

PRODUIRE DES STATISTIQUES ANNUELLES À PARTIR DE DONNÉES RAPPORTÉES SUR DE COURTES PÉRIODES DE TEMPS : L'EXPÉRIENCE DE L'ENQUÊTE SUR LES DÉPENSES DES MÉNAGES

Maryse Champagne¹

RÉSUMÉ

En 2010, l'Enquête sur les dépenses des ménages (EDM) a adopté un modèle de collecte international dans le but de réduire l'erreur de rappel et de diminuer l'effort du répondant. Le modèle est passé d'une période de rappel de 12 mois pour toutes les dépenses à des périodes de rappel plus courtes et variées selon le type de dépenses. Ces changements ont eu des impacts pour les utilisateurs: le pourcentage annuel de ménages avec une dépense et la moyenne annuelle des dépenses parmi les ménages déclarants ne sont plus diffusés. Diverses méthodes d'annualisation ont été évaluées afin de produire à nouveau ces statistiques d'intérêt pour permettre une utilisation plus complète des données de l'EDM et plus adaptée aux besoins des utilisateurs. L'article décrit une solution récemment développée qui utilise la minimisation du changement de comportement des consommateurs ainsi que la modélisation, via une régression segmentée.

MOTS CLÉS: Annualisation, minimisation, régression segmentée

ABSTRACT

Since 2010, the Survey of Household Spending (SHS) has adopted an international data collection model, aimed at reducing the recall error in expenditure data and the recall effort of respondents. It transitioned from a 12 month recall period for all items to a varying shorter recall period based on the type of expenditures. These changes have had impacts on SHS users: the annual proportion of households with a given expenditure and the annual average spending among reporting households are no longer published. Several annualization techniques have been evaluated to produce these statistics of interest once again, thus allowing a more complete use of the SHS data as well as being more adapted to the users' needs. The article describes a solution recently developed that relies on the minimization of the change in the consumers' patterns and on modelling, through segmented regression.

KEY WORDS: Annualization, minimization, segmented regression

1. INTRODUCTION

1.1 Contexte

L'Enquête sur les dépenses des ménages (EDM) est une enquête annuelle effectuée par Statistique Canada depuis 1997. Elle recueille de l'information détaillée sur les dépenses en biens et services effectuées par les ménages canadiens, ainsi que les caractéristiques démographiques du ménage, les caractéristiques du logement occupé par le ménage et des données sur le revenu des membres du ménage. En 2010, l'EDM a connu un remaniement lui faisant adopter le modèle de collecte international ayant pour objectif de réduire l'effort de rappel du répondant. Les longueurs des périodes de référence sont passées de 12 mois pour toutes les dépenses à des longueurs variées selon le type de dépense (1 mois, 3 mois, dernier paiement, 12 mois ou 2 semaines). En général, des périodes de référence plus longues sont utilisées pour des biens et services qui sont plus dispendieux ou pour des achats non fréquents ou irréguliers tel que l'achat d'un réfrigérateur. Par opposition, des périodes de référence plus courtes sont utilisées pour les biens et services moins dispendieux ou pour des achats fréquents ou qui sont effectués à intervalles réguliers, tel que l'achat de nourriture ou encore, les frais de garde d'enfants. Un journal des dépenses que les ménages sélectionnés complètent sur une période de deux semaines suivant l'entrevue personnelle a également été introduit. De plus, jusqu'en 2009, la collecte avait lieu entre janvier et mars de chaque année, et les ménages devaient déclarer les dépenses faites au cours de l'année de référence, soit l'année civile précédente. Depuis 2010, la collecte des données se fait de façon continue tout au long de l'année d'enquête où un douzième de l'échantillon

¹ Maryse Champagne, Statistique Canada, 150 Tunney's Pasture Dr, Ottawa, Ontario, K1A 0T6. Email : maryse.champagne@canada.ca

est collecté chaque mois. Les estimations des dépenses sont ensuite obtenues en combinant les données des 12 mois de collecte et en les annualisant selon les longueurs des périodes de référence pour standardiser les dépenses. Par exemple, une dépense rapportée sur une période de 2 semaines via le journal est multipliée par 26 puisqu'il y a 26 périodes de 2 semaines dans une année. Une détection des valeurs aberrantes est également effectuée afin de réduire le facteur d'annualisation associé à des dépenses extrêmes. Même si, de façon générale, les dépenses recueillies depuis 2010 sont similaires à celles des années précédentes, l'introduction des différentes périodes de référence a eu quelques impacts sur lesquels nous porterons une attention particulière, entre autres au niveau de l'interprétation des données et au niveau des statistiques annuelles par ménage déclarant.

A-Interprétation des données: Auparavant lorsqu'un ménage rapportait une dépense nulle, cela signifiait que le ménage n'avait pas effectué cette dépense au cours de l'année. De plus, lorsqu'une dépense de 2 600 \$ était rapportée, cela signifiait que c'était le montant (ou du moins sa meilleure estimation) que le ménage avait dépensé au cours de l'année. Par contre, depuis l'introduction des nouvelles périodes de référence, un ménage avec une dépense rapportée de 0 \$ signifie seulement que ce ménage n'a pas acheté ce bien ou ce service au cours de la période de référence de l'item (ex. 2 semaines). Toutefois, le ménage peut avoir effectué l'achat à un autre moment de l'année sans que l'enquête puisse le déterminer. Une dépense de 2 600 \$, cette fois, peut signifier simplement que 100 \$ ait été dépensé sur la période de référence de 2 semaines. Ce qui a été réellement dépensé au cours de l'année n'est pas connu (et ne peut être connu avec la structure actuelle de l'enquête).

B-Statistiques annuelles par ménage déclarant: De façon générale, les effets des périodes de référence plus courtes sont une proportion de ménages déclarants plus faible et une moyenne par ménage déclarant plus élevée. Cela s'explique par le fait que ce ne sont pas tous les ménages consommateurs qui achètent à chacune des périodes de référence de l'item. Voyons un exemple pour les dépenses en *Boissons alcoolisées achetées en magasin*. Le tableau 1.1 suivant présente les statistiques entre 2007 et 2015. Lorsque la période de référence était de 12 mois, soit avant 2010, la proportion de ménages consommateurs varie entre 76,14 % et 77,72 %. Cette proportion varie entre 33,28 % et 36,67 % lorsque la période de référence devient 2 semaines à partir de 2010. Les valeurs des dépenses sont ajustées en dollars constants de 2002² pour permettre des comparaisons au cours du temps. En ce qui concerne les moyennes des dépenses par ménage déclarant, on remarque l'effet inverse, soit qu'elles augmentent lorsque la période de référence diminue. Cet effet s'explique par le fait que, bien que les totaux des dépenses soient comparables au cours des cycles (tout comme les moyennes annuelles globales), ils sont maintenant divisés par moins de la moitié des ménages consommateurs.

Tableau 1.1 - Statistiques des dépenses en *Boissons alcoolisées achetées en magasin*

Année	Période de référence	Total annuel (en millions \$)	Moyenne globale (\$)	Proportion de ménages déclarants	Moyenne par ménage déclarant (\$)
2007	12 mois	7 042,95	543,82	77,72 %	699,72
2008	12 mois	7 334,24	557,13	76,33 %	729,89
2009	12 mois	7 270,00	543,26	76,14 %	713,46
2010	2 semaines	8 324,97	626,10	36,67 %	1 707,35
2011	2 semaines	7 920,57	587,64	35,94 %	1 635,12
2012	2 semaines	7 653,81	561,37	36,51 %	1 537,61
2013	2 semaines	8 583,56	622,35	35,60 %	1 748,22
2014	2 semaines	7 814,44	560,18	34,44 %	1 626,41
2015	2 semaines	8 162,14	578,98	33,28 %	1 740,45

Étant donné ce bris au niveau des statistiques par ménage déclarant, plusieurs utilisateurs des données de l'EDM ont limité leur emploi des variables à celles rapportées sur une période de 12 mois, car leur intérêt repose sur les dépenses à long terme et non pas sur celles rapportées sur une courte période de référence.

² Les indices d'inflation utilisés pour ajuster les dépenses proviennent du tableau 18-10-0005-01 de l'Indice des Prix à la Consommation publié par Statistique Canada

Il y a donc une incohérence entre la période d'intérêt analytique et la période de référence. Malheureusement, il n'est pas simple de remédier à cette incohérence avec la structure actuelle de l'enquête, car elle ne permet pas de connaître la variation des dépenses d'une période à l'autre (l'enquête consiste en une seule entrevue et aucune question n'est posée à propos de la fréquence des achats). Bien que plusieurs pays utilisent le même modèle de collecte que l'EDM, aucune solution n'a été apportée jusqu'à maintenant permettant d'obtenir les statistiques annuelles par ménage déclarant. Puisque les méthodes classiques d'estimations directes ne permettent pas de répondre à cette problématique, la méthode présentée dans cet article se tourne vers les méthodes d'estimations synthétiques telles que la modélisation, pour développer une méthodologie d'annualisation innovante.

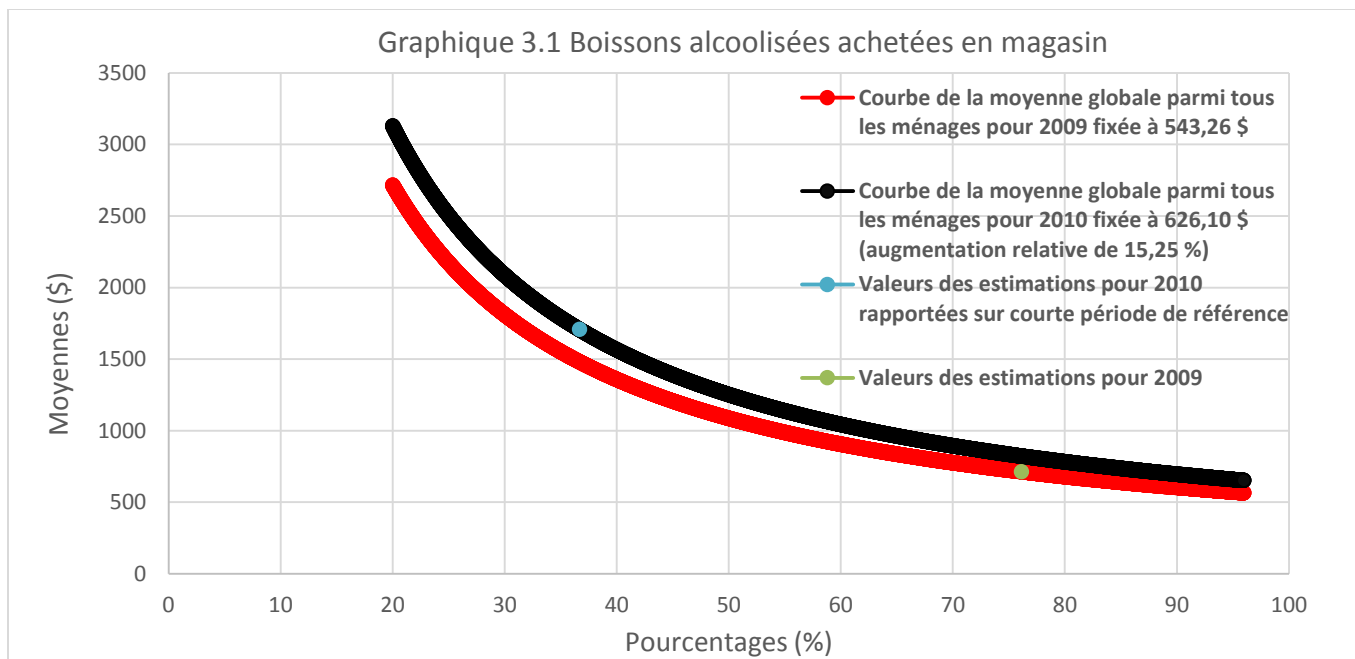
2. MÉTHODOLOGIE

2.1 Hypothèses

La standardisation des dépenses rapportées sur une période de référence plus courte (entre autres par l'annualisation) permet d'obtenir des statistiques agrégées (totaux et moyennes annuelles globales) qui sont toujours valides. De plus, par définition, on a que la moyenne annuelle globale équivaut à la proportion annuelle de ménages déclarants multipliée par la moyenne annuelle des dépenses par ménage déclarant. L'équation (2.1) suivante représente ce qui vient d'être mentionné:

$$\hat{p}_{annuelle}^{ménages\ déclarants} * \hat{x}_{annuelle}^{ménage\ déclarant} = \hat{x}_{annuelle}^{globale} \quad (2.1)$$

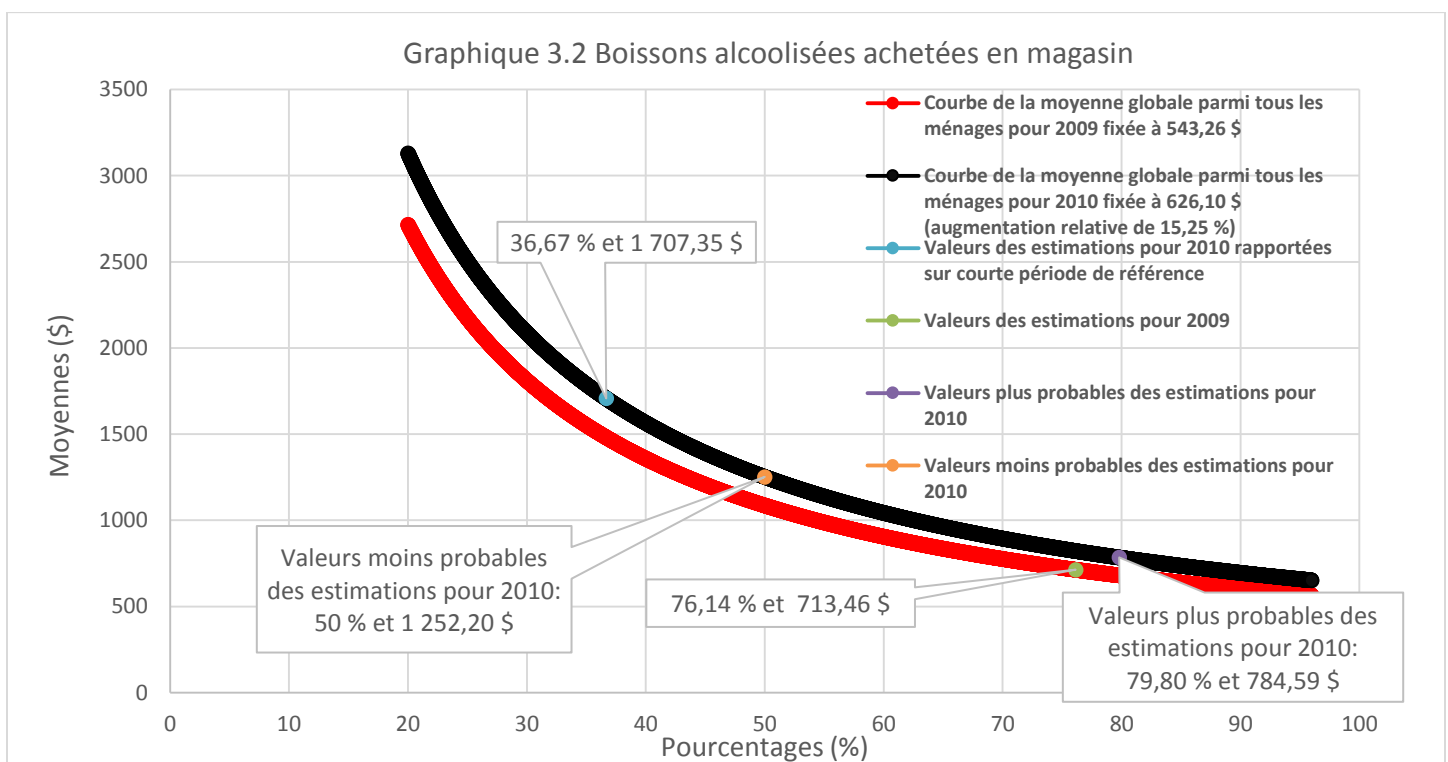
De cette relation, il existe donc un ensemble de solutions possibles pour la proportion annuelle de ménages déclarants ainsi que pour la moyenne annuelle par ménage déclarant. Le graphique 3.1 suivant représente l'ensemble des solutions possibles pour les dépenses en *Boissons alcoolisées achetées en magasin* en 2009 et en 2010:



L'axe horizontal représente la proportion de ménages déclarants (en pourcentage) et l'axe vertical représente la moyenne annuelle des dépenses par ménage déclarant (en dollars constants de 2002). La courbe rouge représente l'ensemble des solutions possibles, pour l'année 2009, qui respectent la moyenne annuelle globale estimée à 543,26 \$. Le point vert représente les valeurs des estimations, soit 76,1 % de ménages déclarants et 713,46 \$ de dépenses moyennes par ménage déclarant, pour 2009 et telles que rapportées sur une période de 12 mois. La courbe noire représente l'ensemble des solutions possibles, pour l'année 2010, qui respectent la moyenne annuelle globale estimée à 626,10 \$. Le point bleu représente les valeurs des estimations, soit 36,7 % de ménages déclarants et 1 707,35 \$ de dépenses moyennes annuelles par ménage déclarant, pour 2010 et telles que rapportées sur une période de 2 semaines. L'augmentation relative de la moyenne annuelle globale entre 2009 et 2010 est d'environ 15,25 % ce qui explique pourquoi la courbe de 2010 est légèrement au-dessus de celle de 2009. À noter qu'il est possible de construire ce genre de courbe pour tous les cycles de l'EDM. L'objectif est de

déterminer le point, sur la courbe des solutions possibles de 2010 (ou toute autre année post-remaniement), qui représente les estimations annuelles et non pas les estimations sur la période de référence. Pour développer la méthode et l'amener vers une solution plausible, il faut maintenant poser deux hypothèses:

- 1) Les estimations des proportions annuelles de ménages déclarants pour 2009 et avant sont valides. On suppose ici qu'un ménage se souvient s'il a ou s'il n'a pas effectué une dépense pour un bien ou un service durant les 12 derniers mois, soit la longueur de la période de référence pour toutes les dépenses avant le remaniement. Étant donné que les totaux sont considérés valides et la relation donnée par l'équation (2.1), on déduit que les moyennes annuelles par ménage déclarant pour 2009 et avant sont également valides. Ainsi, le point vert du graphique 3.1, soit les estimations annuelles de 2009, est un point de départ valide pour construire des estimations pour les années subséquentes.
- 2) La deuxième hypothèse est que le changement de comportement au sein d'un groupe de consommateurs, s'il y en a un, est minimal. Autrement dit, si une solution satisfait la contrainte et nécessite moins de changement de la part de la population qu'une autre solution, alors elle est plus probable de survenir. Par conséquent, c'est celle qui sera choisie pour produire les estimations annuelles. On cherchera donc à minimiser l'écart entre les statistiques, soit le pourcentage de ménages déclarants ainsi que la moyenne par ménage déclarant, pour deux années de données. L'impact de cette hypothèse est illustré par le graphique 3.2 suivant.



Sur le graphique 3.2, on retrouve les deux courbes de solutions possibles pour les cycles de 2009 et de 2010 où le point vert représente les valeurs telles qu'estimées pour 2009 qui seront utilisées comme référence. Une solution possible, pour 2010, est le point orange, c'est-à-dire 50 % de ménages déclarants et 1 252 \$ de dépenses en moyenne par ménage déclarant. Toutefois, cela impliquerait que le comportement des ménages aurait significativement changé en l'espace d'une seule année. Une solution plus plausible est un pourcentage de ménages déclarants de 79,8 % et 784 \$ de dépenses en moyenne par ménage déclarant. Les changements de comportement sous-jacents pour obtenir ces estimations sont minimes et donc plus probables.

2.2 Mesure de distance

En utilisant ces deux hypothèses, on suggère d'estimer les valeurs des deux statistiques pour 2010 en minimisant la distance entre les estimations de 2009 (le point vert) et la courbe de 2010. Pour ce faire, une mesure de distance qui prend en considération le fait que les deux axes ne sont pas de la même échelle doit être déterminée. Plusieurs définitions de distance

ont été essayées, notamment en standardisant les deux variables par leur erreur-type respective. Cependant, les estimations des proportions étant généralement de meilleure qualité que celles de la moyenne des dépenses, cette mesure de distance conservait les proportions de ménages déclarants pratiquement constantes, et tout le changement observé était absorbé par la moyenne. Cette mesure a donc été écartée. Une autre tentative a été d'utiliser la fonction de distance du khi-carré, mais puisque les valeurs de dépenses de la moyenne sont généralement d'une étendue plus large que les valeurs possibles des proportions, qui a un maximum à 100 %, la pénalité associée à la moyenne était souvent plus élevée simplement pour cette raison. Finalement, la mesure de distance retenue est basée sur le fait que les deux échelles permettent l'établissement de ratios: l'ordre de grandeur et les différences sur chaque échelle ne sont pas comparables, mais les ratios peuvent l'être. En effet, sur chacune des échelles, une quantité peut être le double ou encore la moitié d'une quantité précédente, ce qui permet de comparer les changements observés. Il y a également, pour chacune des échelles, la présence d'un zéro significatif (0 % de ménages déclarants et une moyenne de 0 \$ par ménage déclarant). La mesure de distance retenue est donc la distance euclidienne à laquelle le logarithme naturel a été ajouté pour permettre l'établissement de ratio. L'équation (2.2) suivante présente cette mesure :

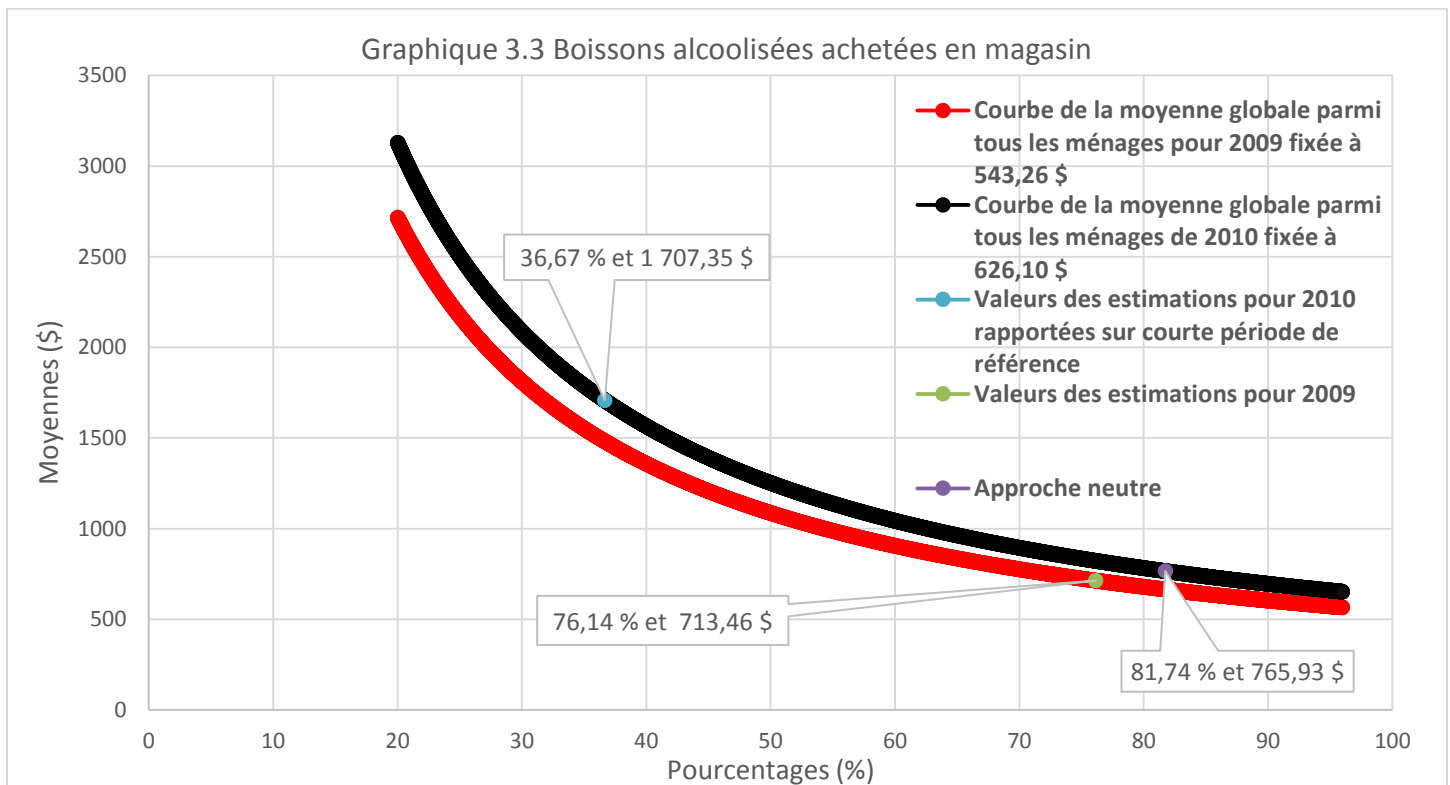
$$\text{Minimiser la fonction } \alpha(\ln(\hat{p}_{an2010}^{mén.décl.}) - \ln(\hat{p}_{an2009}^{mén.décl.}))^2 + (\ln(\hat{x}_{an2010}^{mén.décl.}) - \ln(\hat{x}_{an2009}^{mén.décl.}))^2 \quad (2.2)$$

par rapport à $(\hat{p}_{an2010}^{mén.décl.}, \hat{x}_{an2010}^{mén.décl.})$ sujet à la contrainte $\ln(\hat{p}_{an2010}^{mén.décl.}) + \ln(\hat{x}_{an2010}^{mén.décl.}) = \ln(\hat{x}_{an2010}^{globale})$

2.3 Approche neutre

Lorsque la constante α vaut 1, la fonction (2.2) donne une pénalité équivalente aux deux dimensions, le pourcentage et la moyenne, de sorte que chacune des variables subit le même niveau d'augmentation ou de diminution lors du déplacement vers la courbe de 2010. Il est possible de démontrer que le minimum est atteint lorsque les statistiques sont fixées à

$\hat{p}_{an2010}^{mén.décl.} = \hat{p}_{an2009}^{mén.décl.} * \sqrt{(\hat{x}_{an2010}^{globale} / \hat{x}_{an2009}^{globale})}$ et à $\hat{x}_{an2010}^{mén.décl.} = \hat{x}_{an2009}^{mén.décl.} * \sqrt{(\hat{x}_{an2010}^{globale} / \hat{x}_{an2009}^{globale})}$. On nommera cette fonction *l'approche neutre*. Voici un exemple à l'aide du graphique 3.3 suivant.



Étant donné une augmentation relative de 15,25 % par rapport à l'estimation de 2009, l'approche neutre assigne une augmentation relative de 7,35 % autant à la proportion annuelle de ménages déclarants qu'à la moyenne annuelle par ménage

déclarant pour obtenir les estimations de 2010. On obtient donc une proportion annuelle de ménages déclarants de 81,74 % et une moyenne annuelle par ménage déclarant de 765,93 \$ comme estimations pour 2010.

2.4 Approche orientée

Supposons que de l'information auxiliaire est disponible comme, par exemple, un avis d'expert sur la tendance de certains biens ou services ou encore des données provenant d'une source externe. Dans ce cas, on voudra que la fonction de distance en tienne compte en accordant plus de changement à une dimension plutôt que l'autre, selon l'information auxiliaire. Ceci peut être effectué en utilisant une valeur de α différente de 1. On voudra que la valeur de la constante α soit le reflet de la contribution relative de chaque dimension au changement qui a été subi par la moyenne globale. La définition de α est donc

$$\alpha = \ln\left(\frac{\hat{x}_{an,t+1}^{mén.décl.}}{\hat{x}_{an,t}^{mén.décl.}}\right) / \ln\left(\frac{\hat{p}_{an,t+1}^{mén.décl.}}{\hat{p}_{an,t}^{mén.décl.}}\right) \quad (2.3)$$

Par la définition (2.3), on peut voir que lorsque la constante α est supérieure à 1, un plus grand changement sera accordé à la moyenne. Par contre, si la valeur de α est inférieure à 1, un plus grand changement sera accordé à la proportion. Pour ces situations, on nommera la fonction de distance, *l'approche orientée*. Pour estimer la valeur de la constante α , on propose d'utiliser la régression segmentée pour modéliser les proportions de ménages déclarants (\hat{p}_t) et pour modéliser les moyennes par ménage déclarant (\hat{x}_t). La segmentation permet de tenir compte des périodes de référence qui sont différentes avant et après le remaniement de l'enquête en 2010. Les modèles de régression segmentée proposés sont les suivants:

$$\hat{p}_t = \alpha_1 I_{t \leq 2009} + \beta_1 I_{t \leq 2009}(t - 2010) + \alpha_2 I_{t > 2009} + \beta_2 I_{t > 2009}(t - 2010) + \varepsilon_t \quad (2.4)$$

$$\hat{x}_t = \alpha_1^* I_{t \leq 2009} + \beta_1^* I_{t \leq 2009}(t - 2010) + \alpha_2^* I_{t > 2009} + \beta_2^* I_{t > 2009}(t - 2010) + \delta_t \quad (2.5)$$

À partir des paramètres estimés issus des modèles, il existe plusieurs façons de dériver la valeur de $\hat{\alpha}$. Selon nos observations, une hypothèse qui donne des résultats probables est de supposer que la tendance observée avant 2009 se poursuive au moins jusqu'en 2010. Ainsi, le ratio du changement relatif subi par les deux dimensions peut s'estimer par:

$$\hat{c}_1^* = \frac{\hat{p}_{2010} / \hat{p}_{2009}}{\hat{x}_{2010} / \hat{x}_{2009}} = \frac{\hat{\alpha}_1 / (\hat{\alpha}_1 - \hat{\beta}_1)}{\hat{\alpha}_1^* / (\hat{\alpha}_1^* - \hat{\beta}_1^*)} \quad (2.6)$$

Puisque les deux modèles qui permettent d'obtenir la valeur de \hat{c}_1^* sont appliqués indépendamment, il est nécessaire d'ajuster \hat{c}_1^* pour respecter la contrainte (2.1). L'équation suivante, dérivée de (2.3) et (2.6), permet d'obtenir $\hat{\alpha}$:

$$\hat{\alpha} = \frac{(\ln(\hat{x}_{an,2010}^{globale} / \hat{x}_{an,2009}^{globale}) - \ln(\hat{c}_1^*))}{(\ln(\hat{x}_{an,2010}^{globale} / \hat{x}_{an,2009}^{globale}) + \ln(\hat{c}_1^*))} \quad (2.7)$$

Par exemple, supposons que le résultat des modèles de régression indique qu'il y a eu 1,1 fois plus de changement au niveau des proportions qu'au niveau des moyennes (c'est-à-dire que $\hat{c}_1^* = 1,1$) et que la moyenne annuelle globale a augmenté de 15 % entre 2009 et 2010. En appliquant l'équation (2.7), on trouve une valeur $\hat{\alpha}$ de 0,1890. La valeur de $\hat{\alpha}$ étant inférieure à 1, un plus grand changement sera accordé au niveau des proportions. Il y aura donc une augmentation de 12,47 % assignée à la proportion de 2009 et une augmentation de 2,247 % assignée à la moyenne de 2009 pour obtenir les estimations de 2010.

Par la suite, il est possible d'obtenir les estimations des années subséquentes en utilisant le même procédé. Toutefois, cela voudra dire que la tendance observée entre 2004 et 2009 se poursuit au-delà de plusieurs années ce qui pourrait ne pas être le cas plus nous nous éloignons de 2009. Pour les années après 2010, nous connaissons toutefois les estimations sur les courtes périodes de référence et nous supposons que cette tendance reflète ce qui se produit au niveau annuel. Pour les années après 2010, par exemple pour 2011, le ratio du changement respectif sera estimé par :

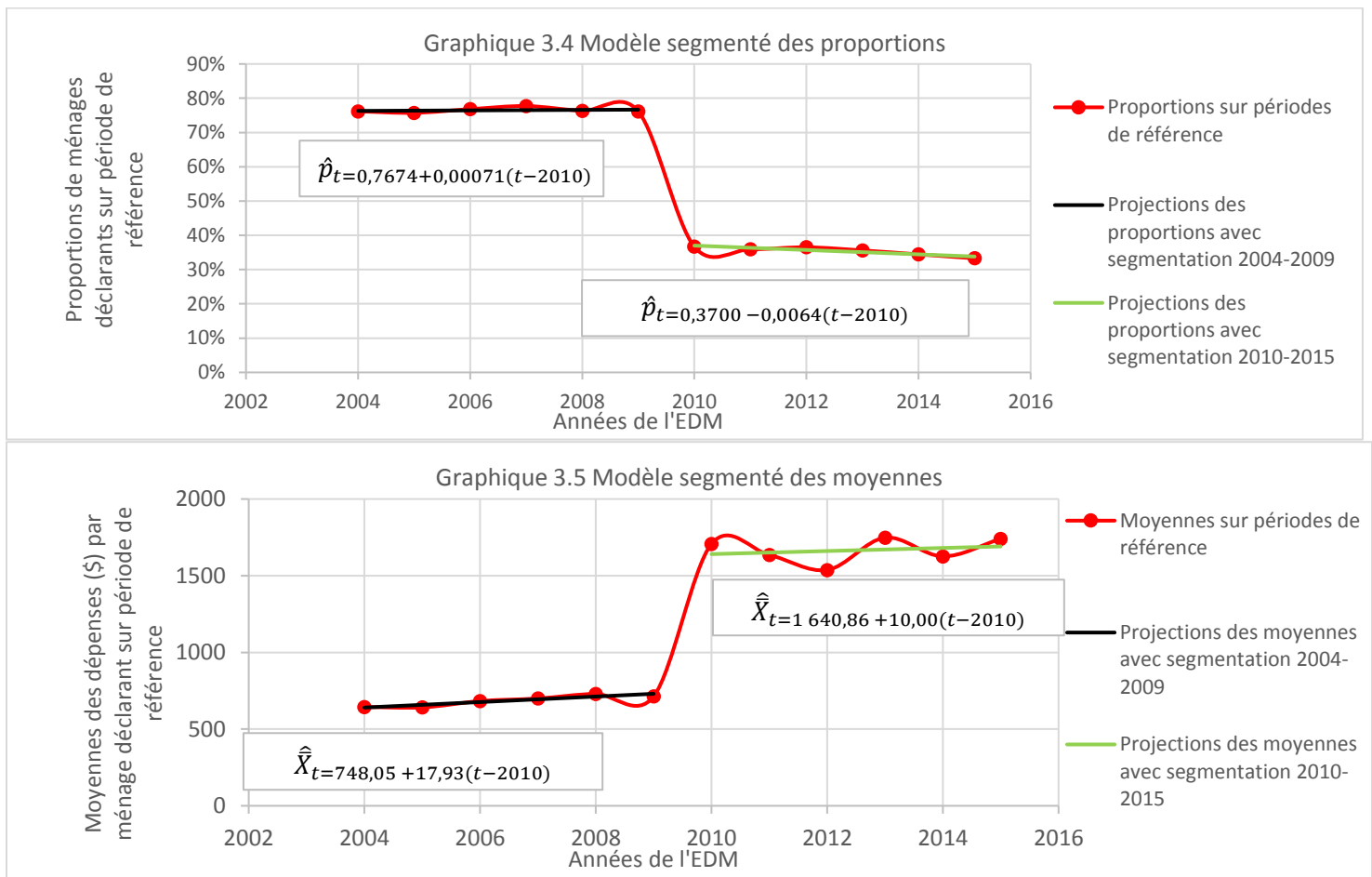
$$\hat{c}_1^{**} = \frac{\hat{p}_{2011}/\hat{p}_{2010}}{\hat{x}_{2011}/\hat{x}_{2010}} = \frac{(\hat{\alpha}_2 + \hat{\beta}_2)/\hat{\alpha}_2}{(\hat{\alpha}_2^* + \hat{\beta}_2^*)/\hat{\alpha}_2^*} \quad (2.8)$$

La valeur $\hat{\alpha}$ s'obtient ensuite de cette façon:

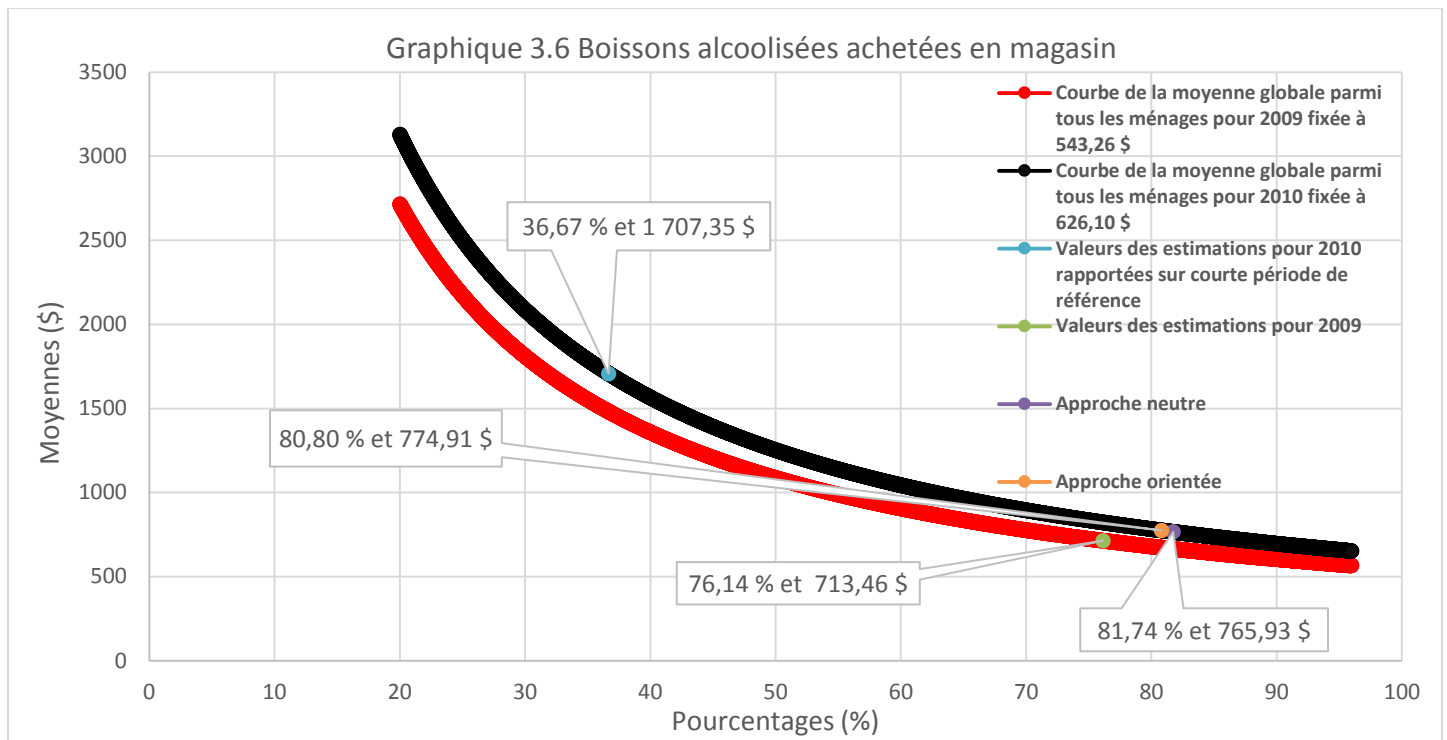
$$\hat{\alpha} = \frac{(\ln(\hat{x}_{an2011}^{globale}/\hat{x}_{an2010}^{globale}) - \ln(\hat{c}_1^{**}))}{(\ln(\hat{x}_{an2011}^{globale}/\hat{x}_{an2010}^{globale}) + \ln(\hat{c}_1^{**}))} \quad (2.9)$$

2.4.1 Application de l'approche orientée

Les estimations des dépenses en *Boissons alcoolisées achetées en magasin* avant et après le remaniement sont utilisées pour appliquer les modèles de régression segmentée tels que vus précédemment. Les équations des tendances obtenues sont illustrées dans les deux prochains graphiques 3.4 et 3.5.



À partir des paramètres issus des deux modèles, et en utilisant l'équation (2.6), on déduit que la constante \hat{c}_1^* vaut 0,9769. Ensuite, en appliquant l'équation (2.7), on déduit la valeur de $\hat{\alpha}$, qui est égale à 1,3934. Cette valeur, étant légèrement supérieure à 1, indique qu'il y a eu un plus grand changement au niveau de la moyenne par ménage déclarant qu'au niveau de la proportion de ménages déclarants. Il ne reste qu'à appliquer l'équation (2.2) sous la contrainte de respecter l'estimation de la moyenne annuelle globale de 2010. Sous l'approche orientée, les estimations de la proportion annuelle de ménages déclarants et de la moyenne annuelle des dépenses par ménage déclarant pour 2010 sont 80,80 % et 774,91 \$. Le graphique 3.6 suivant illustre les résultats obtenus selon les approches neutre et orientée.



On remarque que l'approche orientée décale légèrement les estimations vers la gauche par rapport aux estimations de l'approche neutre étant donné qu'un changement plus important au niveau des moyennes qu'au niveau des proportions a été déduit de la modélisation. Pour obtenir les estimations pour les années subséquentes, il suffit d'utiliser le ratio \hat{c}_1^{**} (2.8) avec les paramètres issus du modèle segmenté 2010-2015 ainsi que les valeurs précédemment trouvées pour 2010. Le tableau 1.2 suivant présente les proportions de ménages déclarants pour les années 2010 à 2015 obtenues selon les deux approches (l'année 2009 est présentée à titre de référence).

Tableau 1.2 – Résultats pour *Boissons alcoolisées achetées en magasin*

Année	Moyenne annuelle globale (\$)	Proportion sur période de référence	Proportion annuelle <i>Approche Neutre</i>	Proportion annuelle <i>Approche Orientée</i>
2009	543,26	76,14 %	76,14 %	76,14 %
2010	626,10	36,67 %	81,74 %	80,80 %
2011	587,64	35,94 %	79,19 %	77,36 %
2012	561,37	36,51 %	77,40 %	74,72 %
2013	622,35	35,60 %	81,50 %	77,74 %
2014	560,18	34,44 %	77,32 %	72,86 %
2015	578,98	33,28 %	78,63 %	73,19 %

2.5 Approche orientée proportionnelle

L'obtention de \hat{c}_1^{**} par l'équation (2.8) permet d'utiliser le ratio des changements observés sur les courtes périodes de référence comme étant indicateur de ce qui se produit également sur la période annuelle. Par contre, cette hypothèse est possiblement trop forte. En effet, elle n'est valide que si les ménages consommateurs achètent à chacune des périodes de référence du bien ou du service et si les ménages non-consommateurs sur la courte période sont aussi des ménages non-consommateurs annuels. Pour certains types de dépenses, il existe des ménages consommateurs occasionnels, c'est-à-dire qui n'achètent pas à chacune des périodes de référence; et la fréquence à laquelle ces ménages achètent peut avoir un grand impact sur la proportion de ménages déclarants sur courte période. Supposons qu'il n'existe que cinq types de

comportements d'achats dans la population et qu'ils soient représentés de façon équivalente. Supposons également que la période de référence pour un certain bien est de 3 mois ce qui donne 4 périodes d'achat durant l'année. Dans le tableau 1.3 suivant, un '1' signifie qu'un achat a été effectué et un '0' signifie qu'aucun achat n'a été effectué durant la période de référence de 3 mois. Les quatre ménages consommateurs achètent à une seule période de référence et le cinquième ménage n'est pas un ménage consommateur.

Tableau 1.3 - Simulation de comportements d'achat dans une population

Ménages	Période 1	Période 2	Période 3	Période 4	Statut de consommateur annuel
1	1	0	0	0	1
2	0	1	0	0	1
3	0	0	1	0	1
4	0	0	0	1	1
5	0	0	0	0	0
Estimations	20 %	20 %	20 %	20 %	80 %

Supposons maintenant que trois des quatre ménages consommateurs achètent plus souvent mais toujours au même niveau (même montant dépensé). Ceci se traduit par une augmentation de la proportion de ménages déclarants sur une courte période mais pas en une augmentation de la proportion annuelle de ménages déclarants tel que l'on peut le voir dans le tableau 1.3.1 suivant.

Tableau 1.3.1 - Suite de la simulation de comportements d'achat dans une population

Ménages	Période 1	Période 2	Période 3	Période 4	Statut de consommateur annuel
1	1	0	1	0	1
2	0	1	0	1	1
3	1	1	1	0	1
4	0	0	0	1	1
5	0	0	0	0	0
Estimations	40 %	40 %	40 %	40 %	80 %

Dans cet exemple particulier, \hat{c}_1^{**} serait artificiellement gonflé et attribuerait une partie de l'augmentation observée dans la moyenne annuelle globale à la proportion de ménages déclarants lorsqu'en fait la totalité de l'augmentation observée devrait être allouée à la moyenne par ménage déclarant. Dans cet exemple, quatre des cinq ménages sont des ménages consommateurs occasionnels qui sont, au final, des ménages consommateurs annuels au même titre que les ménages qui achètent à chacune des périodes de référence. Toutefois, un changement de comportement de ces ménages consommateurs occasionnels peut avoir un impact sur la performance de la méthodologie d'annualisation.

Une façon de diminuer cet impact est de supposer que l'augmentation observée dans les proportions de ménages déclarants sur courte période est distribuée entre les proportions d'acheteurs occasionnels et les non-acheteurs (qui deviennent des ménages consommateurs). N'ayant aucune information sur la distribution, nous suggérons d'utiliser une allocation proportionnelle entre les proportions d'acheteurs occasionnels et les non-acheteurs (d'autres fonctions d'allocation pourraient être considérées). La valeur de $\hat{\beta}_2$ (paramètre reflétant la pente des proportions d'acheteurs post-remaniement) est ainsi révisée pour mieux capturer le changement au niveau des ménages non-acheteurs qui sont devenus consommateurs :

$$\hat{\beta}_2^{révisée} = \left(\frac{P \text{ de non-acheteurs annuels}}{P \text{ de non-acheteurs annuels} + P \text{ acheteurs occasionnels}} \right) * \hat{\beta}_2.$$

L'estimation du pourcentage d'acheteurs occasionnels s'obtient de l'écart entre le pourcentage d'acheteurs annuels et le pourcentage d'acheteurs sur la courte période de référence. Ceci mène également à des valeurs révisées de \hat{c}_1^{**} et $\hat{\alpha}$. Ce raisonnement est aussi valide si une diminution des achats est observée; la répartition se ferait alors entre les acheteurs annuels et les acheteurs occasionnels.

3. CONCLUSION ET TRAVAUX FUTURS

Cette méthodologie présente plusieurs avantages, notamment celui de mieux répondre aux besoins des utilisateurs de données. En effet, il est maintenant possible de leur offrir les mêmes statistiques qu'avant le remaniement de 2010. Toutefois, il est important de les informer de la méthodologie utilisée et du fait que ces statistiques sont synthétiques. Un autre grand avantage de cette méthode est qu'elle fonctionne peu importe la longueur de la période de référence. La méthode s'applique aussi à des domaines, bien que l'additivité des domaines ne soit pas garantie. Elle permet également l'utilisation d'information auxiliaire. Par exemple, il est possible d'utiliser le total provenant d'une source externe advenant l'existence d'une telle source ou encore, si l'utilisateur possède des connaissances de l'industrie, il lui est possible de spécifier la valeur de la constante α . Cependant, cette méthode présente aussi quelques inconvénients. Il est difficile d'en évaluer la performance puisqu'il n'existe pas, ou pratiquement pas, de données de confrontation. De plus, elle repose sur de fortes hypothèses. Toutefois, une façon de valider les modèles et les hypothèses posées serait d'effectuer une enquête à petite échelle, de fréquence sporadique, qui se concentrerait sur les dépenses les plus importantes. Ainsi, il serait possible de valider les tendances obtenues avec la modélisation segmentée.

Les travaux futurs consistent à évaluer la méthodologie sur l'ensemble des variables de dépenses de l'EDM pour déterminer si elles peuvent être utilisées telles quelles ou si quelques modifications devraient d'abord être effectuées. Finalement, la méthode proposée ne répond pas aux questions reliées à la distribution des dépenses telles que '*Quels sont les ménages consommateurs ?*' ni '*Combien chaque ménage dépense-t-il précisément ?*'. Les projets futurs incluent de tenter de développer une méthodologie facilitant l'annualisation des distributions des dépenses permettant de répondre à ces questions.

RÉFÉRENCES

- Champagne, M. et Lafortune, Y. (2017). *Stratégies pour estimer le pourcentage annuel de ménages déclarants et la dépense annuelle moyenne par ménage déclarant dans l'Enquête sur les dépenses des ménages*, DMEM document interne.
- Champagne, M. (2018). *Guide de l'utilisateur pour obtenir les statistiques annuelles de l'Enquête sur les dépenses des ménages – remaniée (EDM-R)*, DMEM document interne.
- Frazis, H. et Stewart, J. (2010). *How to Think About Time-Use Data: What Inferences Can We Make About Long- and Short-Run Time Use from Time Diaries?*, U.S. Bureau of Labor Statistics, Working Paper 442.