAR2
(DATAMEN=BOOT,
POIDS=POIDS0,
IDENT=id,
```

```

MARMEN=MARGES1 ,
M=2 ,
NONREP=OUI ,
DATAPOI=TABCAL ,
POIDSFIN=POIDS1 ,
MAXITER=50 ,
COLIN=OUI ) ;

```

3. Principaux résultats

On va maintenant présenter les principaux résultats de cette simulation (l'intégralité des tableaux de résultats est disponible en annexe). Un des paramètres de la simulation est la valeur du coefficient de corrélation ρ entre la dépense totale de consommation et la variable de calage fictive (10) qui a été créée afin d'avoir autant de variables de non-réponse que de variables de calage. Différentes valeurs de ρ ont été testées : $\rho=0.8, 0.6, 0.4, 0.2, 0.1$ et 0.05 . Un autre paramètre est la méthode de calage.

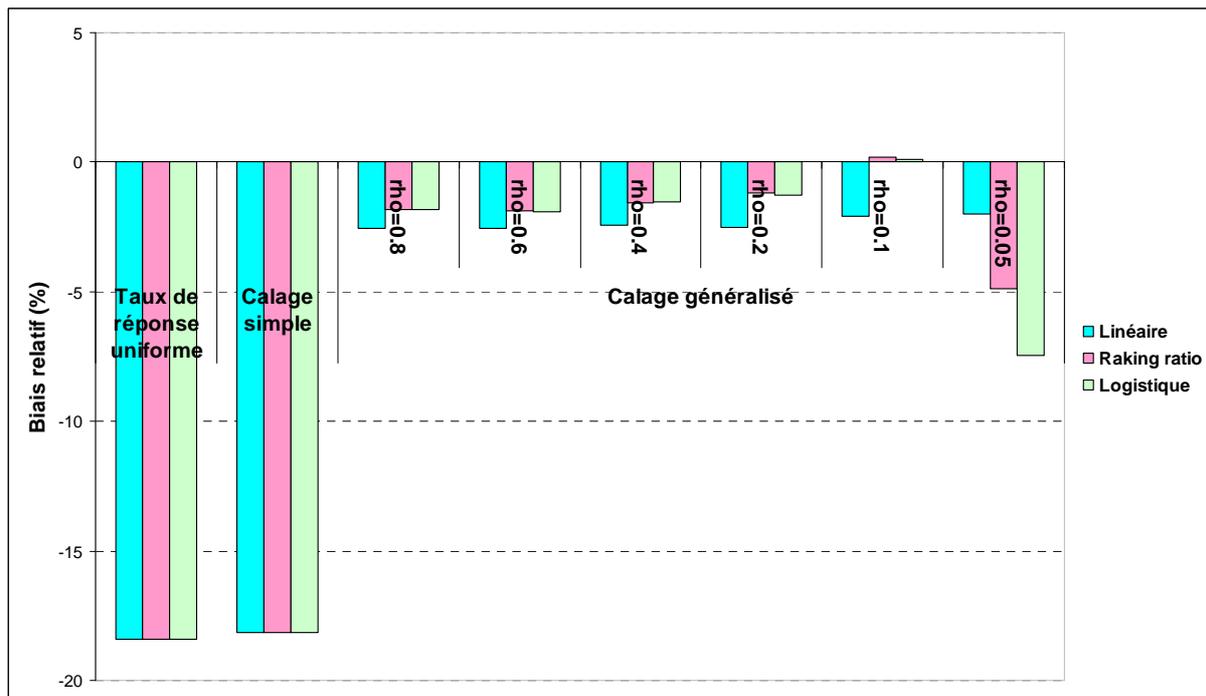
On a testé ici la méthode linéaire, la méthode du raking ratio et la méthode logistique (cf. Tableau 1). Pour cette dernière, les valeurs pour la borne inférieure et la borne supérieure du ratio des poids ont été fixées respectivement à 0.1 et 30 . On a enfin considéré trois modèles de réponse pour les ménages. Ces modèles ont déjà été introduits à la section 2.2. On conservera leur numérotation, bien qu'on les abrègera en $(M1)$, $(M2)$ et $(M3)$. On a vu que les modèles $(M1)$ et $(M2)$ sont non-ignorables, car ils intègrent la dépense de consommation du ménage, tandis que le modèle $(M3)$ ne l'intègre pas et sera donc qualifié d'ignorable.

Le paramètre-cible que l'on considère ici est la dépense moyenne de consommation par ménage.

3.1. Impact sur le biais

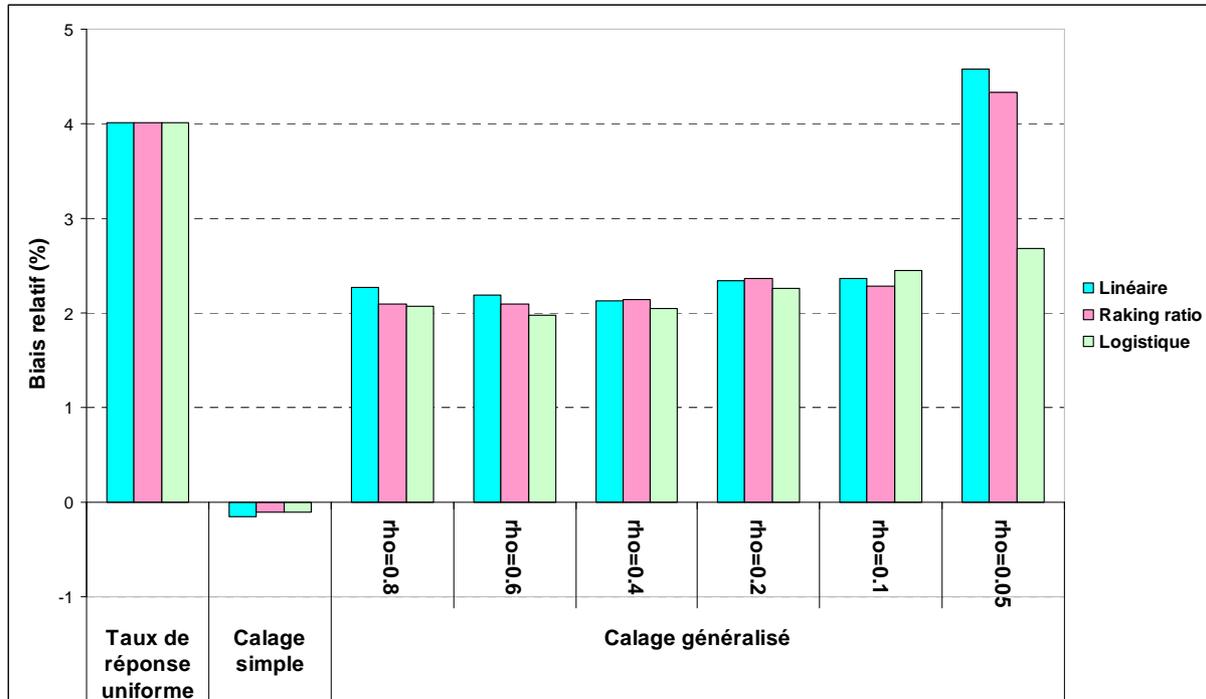
Sous les modèles non-ignorables $(M1)$ et $(M2)$, on observe que le recours au calage généralisé a réduit substantiellement le biais. L'impact semble être indifférent à la méthode de calage utilisée. Le décrochage pour $\rho=0.05$ s'explique par la forte instabilité du calage généralisé lorsque les variables de non-réponse sont très peu corrélées avec les variables de calage (cf. section suivante)

Figure 2 : Biais relatif (%) – Cas non-ignorable (Modèle $(M1)$)



Dans le cas ignorable (modèle (M3)), on voit que le calage généralisé n'a pas dégradé le biais par rapport à la situation d'avant calage. Il faut dire que les variables de non-réponse qui ont été utilisées dans l'équation de calage comprenaient, en plus de la dépense de consommation du ménage, toutes les variables du modèle (M3). C'est bien l'utilisation de ces variables a permis de réduire le biais, mais dans une plus faible mesure par rapport au calage simple. Ce résultat est malgré tout intéressant et confère au calage généralisé un aspect « robuste » qui est particulièrement séduisant pour les utilisateurs.

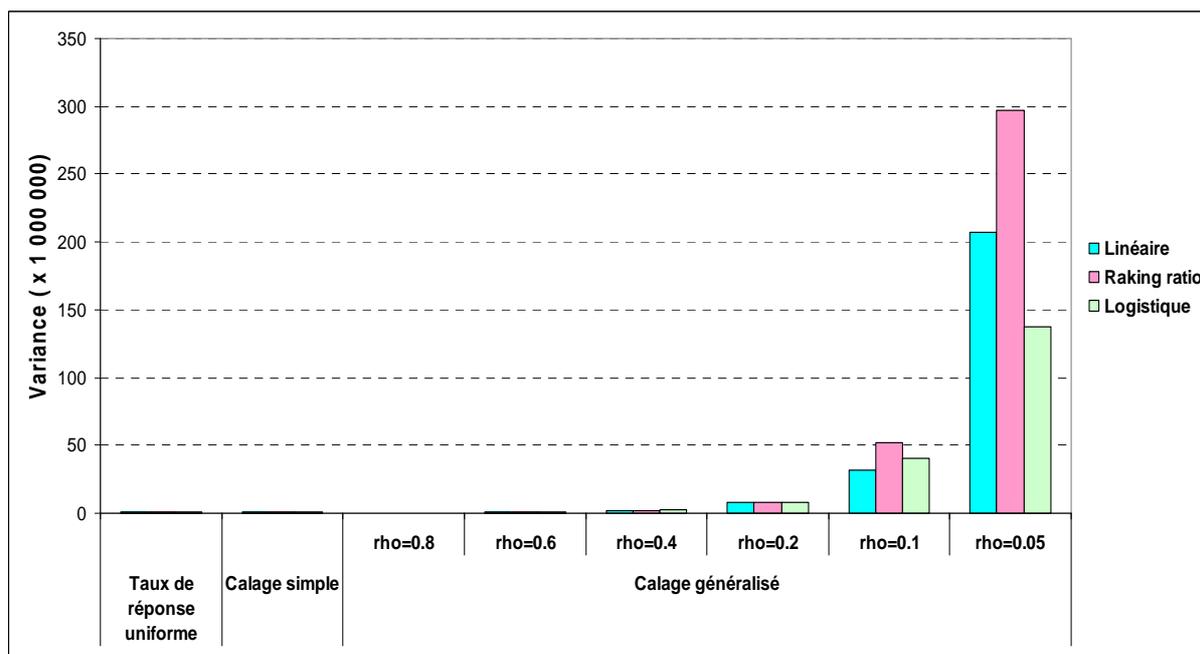
Figure 3 : Biais relatif (%) – Cas ignorable (Modèle (M3))



3.2. Impact sur la variance

Pour les trois modèles considérés, le calage généralisé a conduit à une augmentation de la variance à mesure que le coefficient de corrélation ρ devenait proche de zéro. Il semble ainsi que le gain du calage généralisé en termes de réduction du biais se fasse au prix d'une dégradation de la variance. Ce résultat est à l'envers de ce que l'on obtient généralement avec un calage simple.

Figure 4 : Variance – Cas non-ignorable (Modèle (M1))



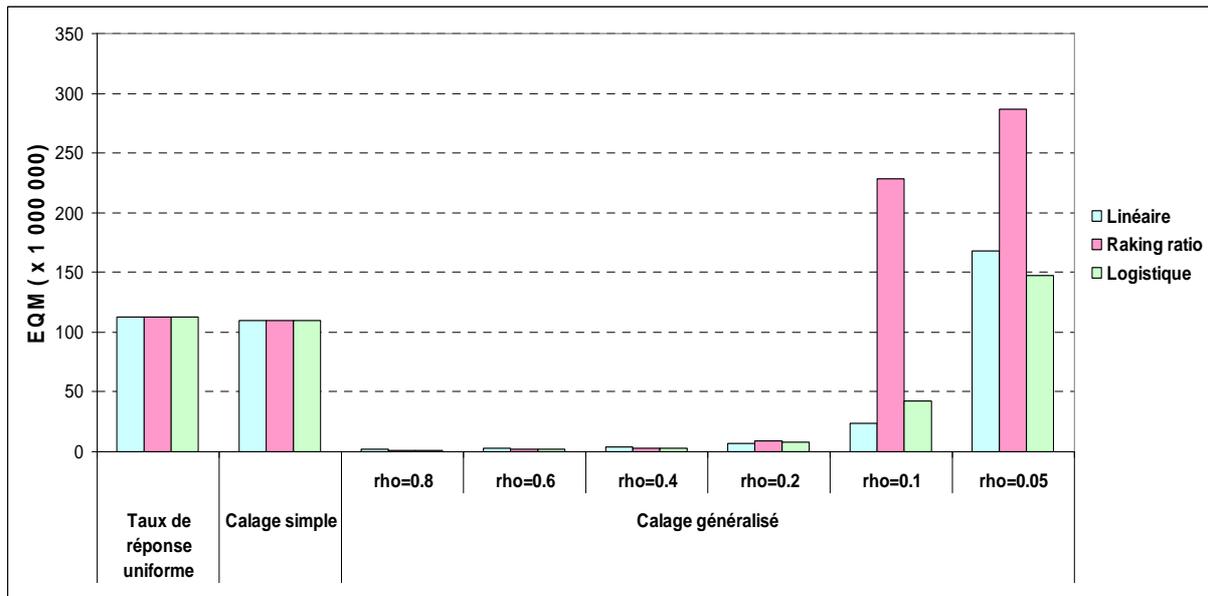
On peut comprendre ce résultat en se rappelant (cf. section 1.3) que la variance de l'estimateur du calage généralisé dépend de la variance des résidus de la régression instrumentale sur les variables de calage, avec les variables de non-réponse comme instruments. A la différence de ce qui se passe avec la régression linéaire classique, il n'existe pas en régression instrumentale de formule de décomposition de la variance qui permette de dire que les résidus sont toujours moins dispersés que la variable d'intérêt. Il est donc parfaitement possible, surtout lorsque les instruments sont très peu corrélés avec les régresseurs, que la dispersion des résidus devienne très importante et dépasse celle de la variable d'intérêt de la régression.

Même si la tendance observée est la même quelle que soit la méthode de calage utilisée, la dégradation de la variance est beaucoup plus forte avec la méthode du raking ratio qu'avec la méthode linéaire ou la méthode logistique. Il faut dire que la méthode logistique permet de contrôler la dispersion des poids calés (au travers de bornes qui sont fixées au départ par l'utilisateur), ce que ne permet pas le raking ratio. La méthode du raking ratio conduit à des poids calés qui sont toujours positifs mais qui peuvent prendre des valeurs très élevées, ce qui ne sera pas bénéfique pour la variance. Afin de « limiter les dégâts » en termes de dégradation de la variance, on aura donc plutôt intérêt à utiliser une méthode de calage de type « bornée » comme la méthode logistique ou la méthode linéaire tronquée (cf. Tableau 1)

3.3. Impact sur l'EQM

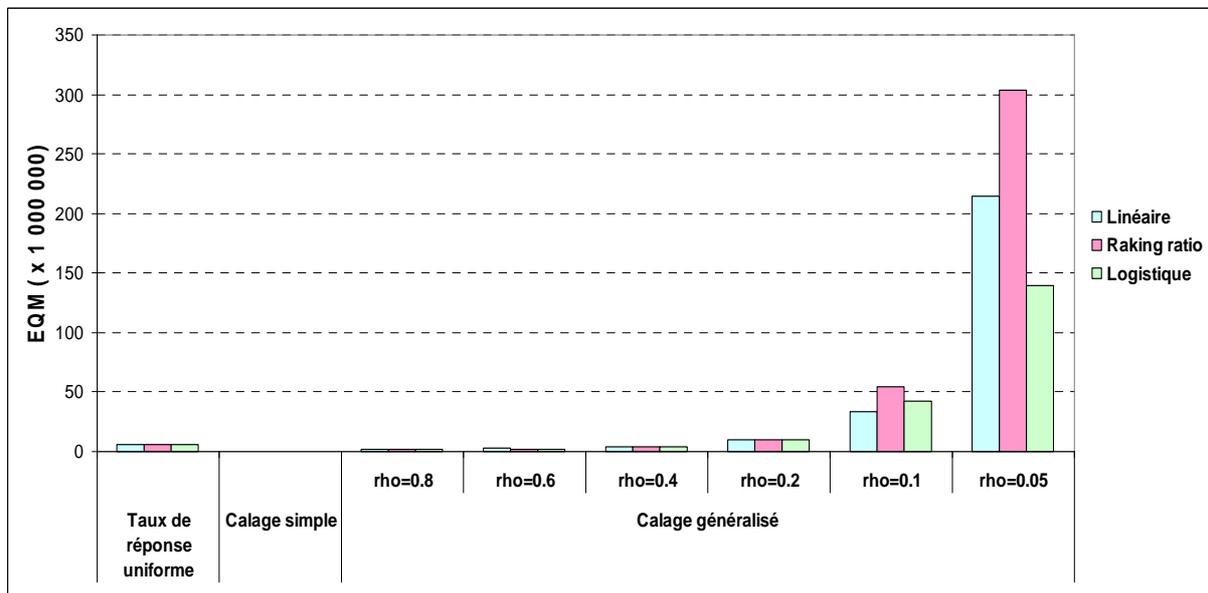
L'erreur quadratique moyenne (EQM) mesure la qualité globale d'un estimateur en tenant compte à la fois de son biais et de sa variance. Dans le cas des modèles ($M1$) et ($M2$), le gain en termes de biais a été prédominant sur l'augmentation de la variance, et le calage généralisé a permis une baisse de l'EQM. On peut donc dire que la qualité de l'estimateur s'est améliorée. Ce n'est que lorsque la corrélation ρ devient trop faible que l'augmentation de variance l'emporte sur la réduction du biais, et fait que le niveau de l'EQM après calage dépasse celui observé avant calage.

Figure 5 : EQM – Cas non-ignorable (Modèle ($M1$))



Dans le cas ignorable ($M3$), il n'y a plus de gain en termes de biais, mais seulement dégradation de la variance. L'EQM va donc augmenter et l'estimateur va perdre en qualité. On voit donc le risque que l'on court à supposer à tort que la non-réponse dépend de certaines variables d'intérêt de l'enquête : on va dégrader la variance pour un gain en termes de biais qui est inexistant. C'est bien là le problème du calage généralisé. La méthode est à double tranchant. Son résultat dépend d'une hypothèse qui, si elle est vraie, permet de dire que le calage a amélioré la qualité de l'estimateur. Dans le cas contraire, le recours au calage généralisé aggrave la situation.

Figure 6 : EQM – Cas ignorable (Modèle ($M3$))



4. Conclusion

Les résultats de cette simulation montrent que le calage généralisé peut réduire sensiblement le biais dû à la non-réponse, en particulier lorsque celle-ci dépend des variables d'intérêt de l'enquête (mécanisme non-ignorable). Il s'agit là d'un résultat assez remarquable, qu'on ne peut pas atteindre avec un calage simple ou en estimant explicitement les probabilités de réponse. Le prix à payer pour ce résultat est une dégradation de la variance d'échantillonnage, d'autant plus importante que les

variables de non-réponse sont peu corrélées avec les variables de calage. L'estimateur calé devient donc de plus en plus instable.

Le calage généralisé est donc une méthode à double tranchant, qu'il faut employer avec un minimum de discernement. Il faut commencer par se convaincre que la non-réponse est influencée par un certain nombre de caractéristiques qui sont observées uniquement sur les unités répondantes (par exemple, des variables qui sont collectées au moment de l'enquête). Cette étape est délicate. Elle constitue un pari, certains parlent même d'un « acte de foi », dans la mesure où les techniques statistiques ne permettent pas de justifier une telle assertion (pour cela il faudrait pouvoir faire des tests cognitifs en laboratoire). Sans autres arguments, il faut donc croire à l'hypothèse que l'on fait.

Même si l'on est convaincu du lien entre la non-réponse et un certain nombre de variables d'intérêt de l'enquête, il faut aussi se demander quelle est l'importance du biais que cela génère. Si le biais est relativement faible, voire inexistant, on ne gagnera finalement pas grand-chose à le corriger par calage généralisé. On risque même de dégrader la qualité d'ensemble de l'estimateur puisque le gain (faible) lié à la réduction du biais sera perdu en raison de la dégradation de la variance. L'erreur quadratique moyenne (EQM), qui est une mesure de la qualité globale d'un estimateur, pourra même être plus élevée qu'avant calage. On voit donc qu'il y a un arbitrage à faire non seulement sur l'existence d'un biais mais aussi sur son importance, et que là aussi cet arbitrage relève du pari, de l'acte de foi.

Malgré tout, il y a des situations où l'utilisation du calage généralisée semble moins controversée. C'est le cas par exemple lorsque les variables de calage proviennent d'une source vieillissante (par exemple un recensement de la population) tandis que les variables de non-réponse correspondent aux mêmes variables collectées au moment de l'enquête. En raison d'une corrélation significative entre les variables de non-réponse et les variables de calage, on peut espérer une augmentation de variance qui soit relativement sous contrôle, et donc une qualité globale (mesurée par l'EQM) de l'estimateur qui soit meilleure après l'utilisation du calage généralisé. Ce problème des bases de sondage vieillissantes avait d'ailleurs fait l'objet d'une simulation [7] avec des résultats plutôt concluants.

Bibliographie

- [1] Ardilly P., « Les techniques de sondage », éditions Technip, avril 2006.
- [2] Caron N., « La correction de la non-réponse par repondération et par imputation », document de travail de l'INSEE n°M0502, novembre 2005.
- [3] Deville J.C., « La correction de la non-réponse par calage généralisé », Actes des Journées de Méthodologie Statistique, 2002.
- [4] Deville J.C., Särndal C.E., « Calibration estimators in survey sampling », *Journal of the American Statistical Association*, vol 87, n°418, pp 376-382, juin 1992.
- [5] Dupont F., « Calage et redressement de la non-réponse totale : validité de la pratique courante de redressement et comparaison des méthodes alternatives pour l'enquête sur la consommation alimentaire de 1989 », Actes des Journées de Méthodologie Statistique, 1993.
- [6] Fuller W.A., « Measurement error models », John Wiley & Sons, 1987.
- [7] Le Guennec J., Sautory O., « Application du calage généralisé à la correction de la non-réponse : une expérimentation », Actes des Journées de Méthodologie Statistique, 2002.
- [8] Le Guennec J., Sautory O., « Calmar 2 : une nouvelle version de la macro Calmar de redressement d'échantillon par calage », Actes des Journées de Méthodologie Statistique, 2002.
- [9] Rubin D.B., « Multiple imputation in sample surveys – a phenomenological Bayesian approach to nonresponse », *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp 20-34, 1978.
- [10] Rubin D.B., « Multiple Imputation for Nonresponse in Surveys », éditions Wiley, 1987.
- [11] Särndal C.E., Lundström S., « Estimation in surveys with Nonresponse », éditions Wiley, juin 2005.
- [12] Särndal C.E., Lundström S., « Calibration as a Standard Method for Treatment of Nonresponse », *Journal of Official Statistics*, vol 15, n°2, pp 305-327, 1999.
- [13] Tillé Y., « Théorie des sondages : Echantillonnage et estimation en populations finies », éditions Dunod, mars 2001.
- [14] « La macro CALMAR – Redressement d'un échantillon par calage sur marges », guide utilisateur, disponible sur www.insee.fr.

Annexe 1 : Résultats de la simulation - modèle (M1)

1. Biais relatif (%)

Méthode Linéaire		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	-18.410	-18.174	-2.564	-2.567	-2.440	-2.531	-2.078	-1.985
t1_1	Produits alimentaires et boissons non alcoolisées	-8.892	-8.519	0.153	0.149	0.098	0.037	0.329	0.016
t1_2	Boissons alcoolisées, tabac et stupéfiants	-15.880	-14.395	-3.008	-3.008	-3.089	-3.138	-2.786	-2.299
t1_3	Articles d'habillement et articles chaussants	-19.458	-19.536	-2.753	-2.757	-2.584	-2.683	-2.162	-1.975
t1_4	Logement, eau, électricité, gaz et autres combustibles	-13.826	-13.623	-2.110	-2.113	-1.952	-2.017	-1.706	-1.904
t1_5	Ameublement, équipement ménager et entretien courant de la maison	-26.781	-26.845	-3.222	-3.213	-3.396	-3.513	-2.639	-1.606
t1_6	Santé	-17.617	-17.479	-2.015	-2.018	-1.783	-1.861	-1.496	-1.673
t1_7	Transports	-26.484	-25.993	-3.754	-3.766	-3.581	-3.709	-3.026	-2.801
t1_8	Communications	-8.477	-8.429	-1.414	-1.416	-1.392	-1.431	-1.179	-1.133
t1_9	Loisirs et culture	-20.102	-20.075	-3.171	-3.172	-2.950	-3.067	-2.662	-2.416

t1_10	Enseignement	-32.259	-32.170	-2.799	-2.799	-2.792	-2.942	-1.497	0.654
t1_11	Restaurants et hôtels	-23.761	-23.206	-3.827	-3.826	-3.699	-3.819	-3.251	-3.060
t1_12	Autres biens et services	-18.821	-18.826	-2.829	-2.832	-2.596	-2.696	-2.380	-2.164

Méthode Raking Ratio		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	-18.405	-18.150	-1.814	-1.882	-1.569	-1.196	0.202	-4.866
t1_1	Produits alimentaires et boissons non alcoolisées	-8.910	-8.507	-0.358	-0.392	-0.296	-0.248	-0.748	-3.067
t1_2	Boissons alcoolisées, tabac et stupéfiants	-16.024	-14.547	-1.777	-1.757	-1.033	-0.685	0.522	-4.149
t1_3	Articles d'habillement et articles chaussants	-19.456	-19.519	-1.899	-1.977	-1.541	-1.051	-0.337	-5.182
t1_4	Logement, eau, électricité, gaz et autres combustibles	-13.801	-13.581	-1.889	-1.952	-1.734	-1.449	-0.767	-4.276
t1_5	Ameublement, équipement ménager et entretien courant de la maison	-26.731	-26.778	1.130	0.982	1.543	2.355	14.724	-0.626
t1_6	Santé	-17.565	-17.413	-2.053	-2.120	-1.691	-1.518	-1.740	-4.700
t1_7	Transports	-26.515	-26.007	-3.260	-3.345	-2.935	-2.381	-1.817	-7.821
t1_8	Communications	-8.474	-8.405	-1.468	-1.498	-1.436	-1.301	-1.526	-3.539
t1_9	Loisirs et culture	-20.074	-20.015	-2.207	-2.261	-1.900	-1.683	-0.839	-5.953
t1_10	Enseignement	-32.070	-31.979	2.123	2.022	2.525	3.449	8.024	4.043
t1_11	Restaurants et hôtels	-23.762	-23.180	-2.575	-2.628	-2.339	-1.907	-0.936	-6.273
t1_12	Autres biens et services	-18.842	-18.832	-1.786	-1.851	-1.515	-1.075	0.074	-5.252

Méthode Logistique (LO=0.1,UP=30)		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	-18.403	-18.156	-1.842	-1.934	-1.537	-1.291	0.085	-7.469
t1_1	Produits alimentaires et boissons non alcoolisées	-8.929	-8.541	-0.340	-0.395	-0.090	-0.058	0.537	-3.390
t1_2	Boissons alcoolisées, tabac et stupéfiants	-15.997	-14.527	-1.477	-1.514	-1.488	-1.290	0.235	-5.851
t1_3	Articles d'habillement et articles chaussants	-19.433	-19.500	-1.877	-1.966	-1.649	-1.402	0.178	-8.000
t1_4	Logement, eau, électricité, gaz et autres combustibles	-13.829	-13.615	-1.996	-2.073	-1.652	-1.446	-0.742	-6.152
t1_5	Ameublement, équipement ménager et entretien courant de la maison	-26.719	-26.776	0.789	0.598	1.268	1.733	4.348	-8.100
t1_6	Santé	-17.469	-17.332	-1.956	-2.045	-1.871	-1.688	-0.394	-7.908
t1_7	Transports	-26.430	-25.935	-3.080	-3.203	-2.927	-2.546	-0.614	-11.092
t1_8	Communications	-8.500	-8.426	-1.443	-1.481	-1.390	-1.293	-0.776	-4.067
t1_9	Loisirs et culture	-20.079	-20.037	-2.397	-2.483	-1.795	-1.677	0.126	-8.108
t1_10	Enseignement	-32.196	-32.059	1.783	1.662	2.952	3.898	7.428	-6.186
t1_11	Restaurants et hôtels	-23.765	-23.196	-2.512	-2.601	-2.134	-1.816	-0.049	-9.720
t1_12	Autres biens et services	-18.884	-18.876	-1.780	-1.869	-1.552	-1.292	0.562	-7.371

2. Variance

Méthode Linéaire		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	255 351	207 149	251 037	643 479	1 526 584	5 113 840	21 913 413	166 636 532
t1_1	Produits alimentaires et boissons non alcoolisées	2 872	2 250	7 107	7 760	9 595	17 262	54 512	390 281
t1_2	Boissons alcoolisées, tabac et stupéfiants	912	984	3 394	3 449	3 286	3 771	5 951	25 019
t1_3	Articles d'habillement et articles chaussants	4 202	3 803	10 646	12 702	16 789	35 042	118 577	848 865
t1_4	Logement, eau, électricité, gaz et autres combustibles	29 602	29 055	65 617	88 057	141 101	338 997	1 235 319	9 171 415
t1_5	Ameublement, équipement ménager et entretien courant de la maison	9 114	8 655	44 345	50 097	55 656	92 012	295 180	1 710 149
t1_6	Santé	1 695	1 696	4 114	4 387	5 234	7 778	19 982	124 071
t1_7	Transports	35 306	34 082	83 809	97 219	134 492	320 556	1 149 184	8 574 947
t1_8	Communications	237	211	379	413	508	838	2 270	15 588
t1_9	Loisirs et culture	5 196	4 869	13 573	16 595	21 636	44 232	161 045	1 120 091
t1_10	Enseignement	443	446	1 848	1 902	2 115	2 418	3 976	18 500
t1_11	Restaurants et hôtels	7 315	7 177	17 730	22 100	31 829	64 078	226 518	1 597 790
t1_12	Autres biens et services	6 005	5 648	16 800	20 016	29 128	58 038	194 731	1 380 578

Méthode Raking Ratio		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	253 042	214 731	354 884	768 913	2 069 912	8 101 994	227 949 659	278 768 028
t1_1	Produits alimentaires et boissons non alcoolisées	2 786	2 200	17 234	17 266	23 170	28 489	321 451	511 655
t1_2	Boissons alcoolisées, tabac et stupéfiants	829	895	6 206	6 559	7 855	10 088	138 521	77 115
t1_3	Articles d'habillement et articles chaussants	4 000	3 710	20 896	22 799	28 117	66 912	1 269 785	1 687 678
t1_4	Logement, eau, électricité, gaz et autres combustibles	29 019	28 524	123 098	130 585	220 772	570 538	12 191 144	16 504 296
t1_5	Ameublement, équipement ménager et entretien courant de la maison	9 233	8 747	239 753	241 527	253 994	396 300	23 859 570	7 176 490
t1_6	Santé	1 722	1 726	5 491	5 648	7 005	9 929	103 596	413 106
t1_7	Transports	34 323	32 132	138 425	156 880	208 448	532 131	8 416 591	14 535 596
t1_8	Communications	231	213	447	476	574	016	13 705	21 073
t1_9	Loisirs et culture	4 842	4 653	29 527	34 103	47 629	73 060	1 640 459	2 083 664
t1_10	Enseignement	485	494	3 516	3 584	4 027	5 366	99 531	166 135
t1_11	Restaurants et hôtels	7 089	7 157	29 615	36 412	54 672	114 277	1 774 091	2 944 948
t1_12	Autres biens et services	6 012	5 811	36 613	39 686	55 783	122 052	2 168 048	2 348 874

Méthode Logistique (LO=0.1,UP=30)		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	255 931	213 382	327 722	746 403	1 986 682	7 681 456	42 094 096	128 717 412
t1_1	Produits alimentaires et boissons non alcoolisées	2 930	2 355	17 524	17 003	22 900	29 482	95 054	282 492
t1_2	Boissons alcoolisées, tabac et stupéfiants	818	887	6 031	6 162	6 160	7 733	17 384	27 285
t1_3	Articles d'habillement et articles chaussants	3 890	3 592	18 673	21 030	26 958	54 956	232 940	676 021
t1_4	Logement, eau, électricité, gaz et autres combustibles	30 343	29 178	114 336	125 177	210 504	541 591	2 038 625	6 699 579
t1_5	Ameublement, équipement ménager et entretien courant de la maison	9 339	8 924	173 943	176 950	194 802	286 548	996 637	1 662 548
t1_6	Santé	1 677	1 671	5 175	5 400	6 321	9 560	32 893	85 664
t1_7	Transports	35 824	34 435	121 785	141 905	201 177	497 641	2 159 718	6 417 097
t1_8	Communications	239	213	409	434	577	1 077	3 620	11 894
t1_9	Loisirs et culture	4 551	4 429	24 520	27 866	43 400	71 075	333 523	921 951
t1_10	Enseignement	507	509	3 595	3 696	3 855	5 194	11 106	21 998
t1_11	Restaurants et hôtels	7 385	7 398	29 979	36 487	51 023	109 954	455 851	1 273 334
t1_12	Autres biens et services	5 716	5 414	33 045	36 961	49 282	102 630	449 376	1 177 937

3. EQM

Méthode Linéaire		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	112 616 017	109 700 967	2 430 077	2 828 311	3 500 035	7 237 657	23 345 062	167 942 932
t1_1	Produits alimentaires et boissons non alcoolisées	201 552	184 592	7 166	7 815	9 619	17 266	54 785	390 281
t1_2	Boissons alcoolisées, tabac et stupéfiants	19 270	16 068	4 052	4 108	3 981	4 488	6 516	25 404
t1_3	Articles d'habillement et articles chaussants	541 796	545 691	21 410	23 497	26 272	45 262	125 214	854 406
t1_4	Logement, eau, électricité, gaz et autres combustibles	6 318 159	6 134 697	212 033	234 894	266 483	472 791	1 331 042	9 290 631
t1_5	Ameublement, équipement ménager et entretien courant de la maison	1 059 135	1 063 702	59 548	65 212	72 542	110 080	305 375	1 713 925
t1_6	Santé	74 690	73 555	5 070	5 345	5 982	8 592	20 509	124 730
t1_7	Transports	5 769 497	5 557 483	198 997	213 191	239 336	433 016	1 224 021	8 639 068
t1_8	Communications	10 942	10 794	677	711	796	143	2 477	15 779
t1_9	Loisirs et culture	759 535	757 189	32 343	35 383	37 885	61 796	174 270	1 130 992
t1_10	Enseignement	8 590	8 548	1 910	1 963	2 176	2 486	3 993	18 503
t1_11	Restaurants et hôtels	1 155 601	1 102 461	47 511	51 869	59 664	93 744	248 012	1 616 830

t1_12	Autres biens et services	917 197	917 266	37 390	40 650	46 463	76 740	209 303	392 625	1
-------	--------------------------	---------	---------	-----------	-----------	-----------	-----------	---------	---------	---

Méthode Raking Ratio

		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	112 551 656	109 413 982	1 445 440	1 942 618	2 885 886	8 576 209	227 963 241	286 615 787
t1_1	Produits alimentaires et boissons non alcoolisées	202 273	184 064	17 555	17 651	23 389	28 644	322 859	535 290
t1_2	Boissons alcoolisées, tabac et stupéfiants	19 522	16 301	6 436	6 783	7 933	10 122	138 541	78 368
t1_3	Articles d'habillement et articles chaussants	541 452	544 653	26 016	28 351	31 489	68 481	1 269 946	1 725 802
t1_4	Logement, eau, électricité, gaz et autres combustibles	6 295 409	6 096 064	240 463	255 987	319 692	639 585	12 210 514	17 105 689
t1_5	Ameublement, équipement ménager et entretien courant de la maison	1 055 403	1 058 552	241 622	242 938	257 479	404 422	24 176 962	7 177 063
t1_6	Santé	74 287	73 042	6 483	6 705	7 678	10 471	104 308	418 301
t1_7	Transports	5 781 837	5 561 885	225 315	248 381	278 859	578 497	8 443 593	15 035 656
t1_8	Communications	10 928	10 737	768	810	881	268	14 052	22 938
t1_9	Loisirs et culture	757 097	752 491	38 622	43 649	54 366	78 350	1 641 774	2 149 831
t1_10	Enseignement	8 538	8 501	3 551	3 616	4 077	5 459	100 035	166 263
t1_11	Restaurants et hôtels	1 155 544	1 100 028	43 097	50 454	65 801	121 676	1 775 872	3 024 973
t1_12	Autres biens et services	919 200	918 037	44 817	48 494	61 683	125 022	2 168 062	2 419 811

Méthode Logistique (LO=0.1,UP=30)

		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	112 521 812	109 485 857	1 452 161	1 986 755	2 769 990	8 234 361	42 096 496	147 212 441
t1_1	Produits alimentaires et boissons non alcoolisées	203 252	185 636	17 815	17 395	22 921	29 490	95 779	311 367
t1_2	Boissons alcoolisées, tabac et stupéfiants	19 446	16 250	6 189	6 329	6 322	7 854	17 388	29 777
t1_3	Articles d'habillement et articles chaussants	540 087	543 464	23 674	26 516	30 816	57 746	232 985	766 885
t1_4	Logement, eau, électricité, gaz et autres combustibles	6 322 232	6 127 462	245 377	266 592	300 338	610 400	2 056 715	7 944 580
t1_5	Ameublement, équipement ménager et entretien courant de la maison	1 054 538	1 058 616	174 855	177 473	197 154	290 946	1 024 311	1 758 596
t1_6	Santé	73 455	72 321	6 075	6 383	7 144	10 230	32 929	100 373
t1_7	Transports	5 746 794	5 533 342	199 345	225 769	271 196	550 636	2 162 803	7 422 873
t1_8	Communications	11 001	10 788	719	761	865	1 326	3 709	14 357
t1_9	Loisirs et culture	757 197	753 904	35 242	39 373	49 413	76 326	333 553	1 044 663

t1_10	Enseignement	8 623	8 556	3 620	3 718	3 923	5 313	11 538	22 297
t1_11	Restaurants et hôtels	156 048	101 793	42 815	50 243	60 283	116 660	455 856	465 479
t1_12	Autres biens et services	922 944	921 920	41 192	45 949	55 480	106 925	450 189	317 687

Annexe 2 : Résultats de la simulation - modèle (M2)

1. Biais relatif (%)

Méthode Linéaire		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	-14.010	-14.631	-1.342	-1.389	-1.266	-1.180	-0.857	0.108
t1_1	Produits alimentaires et boissons non alcoolisées	-6.216	-6.995	0.077	0.055	0.033	0.061	0.273	0.522
t1_2	Boissons alcoolisées, tabac et stupéfiants	-12.578	-11.625	-1.694	-1.736	-1.609	-1.520	-1.872	-0.718
t1_3	Articles d'habillement et articles chaussants	-14.362	-15.621	-1.274	-1.325	-1.185	-1.088	-0.665	0.458
t1_4	Logement, eau, électricité, gaz et autres combustibles	-10.724	-11.023	-1.376	-1.413	-1.328	-1.263	-1.091	-0.646
t1_5	Ameublement, équipement ménager et entretien courant de la maison	-20.493	-21.756	-1.026	-1.114	-0.780	-0.677	-0.002	2.227
t1_6	Santé	-13.352	-14.140	-1.258	-1.298	-1.140	-1.059	-0.560	0.043
t1_7	Transports	-20.177	-20.783	-1.768	-1.831	-1.589	-1.453	-1.122	0.417
t1_8	Communications	-5.940	-6.726	-0.892	-0.909	-0.873	-0.837	-0.790	-0.356
t1_9	Loisirs et culture	-15.067	-16.062	-1.654	-1.704	-1.650	-1.573	-0.951	0.328

t1_10	Enseignement	-25.428	-26.662	-0.804	-0.915	-0.262	-0.031	1.055	4.072
t1_11	Restaurants et hôtels	-18.382	-18.528	-1.901	-1.956	-1.919	-1.798	-1.347	-0.184
t1_12	Autres biens et services	-14.284	-15.209	-1.493	-1.542	-1.400	-1.308	-0.978	0.109

Méthode Raking Ratio		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	-14.043	-14.644	-0.937	-0.993	-0.772	-0.240	0.496	-3.444
t1_1	Produits alimentaires et boissons non alcoolisées	-6.308	-7.060	-0.307	-0.351	-0.347	-0.221	0.018	-2.799
t1_2	Boissons alcoolisées, tabac et stupéfiants	-12.903	-11.880	-0.891	-0.893	-0.668	-0.104	-0.154	-0.348
t1_3	Articles d'habillement et articles chaussants	-14.346	-15.625	-0.826	-0.877	-0.733	-0.141	0.423	-3.788
t1_4	Logement, eau, électricité, gaz et autres combustibles	-10.731	-11.008	-1.200	-1.243	-1.187	-0.831	-0.317	-3.325
t1_5	Ameublement, équipement ménager et entretien courant de la maison	-20.621	-21.850	1.617	1.534	2.883	4.198	10.733	1.141
t1_6	Santé	-13.383	-14.152	-1.396	-1.456	-1.364	-0.997	-1.121	-5.032
t1_7	Transports	-20.182	-20.772	-1.644	-1.721	-1.491	-0.770	-0.943	-5.206
t1_8	Communications	-6.031	-6.802	-1.037	-1.059	-0.948	-0.761	-1.055	-2.993
t1_9	Loisirs et culture	-15.022	-15.994	-1.008	-1.074	-0.912	-0.414	0.061	-3.959
t1_10	Enseignement	-25.682	-26.941	1.010	0.877	1.676	2.844	5.355	1.484
t1_11	Restaurants et hôtels	-18.455	-18.562	-1.287	-1.340	-1.061	-0.415	-0.247	-4.872
t1_12	Autres biens et services	-14.324	-15.249	-0.953	-1.014	-0.769	-0.160	0.507	-3.281

Méthode Logistique (LO=0.1,UP=30)		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	-14.009	-14.642	-0.915	-0.945	-0.673	-0.404	1.355	-4.500
t1_1	Produits alimentaires et boissons non alcoolisées	-6.273	-7.058	-0.279	-0.297	-0.205	-0.139	0.681	-2.427
t1_2	Boissons alcoolisées, tabac et stupéfiants	-12.831	-11.860	-0.845	-0.788	-0.692	-0.416	1.196	-3.675
t1_3	Articles d'habillement et articles chaussants	-14.385	-15.701	-1.042	-1.082	-0.633	-0.294	1.694	-4.624
t1_4	Logement, eau, électricité, gaz et autres combustibles	-10.719	-11.016	-1.255	-1.283	-1.099	-0.906	0.135	-4.055
t1_5	Ameublement, équipement ménager et entretien courant de la maison	-20.467	-21.759	2.134	2.038	2.609	3.173	7.207	-3.034
t1_6	Santé	-13.375	-14.167	-1.527	-1.557	-1.179	-0.968	0.456	-5.313
t1_7	Transports	-20.148	-20.778	-1.642	-1.672	-1.421	-1.045	1.271	-6.755
t1_8	Communications	-5.994	-6.801	-1.043	-1.052	-0.852	-0.764	-0.252	-2.675
t1_9	Loisirs et culture	-14.994	-15.992	-0.973	-0.998	-0.796	-0.579	1.388	-4.889
t1_10	Enseignement	-25.485	-26.770	1.783	1.743	2.567	3.274	7.227	-1.530
t1_11	Restaurants et hôtels	-18.376	-18.516	-1.208	-1.233	-0.758	-0.456	1.836	-5.579
t1_12	Autres biens et services	-14.288	-15.242	-0.871	-0.891	-0.582	-0.253	1.768	-4.338

2. Variance

Méthode Linéaire		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	313 305	260 275	286 907	698 003	1 740 352	5 896 419	25 420 890	188 074 407
t1_1	Produits alimentaires et boissons non alcoolisées	3 563	2 716	6 828	7 725	10 046	17 907	62 245	416 998
t1_2	Boissons alcoolisées, tabac et stupéfiants	1 123	1 181	3 299	3 334	3 693	4 312	6 428	29 443
t1_3	Articles d'habillement et articles chaussants	5 282	4 867	11 264	13 426	17 617	38 162	135 263	953 161
t1_4	Logement, eau, électricité, gaz et autres combustibles	39 264	37 664	72 994	90 059	143 705	366 541	1 364 998	10 012 123
t1_5	Ameublement, équipement ménager et entretien courant de la maison	11 890	10 605	43 915	47 221	64 753	108 481	355 559	2 103 044
t1_6	Santé	2 210	2 150	4 386	4 669	5 168	7 767	21 617	135 348
t1_7	Transports	41 280	39 234	82 091	105 775	154 211	371 665	1 338 674	9 753 569
t1_8	Communications	266	234	352	400	524	899	2 457	16 595
t1_9	Loisirs et culture	6 116	5 798	12 661	15 280	24 049	49 875	185 338	1 294 176
t1_10	Enseignement	672	654	107	123	267	779	4 637	20 511
t1_11	Restaurants et hôtels	8 629	8 324	16 976	21 416	31 488	72 488	264 338	1 806 514
t1_12	Autres biens et services	7 060	6 661	15 342	19 069	27 513	64 335	226 398	1 559 618

Méthode Raking Ratio		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	351 664	285 697	333 837	789 870	1 977 739	8 890 634	151 288 522	302 855 934
t1_1	Produits alimentaires et boissons non alcoolisées	3 547	2 718	14 650	14 238	16 864	25 449	447 864	484 744
t1_2	Boissons alcoolisées, tabac et stupéfiants	1 052	1 121	5 065	5 180	5 477	8 151	27 120	181 156
t1_3	Articles d'habillement et articles chaussants	5 462	4 954	16 656	19 072	24 530	63 051	1 039 952	1 857 143
t1_4	Logement, eau, électricité, gaz et autres combustibles	40 520	38 223	105 941	126 947	184 736	536 853	8 318 215	17 130 472
t1_5	Ameublement, équipement ménager et entretien courant de la maison	13 076	11 785	133 468	141 951	189 285	379 858	13 492 241	7 633 277
t1_6	Santé	2 027	1 988	4 819	4 956	6 125	9 705	75 328	198 616
t1_7	Transports	44 923	42 235	111 454	132 476	183 707	540 699	5 926 820	17 693 188
t1_8	Communications	292	266	436	471	554	071	8 056	20 151
t1_9	Loisirs et culture	6 200	5 734	21 351	23 931	37 468	80 012	1 042 956	2 497 247
t1_10	Enseignement	663	641	684	726	194	328	37 404	75 366
t1_11	Restaurants et hôtels	8 668	8 267	21 965	27 973	45 957	112 820	1 099 298	2 713 157
t1_12	Autres biens et services	7 743	7 175	26 014	29 093	45 863	114 010	1 332 870	2 962 789

Méthode Logistique (LO=0.1,UP=30)		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	336 991	273 686	339 682	846 722	2 090 316	7 822 955	47 631 021	143 330 772
t1_1	Produits alimentaires et boissons non alcoolisées	3 597	2 801	14 760	15 525	16 248	24 050	115 138	280 992
t1_2	Boissons alcoolisées, tabac et stupéfiants	1 054	1 118	4 682	5 022	5 396	6 835	16 800	28 390
t1_3	Articles d'habillement et articles chaussants	5 120	4 553	15 269	17 263	25 837	57 915	260 792	736 006
t1_4	Logement, eau, électricité, gaz et autres combustibles	39 162	37 312	104 059	125 001	190 906	489 286	2 233 462	7 560 047
t1_5	Ameublement, équipement ménager et entretien courant de la maison	12 421	11 071	133 830	138 212	165 993	263 908	1 211 762	1 940 700
t1_6	Santé	2 276	2 241	5 096	5 387	6 414	9 989	36 156	91 482
t1_7	Transports	46 304	43 051	108 585	134 468	189 291	479 919	2 273 938	7 238 693
t1_8	Communications	279	251	426	476	572	1 009	3 579	12 074
t1_9	Loisirs et culture	6 251	5 825	21 879	25 974	38 250	71 085	367 387	1 005 239
t1_10	Enseignement	655	633	863	962	3 248	4 187	9 861	28 555
t1_11	Restaurants et hôtels	9 046	8 787	23 455	29 623	46 553	103 656	522 625	1 435 718
t1_12	Autres biens et services	7 488	7 021	26 795	32 446	44 374	99 742	514 517	1 298 304

3. EQM

Méthode Linéaire		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	65 383 902	71 221 483	883 792	1 337 993	2 271 373	6 357 840	25 664 494	188 078 279
t1_1	Produits alimentaires et boissons non alcoolisées	100 656	125 666	6 843	7 733	10 049	17 916	62 433	417 684
t1_2	Boissons alcoolisées, tabac et stupéfiants	12 641	11 019	3 508	3 554	3 881	4 481	6 683	29 481
t1_3	Articles d'habillement et articles chaussants	298 167	351 327	13 570	15 920	19 611	39 843	135 892	953 460
t1_4	Logement, eau, électricité, gaz et autres combustibles	3 822 934	4 035 115	135 256	155 739	201 691	419 032	1 404 181	10 025 838
t1_5	Ameublement, équipement ménager et entretien courant de la maison	626 723	703 571	45 457	49 036	65 643	109 153	355 559	2 110 305
t1_6	Santé	44 138	49 177	4 758	5 065	5 473	8 030	21 690	135 349
t1_7	Transports	3 369 620	3 570 539	107 636	133 173	174 852	388 928	1 348 968	9 754 987
t1_8	Communications	5 522	6 973	471	523	638	003	2 550	16 614
t1_9	Loisirs et culture	429 900	487 383	17 768	20 700	29 131	54 495	187 027	1 294 377
t1_10	Enseignement	5 735	6 219	2 112	2 129	2 267	2 779	4 645	20 641
t1_11	Restaurants et hôtels	695 917	706 523	24 329	29 201	38 977	79 067	268 031	1 806 583

t1_12	Autres biens et services	531 896	601 614	21 077	25 182	32 553	68 734	228 859	1 559 649
-------	--------------------------	---------	---------	-----------	-----------	-----------	-----------	---------	--------------

Méthode Raking Ratio

		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	65 728 140	71 379 925	625 173	1 116 493	2 175 414	8 909 767	151 370 058	306 787 298
t1_1	Produits alimentaires et boissons non alcoolisées	103 530	127 965	14 887	14 547	17 167	25 572	447 865	504 426
t1_2	Boissons alcoolisées, tabac et stupéfiants	13 172	11 395	5 123	5 238	5 509	8 152	27 122	181 165
t1_3	Articles d'habillement et articles chaussants	297 670	351 594	17 624	20 164	25 294	63 079	1 040 206	1 877 514
t1_4	Logement, eau, électricité, gaz et autres combustibles	3 828 750	4 024 428	153 325	177 740	231 095	559 581	8 321 521	17 494 098
t1_5	Ameublement, équipement ménager et entretien courant de la maison	635 657	710 746	137 298	145 398	201 455	405 665	13 660 882	7 635 183
t1_6	Santé	44 153	49 092	5 278	5 455	6 563	9 938	75 623	204 573
t1_7	Transports	3 374 787	3 569 717	133 564	156 679	201 871	545 546	5 934 085	17 914 732
t1_8	Communications	5 710	7 157	596	638	687	157	8 221	21 486
t1_9	Loisirs et culture	427 459	483 264	23 247	26 083	39 023	80 331	1 042 963	2 526 502
t1_10	Enseignement	5 826	6 323	2 692	2 732	3 216	4 392	37 629	75 383
t1_11	Restaurants et hôtels	701 421	709 071	25 334	31 622	48 249	113 170	1 099 422	2 761 433
t1_12	Autres biens et services	535 511	605 265	28 352	31 738	47 385	114 076	1 333 531	2 990 486

Méthode Logistique (LO=0.1,UP=30)

		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	65 396 395	71 341 670	617 041	1 142 704	2 240 469	7 876 988	48 239 217	150 043 183
t1_1	Produits alimentaires et boissons non alcoolisées	102 473	127 980	14 955	15 747	16 354	24 098	116 304	295 796
t1_2	Boissons alcoolisées, tabac et stupéfiants	13 038	11 357	4 734	5 067	5 431	6 847	16 904	29 374
t1_3	Articles d'habillement et articles chaussants	298 941	354 576	16 809	18 925	26 407	58 038	264 864	766 361
t1_4	Logement, eau, électricité, gaz et autres combustibles	3 819 344	4 029 621	155 847	179 157	230 667	516 288	2 234 060	8 101 006
t1_5	Ameublement, équipement ménager et entretien courant de la maison	625 740	704 233	140 496	144 291	175 962	278 649	1 287 797	1 954 180
t1_6	Santé	44 350	49 444	5 644	5 958	6 741	10 209	36 205	98 121
t1_7	Transports	3 365 190	3 572 537	130 615	157 332	205 794	488 852	2 287 148	7 611 695
t1_8	Communications	5 631	7 141	588	641	680	1 096	3 588	13 140
t1_9	Loisirs et culture	425 924	483 253	23 647	27 833	39 434	71 712	1 370 982	1 049 855

t1_10	Enseignement	5 740	6 243	888 26	2 986	3 300	4 271	10 270	28 574
t1_11	Restaurants et hôtels	695 866	706 093	422 28	32 714	47 723	104 079	529 480	499 030
t1_12	Autres biens et services	532 569	604 625	747	34 486	45 245	99 906	522 556	346 697

Annexe 3 : Résultats de la simulation - modèle (M3)

1. Biais relatif (%)

Méthode Linéaire		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	4.011	-0.153	2.269	2.192	2.134	2.337	2.367	4.577
t1_1	Produits alimentaires et boissons non alcoolisées	4.069	-1.185	-0.055	-0.096	-0.124	-0.043	-0.019	0.963
t1_2	Boissons alcoolisées, tabac et stupéfiants	1.355	0.167	2.128	2.058	1.970	2.159	2.240	4.270
t1_3	Articles d'habillement et articles chaussants	6.320	0.226	2.835	2.747	2.672	2.890	2.851	5.501
t1_4	Logement, eau, électricité, gaz et autres combustibles	1.683	-0.465	1.203	1.149	1.083	1.215	1.268	2.480
t1_5	Ameublement, équipement ménager et entretien courant de la maison	6.927	0.108	4.533	4.403	4.201	4.644	4.958	10.572
t1_6	Santé	3.519	-1.019	1.158	1.093	1.320	1.490	1.160	2.801
t1_7	Transports	5.780	0.424	3.841	3.727	3.737	4.022	3.948	6.686
t1_8	Communications	3.701	-0.112	0.858	0.829	0.828	0.905	0.985	1.925
t1_9	Loisirs et culture	5.506	0.270	2.910	2.829	2.863	3.094	3.085	5.715

t1_10	Enseignement	5.254	-3.029	2.080	1.965	1.331	1.718	1.972	7.913
t1_11	Restaurants et hôtels	4.075	0.413	3.489	3.393	3.219	3.477	3.540	6.446
t1_12	Autres biens et services	4.479	-0.027	2.524	2.446	2.353	2.554	2.614	5.085

Méthode Raking Ratio		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	4.008	-0.105	2.092	2.090	2.146	2.368	2.283	4.325
t1_1	Produits alimentaires et boissons non alcoolisées	4.051	-1.096	-0.062	-0.070	-0.049	0.007	-0.110	-0.251
t1_2	Boissons alcoolisées, tabac et stupéfiants	1.200	0.041	1.882	1.870	2.245	2.456	2.101	4.003
t1_3	Articles d'habillement et articles chaussants	6.291	0.214	2.560	2.555	2.551	2.774	2.471	5.131
t1_4	Logement, eau, électricité, gaz et autres combustibles	1.657	-0.468	1.102	1.096	1.160	1.317	1.194	2.206
t1_5	Ameublement, équipement ménager et entretien courant de la maison	7.029	0.371	4.311	4.360	4.433	5.131	7.462	15.885
t1_6	Santé	3.809	-0.681	1.295	1.284	1.113	1.272	0.992	1.633
t1_7	Transports	5.820	0.521	3.582	3.572	3.655	3.909	3.246	6.294
t1_8	Communications	3.747	-0.009	0.878	0.873	0.823	0.884	0.643	0.723
t1_9	Loisirs et culture	5.486	0.317	2.684	2.684	2.818	3.047	2.780	4.889
t1_10	Enseignement	5.108	-3.243	1.298	1.310	1.290	1.827	1.547	7.415
t1_11	Restaurants et hôtels	4.001	0.452	3.192	3.192	3.174	3.411	3.140	4.504
t1_12	Autres biens et services	4.447	-0.101	2.216	2.206	2.301	2.519	2.450	4.517

Méthode Logistique (LO=0.1,UP=30)		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	4.015	-0.108	2.069	1.981	2.046	2.255	2.447	2.677
t1_1	Produits alimentaires et boissons non alcoolisées	4.011	-1.159	-0.146	-0.191	-0.108	-0.058	0.005	-0.086
t1_2	Boissons alcoolisées, tabac et stupéfiants	1.425	0.228	2.094	2.009	2.205	2.444	2.481	3.188
t1_3	Articles d'habillement et articles chaussants	6.346	0.258	2.580	2.487	2.515	2.728	2.861	3.212
t1_4	Logement, eau, électricité, gaz et autres combustibles	1.695	-0.418	1.140	1.066	1.070	1.228	1.258	1.175
t1_5	Ameublement, équipement ménager et entretien courant de la maison	6.954	0.261	4.165	4.054	4.181	4.793	7.074	9.420
t1_6	Santé	3.715	-0.747	1.217	1.130	1.161	1.283	1.044	0.901
t1_7	Transports	5.786	0.444	3.475	3.348	3.476	3.704	3.598	3.614
t1_8	Communications	3.740	-0.050	0.825	0.785	0.838	0.893	0.779	0.824
t1_9	Loisirs et culture	5.494	0.317	2.664	2.572	2.763	2.953	3.041	3.421
t1_10	Enseignement	4.913	-3.541	0.924	0.810	1.110	1.508	2.046	5.166
t1_11	Restaurants et hôtels	4.030	0.484	3.191	3.085	3.138	3.394	3.516	3.531
t1_12	Autres biens et services	4.484	-0.073	2.226	2.136	2.163	2.389	2.650	2.996

2. Variance

Méthode Linéaire		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	537 968	441 225	325 786	893 430	2 098 632	7 964 503	31 516 427	207 559 670
t1_1	Produits alimentaires et boissons non alcoolisées	5 875	4 538	5 042	5 622	7 900	17 523	58 671	351 817
t1_2	Boissons alcoolisées, tabac et stupéfiants	2 033	2 111	2 435	2 509	2 623	3 807	7 120	34 804
t1_3	Articles d'habillement et articles chaussants	8 732	7 808	8 003	10 783	17 895	46 953	159 804	054 156
t1_4	Logement, eau, électricité, gaz et autres combustibles	63 105	62 300	59 700	84 691	145 835	423 389	1 569 558	10 098 234
t1_5	Ameublement, équipement ménager et entretien courant de la maison	26 481	21 395	26 604	35 917	51 833	135 432	477 197	2 918 052
t1_6	Santé	3 335	3 204	3 517	3 810	4 676	8 152	22 004	128 918
t1_7	Transports	73 565	66 398	69 834	95 275	151 600	456 615	1 644 596	10 653 693
t1_8	Communications	380	342	351	390	488	934	2 690	15 699
t1_9	Loisirs et culture	10 725	9 482	9 824	14 066	22 029	61 946	217 006	1 404 716
t1_10	Enseignement	1 224	1 065	1 233	1 345	1 567	2 170	4 747	24 350
t1_11	Restaurants et hôtels	15 796	13 781	13 776	19 781	32 570	92 652	325 414	2 100 766
t1_12	Autres biens et services	14 153	12 954	13 396	18 817	28 069	76 501	281 761	1 766 358

Méthode Raking Ratio		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	597 785	464 942	308 595	864 722	2 130 185	8 001 819	52 304 229	297 076 056
t1_1	Produits alimentaires et boissons non alcoolisées	5 898	4 565	5 225	6 030	8 184	18 130	86 425	382 267
t1_2	Boissons alcoolisées, tabac et stupéfiants	2 052	2 164	2 505	2 560	2 606	3 731	12 735	64 105
t1_3	Articles d'habillement et articles chaussants	9 333	8 276	8 515	11 494	16 826	45 206	190 224	2 050 496
t1_4	Logement, eau, électricité, gaz et autres combustibles	64 183	61 741	60 605	85 350	150 978	459 153	3 044 693	15 010 929
t1_5	Ameublement, équipement ménager et entretien courant de la maison	27 973	22 561	26 247	36 398	58 867	157 460	4 106 366	9 964 169
t1_6	Santé	3 747	3 568	3 695	3 936	4 474	7 990	25 982	173 989
t1_7	Transports	83 418	72 368	71 034	98 893	150 669	439 952	1 757 435	17 151 519
t1_8	Communications	394	338	340	375	474	884	2 907	17 277
t1_9	Loisirs et culture	10 741	9 346	9 204	13 285	21 119	59 913	262 982	1 948 755
t1_10	Enseignement	1 296	1 091	1 269	1 336	1 500	2 163	5 732	34 627
t1_11	Restaurants et hôtels	15 382	13 577	13 273	18 934	31 250	83 997	377 408	2 162 747
t1_12	Autres biens et services	14 042	12 734	12 851	17 223	29 316	77 569	351 023	2 471 190

Méthode Logistique (LO=0.1,UP=30)		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	575 557	450 645	318 777	868 622	2 225 471	7 768 794	40 340 264	137 496 583
t1_1	Produits alimentaires et boissons non alcoolisées	5 639	4 368	4 930	5 857	7 979	17 072	88 731	268 588
t1_2	Boissons alcoolisées, tabac et stupéfiants	2 151	2 213	2 580	2 638	2 762	3 991	9 429	26 215
t1_3	Articles d'habillement et articles chaussants	8 755	7 732	7 891	11 023	18 308	45 127	205 565	702 169
t1_4	Logement, eau, électricité, gaz et autres combustibles	61 275	60 759	58 700	79 578	157 971	459 838	2 018 533	7 510 688
t1_5	Ameublement, équipement ménager et entretien courant de la maison	27 255	22 135	26 945	35 761	54 925	140 730	840 893	2 149 360
t1_6	Santé	3 696	3 598	3 852	4 104	4 873	8 002	24 961	87 033
t1_7	Transports	77 967	66 591	69 634	95 229	156 298	424 103	1 838 946	6 380 530
t1_8	Communications	399	349	352	381	494	866	3 039	10 734
t1_9	Loisirs et culture	10 894	9 392	9 338	13 087	21 666	57 010	274 942	973 084
t1_10	Enseignement	1 276	1 092	1 262	1 374	1 430	2 039	5 333	17 577
t1_11	Restaurants et hôtels	15 536	13 494	13 122	19 303	32 577	84 770	391 359	269 822
t1_12	Autres biens et services	13 604	12 285	12 394	17 767	29 754	77 455	369 435	1 197 100

3. EQM

Méthode Linéaire		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	5 870 803	448 936	2 031 865	2 485 523	3 608 036	9 775 498	33 374 267	214 502 896
t1_1	Produits alimentaires et boissons non alcoolisées	47 468	8 066	5 050	5 646	7 939	17 528	58 671	354 146
t1_2	Boissons alcoolisées, tabac et stupéfiants	2 167	2 113	2 765	2 817	2 905	4 146	7 485	36 131
t1_3	Articles d'habillement et articles chaussants	65 438	7 881	19 414	21 498	28 032	58 810	171 341	097 118
t1_4	Logement, eau, électricité, gaz et autres combustibles	156 289	69 402	107 272	128 103	184 406	471 993	1 622 416	10 300 641
t1_5	Ameublement, équipement ménager et entretien courant de la maison	96 738	21 412	56 683	64 298	77 672	167 010	513 187	3 081 697
t1_6	Santé	6 248	3 448	3 833	4 091	5 086	8 674	22 321	130 763
t1_7	Transports	346 700	67 866	190 443	208 845	265 749	588 890	1 772 051	11 019 105
t1_8	Communications	2 420	2 343	2 461	2 492	2 590	1 056	2 834	16 251
t1_9	Loisirs et culture	67 323	9 618	25 631	29 005	37 327	79 815	234 769	465 696
t1_10	Enseignement	1 440	1 136	1 267	1 375	1 581	2 193	4 778	24 840
t1_11	Restaurants et hôtels	49 576	14 129	38 540	43 199	53 648	117 241	350 907	2 185 269

t1_12	Autres biens et services	65 763	12 956	29 782	34 211	42 312	93 278	299 334	832 861	1
-------	--------------------------	-----------	-----------	-----------	-----------	-----------	-----------	---------	---------	---

Méthode Raking Ratio

		Taux de réponse uniforme	Calage simple	Calage généralisé						
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$	
t_00	Dépenses de consommation au total	5 922 399	468 625	1 759 707	2 312 161	3 657 041	9 860 322	54 031 907	303 275 719	
t1_1	Produits alimentaires et boissons non alcoolisées	47 135	7 581	5 235	6 043	8 190	18 130	86 455	382 425	
t1_2	Boissons alcoolisées, tabac et stupéfiants	2 157	2 164	2 763	2 814	2 973	4 170	13 057	65 271	
t1_3	Articles d'habillement et articles chaussants	65 520	8 341	17 823	20 766	26 062	56 131	198 895	087 876	2
t1_4	Logement, eau, électricité, gaz et autres combustibles	154 473	68 933	100 527	124 858	195 277	516 240	3 091 565	15 171 017	
t1_5	Ameublement, équipement ménager et entretien courant de la maison	100 314	22 763	53 455	64 224	87 631	196 012	4 187 892	10 333 592	
t1_6	Santé	7 160	3 678	4 089	4 324	4 765	8 371	26 213	174 616	
t1_7	Transports	360 378	74 584	175 949	203 176	259 861	564 901	1 843 554	17 475 349	
t1_8	Communications	2 485	3 338	4 454	4 489	4 575	8 000	2 968	17 355	
t1_9	Loisirs et culture	66 922	9 533	22 657	26 736	35 947	77 245	277 414	993 372	
t1_10	Enseignement	1 500	1 173	1 282	1 349	1 513	2 189	5 750	35 057	
t1_11	Restaurants et hôtels	47 939	13 993	33 996	39 662	51 735	107 666	397 463	203 999	
t1_12	Autres biens et services	64 910	12 760	25 485	29 744	42 936	93 890	366 463	523 660	2

Méthode Logistique (LO=0.1,UP=30)

		Taux de réponse uniforme	Calage simple	Calage généralisé					
				$\rho=0.8$	$\rho=0.6$	$\rho=0.4$	$\rho=0.2$	$\rho=0.1$	$\rho=0.05$
t_00	Dépenses de consommation au total	5 920 008	454 530	1 738 126	2 168 957	3 613 686	9 454 376	42 325 491	139 872 896
t1_1	Produits alimentaires et boissons non alcoolisées	46 057	7 743	4 984	5 949	8 008	17 081	88 731	268 607
t1_2	Boissons alcoolisées, tabac et stupéfiants	2 299	2 217	2 899	2 932	3 116	4 425	9 877	26 955
t1_3	Articles d'habillement et articles chaussants	65 926	7 827	17 343	19 802	27 292	55 694	217 184	716 820
t1_4	Logement, eau, électricité, gaz et autres combustibles	155 738	66 509	101 473	116 985	195 631	509 409	2 070 579	7 556 088
t1_5	Ameublement, équipement ménager et entretien courant de la maison	98 052	22 235	52 344	59 825	80 516	174 370	914 159	279 287
t1_6	Santé	6 942	3 729	4 201	4 405	5 191	8 389	25 217	87 224
t1_7	Transports	351 690	68 199	168 358	186 864	255 081	536 282	1 944 771	6 487 304
t1_8	Communications	2 482	3 349	4 454	4 473	5 598	8 985	3 129	10 835
t1_9	Loisirs et culture	67 236	9 579	22 585	25 436	35 915	73 288	292 211	994 934

t1_10	Enseignement	1	1	1	1	1	2		
		465	190	269	380	439	056	5 366	17 786
t1_11	Restaurants et hôtels	48	13	33	38	52	108		1
		575	970	834	664	604	193	416 502	295 182
t1_12	Autres biens et services	65	12	25	29	41	92		1
		315	298	142	508	790	135	387 503	220 189