Application d'un modèle à classes latentes aux achats en grande distribution et comparaison avec la formulation Pareto/NBD

Application of latent class models to purchases in the retailing sector and comparison with the Pareto/NBD formulation

Herbert CASTERAN, Groupe ESC PAU & Université Toulouse III Paul Sabatier, EA 2043, Laboratoire Gestion & Cognition (LGC). Postal address: ESC PAU, 3 r St John Perse, BP 7512, 64 075 Pau Cedex – FRANCE; E-Mail: herbert.casteran@esc-pau.fr, Tel: +33(0)559.92.64.97

Lars MEYER-WAARDEN, Université Toulouse III Paul Sabatier, EA 2043, Laboratoire Gestion & Cognition (LGC). Postal address: IUT-GEA Ponsan, 115E route de Narbonne BP 67701, 31077 Toulouse cedex 4 ; E-Mail: lars.meyer-waarden@iut-tlse3.fr, Tel : +33(0) 6 80 37 42 08

Christophe BENAVENT, Université Paris X, CEROS. Postal address: Université Paris X- Nanterre, CEROS, 200 avenue de la république, 92001 NANTERRE cedex ; E-Mail: christophe.benavent@wanadoo.fr; Tel : +33(0) 6 23 23 62 056

Contact :

Herbert CASTERAN, Groupe ESC PAU, IRMAPE & Université Toulouse III Paul Sabatier, EA 2043, Laboratoire Gestion & Cognition (LGC). Postal address: ESC PAU, 3 r St John Perse, BP 7512, 64 075 Pau Cedex – FRANCE; E-Mail: herbert.casteran@esc-pau.fr, Tel: +33(0)559.92.64.97

Application de modèles à classes latentes Pareto/NBD à des achats en grande distribution

Résumé :

L'étude des transactions dans un cadre non contractuel s'opère de plus en plus d'après des modèles mixtes continus. Or leurs implications managériales sont limitées du fait de leur nature exclusivement stochastique.

Les formulations mixtes finies offrent une plus grande portée opérationnelle du fait d'une segmentation et de variables explicatives. Nous comparons les conditions d'implémentation et les implications managériales d'un modèle Pareto/NBD (mixte continu) et d'un modèle à classes latentes (mixte fini) sur des données de panel.

Au-delà d'un aspect opérationnel sans surprise plus abouti, les modèles à classes latentes présentent, jusqu'à un délai long (75 semaines), un meilleur ajustement et une mise en œuvre incomparablement plus aisée.

Mots-clefs : données de panel, modèle mixte fini, modèle Pareto/NBD

Application of latent classes models to purchases in the retailing sector and comparison with the Pareto/NBD formulation

Abstract: the transactions in a "noncontractual" setting are treated more and more with continuous mixture models. However, their managerial implications are limited due to the parametric distributions of the parameters and the lack of explanatory variables. If generalizations are always possible, they are however very complex.

The finite mixture models give a more operational point of view due to segmentation and explanatory variables. Moreover, they are intrinsically semi-parametric.

We compare the conditions of implementation and managerial implications of a Pareto/NBD model and a latent class model (finite mixture) on panel data.

Keywords: panel data, finite mixture model, Pareto/NBD model

INTRODUCTION

Les transactions dans un cadre non contractuel font l'objet d'une littérature de plus en plus abondante (Ayache A. et alii, 2006, Villanueva J. et Hansens D., 2007). Largement présentes dans le monde professionnel, elles sont généralement traitées par le biais des données de comptage avec ou non intégration d'une notion d'inactivité. Ces données servent de base à la mesure de la valeur actualisée client et du capital client.

Dans ce cadre, les modèles mixtes continus sont considérés comme une des voies de recherche les plus prometteuses et plus spécialement la formalisation négative binomiale (modèle NBD). La distribution Poisson des données y est mixée avec une distribution gamma de la fréquence d'achat. Cette approche, initiée par Ehrenberg (1959), a été enrichie par une prise en compte du phénomène d'inactivité : modèle Pareto/NBD (Schmittlein, Morrison et Colombo, 1987 ; Schmittlein et Peterson, 1994 ; Crié, 1999) ou modèle bêta-géométrique/NBD (BG/NBD par Fader, Hardie et Lok Lee, 2005). Ainsi le comportement des consommateurs est représenté par une formalisation continue permettant, en théorie, la prise en compte de l'ensemble des spécificités individuelles. Cette possibilité d'une approche individuelle des comportements apparaît séduisante.

Toutefois les modèles mixtes continus supposent une spécification paramétrique complète (généralement Poisson avec une distribution spécifique du paramètre de fréquence) par nature restrictive et bien souvent inadaptée eu égard aux hypothèses fondamentales de ces distributions. Des généralisations semi ou non-paramétriques sont naturellement possibles. Mais leur introduction ne peut s'opérer qu'au profit d'une conceptualisation mathématique lourde. De même, l'introduction de variables explicatives est également possible mais toujours au prix d'une formalisation mathématique exigeante (Castéran, Meyer-Waarden et Bénavent, 2007a). De ce fait la portée opérationnelle et managériale de ces modèles est considérablement réduite.

Les modèles mixtes finis sont eux d'un usage déjà ancien. Les premiers principes ont été posés par Newcomb (1886) et Pearson (1894). Les modèles mixtes finis constituent un cas particulier de ces modèles à classes latentes (Baltagi, 2003). Ils postulent l'existence de classes latentes au sein de la population étudiée et d'un lien spécifique, au sein de chacune de ces classes, entre variables expliquée et explicatives. A ce titre ils sous-tendent l'existence de segments avec des logiques comportementales spécifiques ; les implications marketing apparaissent ici clairement.

Toutefois, les applications dans un cadre spécifiquement marketing ont été initiées tardivement, principalement par Wedel et alii (1993). Elles permettent une segmentation de la population, au-delà de la segmentation comportementale classique. Une segmentation, si elle semble ne pas offrir la même finesse d'analyse qu'une approche continue, permet une interprétation claire des résultats obtenus. Elle fournit des implications managériales et opérationnelles directement accessibles. Ces implications sont renforcées par la présence de variables explicatives. Chaque segment peut être étudié en fonction de ses caractéristiques comportementales propres, celles-ci étant expliquées par des variables explicatives. Ces variables explicatives permettent de déterminer les actions marketing les plus efficaces à l'échelle de chaque segment. Les modèles mixtes finis apparaissent ainsi comme une alternative prometteuse aux modèles mixtes continus.

Pour autant, l'efficacité comparative de ces modèles n'a jamais été établie. En particulier, aucune comparaison en termes de validité prédictive n'a été opérée entre les modèles mixtes finis et les modèles de la famille NBD (NBD simple, Pareto/NBD, BG/NBD). Nous nous livrons donc ici à une telle comparaison sur des données issues du panel Behavior Scan dans le cadre de la grande distribution alimentaire. Nous retenons comme base de comparaison pour les modèles mixtes continus, le modèle Pareto/NBD. Ce choix est dicté par la meilleure adéquation de ce modèle avec ces données par rapport au modèle NBD basique et BG/NBD (Castéran, Meyer-Waarden et Bénavent, 2007b).

La première partie de ce document présente les principes théoriques de ces modèles : fondements et méthodes d'estimation. La seconde partie fournit les résultats des différentes estimations et met en comparaison les résultats tant sous l'angle des indicateurs traditionnels que de leur ajustement réel avec les données. Enfin la troisième partie dégage les principales conclusions et les axes de recherche.

LES MODELES EN PRESENCE

Le modèle Pareto/NBD

Fondements

Ce modèle a été développé en 1987 par Schmittlein, Morrison et Colombo. Il obéit à cinq hypothèses.

1. Le nombre de transactions Y durant une période de durée t (entre une date origine fixée à 0 et la date t) est réparti aléatoirement dans le temps suivant une distribution Poisson :

$$P(Y = y) = \frac{(\lambda t)^y}{y!} \mathrm{e}^{-\lambda t}, \; y=0, 1, 2\ldots \quad (1)$$

2. L'hétérogénéité des clients en termes de fréquence d'achat est retranscrite : le taux λ se répartit de façon inégale entre les clients. Cette hétérogénéité est représentée par une fonction gamma de densité :

$$f(\lambda) = \frac{\alpha^r}{\Gamma(r)} \lambda^{r-1} e^{-\alpha\lambda} \text{ avec } \lambda>0, r>0, \alpha>0 \quad (2)$$

avec $\Gamma(r)$ qui est la fonction gamma de paramètre r. Le choix de la fonction gamma tient à sa grande souplesse d'usage.

3. La durée de vie des clients : au delà d'un certain délai, le client n'est plus actif, quelle qu'en soit la raison. Cette inactivité du client intervient aléatoirement selon un taux μ. La durée de vie τ d'un client est une variable aléatoire exponentielle de densité :

$$g(\tau) = \mu e^{-\mu\tau} \quad (3)$$

4. Là encore, il faut représenter l'hétérogénéité inter-individuelle de l'inactivité. Elle se traduit, de la même manière que pour la fréquence d'achat, par une distribution gamma du paramètre μ :

$$f(\mu) = \frac{\beta^s}{\Gamma(s)} \mu^{s-1} e^{-\beta\mu} \text{ avec } \mu>0, s>0, \beta>0 \quad (4)$$

Une propriété importante de ces formulations est l'indépendance entre fréquence d'achat λ et taux d'inactivité μ. La date τ à laquelle un client devient inactif suit une distribution exponentielle avec les paramètres suivants : $E(\tau|\mu)=1/\mu$ et $V(\tau|\mu)=1/\mu^2$.

L'estimation s'opère d'après le critère du maximum de la log-vraisemblance. La formule est présentée en annexe 1.

Compte tenu des cinq hypothèses précédentes, le nombre d'achats effectués durant une période]0 ; T] devient :

$$E(Y|r,\alpha,s,T) = \frac{r\beta}{\alpha(s-1)}\left[1-\left(\frac{\beta}{\beta+T}\right)^{s-1}\right] \quad (5)$$

Il s'agit ici d'une prévision agrégée de la demande. La date 0 représente le premier achat réalisé par chaque client. Il faut donc sommer par date de premier achat l'ensemble des espérances individuelles. Le corollaire est la nécessaire identification des dates réelles d'activité de chaque client, faute de quoi la somme sera, au moins dans un premier temps, sous-estimée. Cette condition est bien entendu délicate dans un cadre non contractuel.

L'espérance conditionnelle des achats futurs *Y** réalisés par un individu donné dans une période à venir de durée *t* s'écrit :

$$E(Y^*|r,\alpha,s,\beta,Y=y,t,T)=E(Y|r+x,\alpha+T,s,T+\beta,t)\ P(\tau>T|r,\alpha,s,\beta,Y=y,t,T) \quad (6)$$

$P(\tau>T|r,\alpha,s,\beta,Y=y,t,T)$ représente la probabilité que le client soit encore actif après une date *T*. L'expression de cette probabilité se trouve en annexe 2.

Intérêts de la formulation

Le modèle Pareto/NBD aboutit à une approche probabiliste raffinée avec un corpus d'hypothèses restreint. Cependant, il présente des difficultés pratiques d'implémentation dans les bases de données. Malgré l'étendue des applications potentielles, la très grande complexité de l'estimation des paramètres (fonctions hypergéométrique, log-gamma) constitue un facteur de blocage majeur. D'autre part, de nature purement stochastique, il ne permet pas une interprétation du rôle des facteurs personnels et des actions marketing.

Le modèle à classes latentes

Fondements

De la même façon que dans le cadre Pareto/NBD, le nombre de transactions *Y* est supposé suivre une loi de Poisson de paramètre λ. Cependant ces paramètres varient conditionnellement à une distribution sur l'échantillon. Cette distribution est supposée être discrète ce qui conduit à une formulation mixte finie.

Il existe ainsi *C* classes latentes inconnues ; chaque individu appartient à une classe et à une seule, inconnue a priori. La probabilité a priori d'appartenance à une classe *c* se note π_c avec $\sum_{c=1}^{C}\pi_c=1$.

La formulation Poisson s'exprime de la manière suivante pour une durée t_i et pour chaque classe *c* et individu *i* :

$$P(y_i|\lambda_{i|c},t_i)=\frac{e^{-\lambda_{i|c}t_i}\left(\lambda_{i|c}t_i\right)^{y_i}}{y_i!} \quad (7)$$

y_i et $\lambda_{i|c}$ représentent respectivement un nombre d'achats pour le consommateur *i* et le paramètre de la loi de Poisson pour le même individu conditionnellement à son appartenance à la classe *c*. Naturellement ces éléments peuvent être régressés par rapport à des variables

explicatives représentées par les vecteurs x_1 et x_2. Ainsi la fréquence d'achat moyenne devient :

$$\lambda_{i|c} = \beta_{0c} \exp(\beta_{1c} x_1) \tag{8}$$

avec β_{0c} constante. Classiquement, le modèle multinomial logit est retenu pour le calcul des probabilités d'appartenance à chaque classe c :

$$\pi_c\left(x_2, \beta_{2c}\right) = \frac{\exp(k_c + \beta_{2c} x_2)}{\sum_{j=1}^{C} \exp(k_j + \beta_{2j} x_2)} \text{ avec } k_j \text{ constante.} \tag{9}$$

La densité mixte s'écrit donc de la manière suivante :

$$f(y|x_1, x_2, \Theta) = \sum_{c=1}^{C} \pi_c(x_2, \beta_{2c}) \mathrm{P}\left(y|\lambda_c(x_1)\right) \tag{10}$$

où $P\left(y|\lambda_c(x_1)\right)$ est la distribution Poisson conditionnellement aux variables explicatives et Θ les paramètres.

Dans une optique bayésienne, les π_c peuvent être vues comme des probabilités a priori d'appartenance à une classe. Elles sont agrégées. Les probabilités a posteriori sont les probabilités qu'a un individu i d'appartenir à une classe c.

$$P(c_i|y_i, x_{1i}, x_{2i}, \Theta) = \frac{\pi_c P\left(y_i|\lambda_{ci}(x_{1i})\right)}{\sum_{j=1}^{C} \pi_j P\left(y_i|\lambda_{ji}(x_{1i})\right)} \tag{11}$$

L'estimation de Θ s'effectue par maximisation de la log-vraisemblance. Celle-ci s'écrit avec N le nombre d'observations :

$$LL = \sum_{n=1}^{N} \log\left[f(y_n|x_{1n}, x_{2n}, \Theta)\right] = \sum_{n=1}^{N} \log\left[\sum_{c=1}^{C} \pi_c(x_{2n}, \beta_2) P\left(y_n|\lambda_c(x_{1n})\right)\right] \tag{12}$$

On applique pour ce faire l'algorithme itératif EM (Dempster, Laird et Rubin, 1977). Cet algorithme se décline en deux temps. Le premier temps réside dans l'estimation (étape E) des probabilités a posteriori d'affectation par classes (à un niveau individuel) $\hat{p}_{ic} = P(c_i|y_i, x_{1i}, x_{2i}, \Theta)$ en utilisant (11). Les probabilités a priori sont obtenues par $\pi_c = \frac{1}{N}\sum_{i=1}^{N} \hat{p}_{ic}$. Le second temps consiste en une maximisation (étape M) de la log-vraisemblance pour chaque classe en utilisant les probabilités a posteriori comme poids :

$$\max_{\Theta_c} \sum_{n=1}^{N} \hat{p}_{nc} \log\left[f(y_n|x_{1n}, x_{2n}, \Theta)\right] \tag{13}$$

Ces étapes sont répétées jusqu'à l'amélioration de la log-vraisemblance soit inférieure à un seuil donné.

La détermination du nombre de classes s'opère par comparaison des modèles avec un nombre de classes variable. Cette sélection est généralement basée sur le critère du minimum de Schwarz-BIC (Oliveira-Brochado et Martins, 2005).

$$\mathrm{BIC} = -2LL + k\ln(N) \tag{14}$$

avec k le nombre de paramètres à estimer, N le nombre d'individus et LL la valeur du maximum de la log-vraisemblance pour le modèle.

Dans une optique bayésienne, les π_c peuvent être vues comme des probabilités a priori d'appartenance à une classe. Elles sont agrégées. Les probabilités a posteriori sont les probabilités qu'a un individu i d'appartenir à une classe c.

$$P(c_i \mid y_i, x_{1i}, x_{2i}, \Theta) = \frac{\pi_c P\left(y_i \mid \lambda_{ci}(x_{1i})\right)}{\sum_{j=1}^{C} \pi_j P\left(y_i \mid \lambda_{ji}(x_{1i})\right)} \tag{15}$$

L'espérance du nombre d'achats pour un individu i donné s'écrit alors :

$$E(Y_i) = \sum_{c=1}^{C} \lambda_{i|c} P(c_i \mid y_i, x_{1i}, x_{2i}, \Theta) \tag{16}$$

Intérêts de la formulation

La souplesse de cette approche est considérable. Elle peut être purement stochastique sans prise en compte d'aucune variable explicative : seules les constantes sont intégrées dans les formulations. L'écriture devient alors $\lambda_{i|c} = \beta_{0c}$ et $\pi_c = \frac{\exp(k_c)}{\sum_{j=1}^{C} \exp(k_j)}$.

Naturellement il est tout à fait envisageable de n'intégrer des variables explicatives que dans une seule de ces expressions.

Cette formulation offre une compréhension intuitive de l'hétérogénéité à travers un nombre limité de segments. Mais cette facilité d'interprétation ne se fait pas au détriment de la rigueur et de la qualité. Laird (1978) ainsi qu'Heckman et Singer (1984) ont montré qu'un modèle mixte fini fournit une bonne approximation numérique de la distribution « réelle » même si la distribution mixte sous-jacente est continue. Chaque élément de la densité mixte constitue une approximation locale de cette distribution. Cette approche est ainsi une réelle alternative à une estimation purement non paramétrique.

APPLICATION

Les données

Notre investigation empirique est conduite à partir du marché test *BehaviorScan* à Angers. L'analyse s'étend sur un hypermarché de la zone du panel pour lequel nous disposons de 81 709 actes d'achat (produits alimentaires de grande consommation) effectués par un échantillon de 2 031 ménages sur une période de 140 semaines (1998-2001).

Nous travaillons sur les achats répétés i.e. les achats réalisés après l'observation de la première transaction. Nous avons distingué deux périodes.

- Une période d'estimation d'une durée de 8 mois (35 semaines). C'est sur cette période que sont estimées les valeurs des paramètres.
- Une période de validation servant à établir la validité prédictive des modèles (105 semaines).

Ce déséquilibre entre période d'estimation et période de validation s'explique par la nécessité de conserver des valeurs limitées au nombre de transactions pour l'estimation du modèle Pareto/NBD (Castéran, Meyer-Waarden et Bénavent, 2007b). En effet, compte tenu de sa formulation, un nombre trop important d'achats ne peut être pris en compte. Point plus intéressant, une longue période de validation permet d'établir la validité prédictive long terme du modèle.

Les variables explicatives suivantes sont disponibles : durée entre le premier achat et fin de la période d'estimation (en semaines) durée entre le dernier achat et la fin de la période d'estimation (en semaines), taille du foyer, revenu mensuel du foyer (en €), catégories socioprofessionnelles des membres du couple (variables indicatrices), âges (par tranche) des membres du couple (variables indicatrices), possession de la carte de fidélité du magasin, nombre total de cartes de fidélité possédées, distance et durée.

L'ensemble des estimations est réalisé sous le logiciel *R*. L'optimisation de la log-vraisemblance du modèle Pareto/NBD s'effectue par le biais de la commande *nlminb* permettant une optimisation sous contrainte. En ce qui concerne les modèles mixtes finis, le package *Flexmix* sous *R* (Grün et Leisch, 2007) permet une estimation directe par le biais de la commande *stepFlexmix*.

Les modèles mixtes finis

Sélection

Pour déterminer le nombre de classes à tester pour chaque modèle, nous avons observé la distribution des achats répétés sur la période d'estimation. Cette distribution compte un certain nombre de piques désignées par les flèches dans le graphique 1 pouvant caractériser autant de classes de consommation.

Les axes du graphique, pour une meilleure lisibilité, ont été tronqués : le maximum sur les abscisses devrait se situer à 144 (nombre d'achats répétés en 8 mois), sur l'axe des ordonnées (occurrences) à 321 (chiffre correspondant à l'absence d'achat répété sur la période d'estimation).

Insérer figure 1

Pour chaque type de modèles, nous avons testé une partition de l'échantillon entre une et huit classes. Plus de huit classes apparaissait à la fois peu vraisemblable compte tenu de la figure 1 et peu souhaitable (dans un souci de clarté des interprétations).

Quatre types de modèles ont été testés suivant la prise en compte de variables explicatives tout à la fois dans la formulation multinomiale et dans le paramètre λ.

Les principaux résultats sont résumés dans le tableau 1.

Insérer tableau 1

Assez logiquement, l'ensemble des modèles mixtes finis fait nettement mieux que le modèle de Poisson simple présenté comme simple benchmark. Même une approche basée sur de simples constantes se révèle plus efficace.

Le nombre de classes optimal varie suivant la spécification paramétrique retenue. Toutefois le nombre de classes le plus fréquemment (trois modèles sur quatre) retenu est de sept. Le BIC étant le critère de sélection des modèles, le meilleur modèle est le modèle combinant à la fois des variables explicatives dans le paramètre de fréquence et le paramètre de π_c et de λ. Le nombre de variables explicatives y est important. Le paramètre de fréquence intègre ainsi 25 variables explicatives. Cette valeur est liée au nombre des classes et à l'impossibilité de distinguer des formulations différentes par classe : si les coefficients sont estimés par classe, les variables explicatives restent identiques d'une classe à l'autre. Certaines variables sont donc significatives dans certaines classes et pas dans d'autres. La liste complète de ces variables et de leurs coefficients se situe en annexe 3. Les variables *carte de fidélité du magasin, ancienneté du premier achat observé* et *récence du dernier achat* ont été intégrées pour le calcul des probabilités a priori d'appartenance à un segment (modèle

multinomial logit). En l'absence de calcul explicite des coefficients pour le modèle logit, ces variables apparaissaient comme les plus pertinentes. Les différentes simulations réalisées avec d'autres variables ont établi, de manière très empirique, la pertinence de ce choix.

Toutefois, le modèle où les probabilités a priori sont définies par une constante se situe à un niveau à peu près équivalent en termes de BIC. La différence se situe en termes de log-vraisemblance avec un clair avantage pour les modèles explicatifs. Ce résultat atteste de l'efficacité limitée d'une formulation multinomiale explicite.

Interprétation

Insérer tableau 2

La probabilité d'affectation a priori par classe dépasse systématiquement les 5%. Lorsqu'on calcule le ratio entre le nombre d'individus affectés par classe et ceux dont la probabilité a posteriori dépasse 10^{-41}, la classe 1 apparaît comme spécialement discriminante. A l'inverse, la classe 6 est beaucoup moins segmentante avec 1 198 individus qui ont une probabilité supérieure à 10^{-4} pour seulement 59 réellement affectés.

La séparation des classes est établie à travers un rotogramme. Ce graphique présente la distribution des probabilités d'affectation (probabilités a posteriori) par classe. La différence par rapport à un histogramme classique tient au fait qu'est représentée la racine carrée des comptages plutôt que les comptages eux-mêmes. Cette transformation permet une meilleure représentation des petites valeurs.

Un pic à la valeur 1 (probabilité certaine) atteste d'une claire séparation des classes ; à l'inverse une répartition uniforme autour de valeurs moyennes est le signe de classes peu distinctes (Leisch, 2007).

Insérer figure 2

Les classes 1 et 4 sont clairement distinctes des autres. Par contre, les classes 3 et 7 apparaissent assez peu discriminantes.

---

[1] plus d'une chance sur 10 000 d'appartenir à la classe

Compte tenu du nombre de classes et de variables explicatives pour le paramètre de fréquence, le commentaire de chaque coefficient serait trop fastidieux et peu éclairant. Toutefois la valeur de ceux-ci ainsi que leur significativité figurent en annexe 3.

Certains résultats restent constants et cohérents d'une classe à l'autre. Ainsi conformément à l'intuition, la possession d'une carte de fidélité et la récence du dernier achat accroissent la probabilité d'achats fréquents. A l'opposé, un acheteur multi-fidèle (mesuré par le nombre de cartes de fidélité possédées) aura des achats moins fréquents. L'impact des autres variables varie suivant les classes.

La nature des classes apparaît clairement à travers l'étude de l'évolution des probabilités postérieures en fonction du nombre de transactions. Nous avons limité ce nombre à 10 dans le graphique suivant afin de faciliter la lecture.

Insérer figure 3

Les classes 1 et 4, déjà repérées pour leur fort pouvoir discriminant, sont également celles qui ont le comportement le plus marqué. La classe 1 rassemble pour l'essentiel les non acheteurs i.e. les clients n'ayant pas renouvelé d'achat. La classe 4, la plus importante, caractérise les acheteurs moyens avec une probabilité maximale pour un nombre d'achats répétés compris entre quatre et six. Les autres classes sont nettement moins représentées ; les classes 3 et 7 regroupent les acheteurs les plus fréquents.

Comparaison avec le modèle Pareto/NBD

L'estimation du modèle Pareto/NBD est extrêmement lourde : cinq heures sur un processeur Xéon double cœur dédié entièrement à son estimation contre quelques minutes pour les modèles mixtes finis. Cette longueur d'estimation est compliquée par l'extrême sensibilité du modèle Pareto/NBD aux valeurs de départ, ce qui implique des tâtonnements assez fastidieux pour éviter une non-convergence du processus d'estimation. Ces contraintes sont naturellement autant de limites à une implémentation dans un cadre professionnel.

Il n'y a pas d'interprétation immédiate des coefficients du modèle Pareto/NBD en l'absence de variables explicatives. Les valeurs des paramètres des lois gamma sont les suivantes.

Insérer tableau 3

L'importance du paramètre β est liée à la durée de vie importante des clients.

La comparaison de la validité prédictive des modèles doit s'opérer à deux niveaux : à un niveau purement statistique tout d'abord par la prise en compte d'indicateurs d'ajustement et à un niveau plus empirique en second lieu.

Les indicateurs statistiques

La log-vraisemblance et le critère BIC, déjà présenté plus haut, permettent de se faire de mesurer respectivement l'adéquation dans l'absolu du modèle avec les données et de « l'efficacité » relative du modèle.

Sur la base de ces critères, le modèle mixte fini est nettement plus efficace et informatif que le modèle Pareto/NBD.

Insérer tableau 4

Toutefois ces indicateurs ne permettent de se prononcer uniquement sur l'ajustement des modèles aux données de la période d'estimation. Pour autant, ils ne permettent pas de se prononcer complètement sur la validité prédictive du modèle.

Les mesures empiriques

Elles sont plus nombreuses et plus proches des préoccupations managériales : somme des achats prévus à un niveau agrégé, corrélation des achats prévus par individu avec les achats effectivement réalisés, moyenne des valeurs absolues des écarts prévision/réalisé par individu et enfin espérance conditionnelle des achats en fonction de l'historique consommateur.

Le premier critère est la capacité à reproduire de manière agrégée les achats.

Insérer figure 4

Sans réelle surprise, le modèle Pareto/NBD retranscrit mal la période d'estimation. En effet, comme signalé, il nécessite la prise en compte du début d'activité des clients c'est-à-dire, dans notre cas, le premier achat réalisé en magasin. Cette information n'étant pas disponible, les clients ont été considérés comme actifs dès leur premier achat observé. Ce décalage début d'activité /premier achat observé explique la médiocre reconstitution de la période estimée par le modèle Pareto/NBD. Pour autant, l'intérêt majeur réside dans la qualité de prévision.

A cet égard, le modèle Pareto/NBD gagne en efficience. Le modèle en classes latentes obtient cependant toujours de meilleurs résultats de la semaine 36 à la semaine 105 soit sur 70 semaines ! Toutefois par définition les prévisions du modèle à classes latentes à base de distribution Poisson (dont une des propriétés contraignantes est « l'absence de mémoire ») restent stables. L'inactivité progressive des clients est ici totalement ignorée. C'est donc sur la durée que le modèle Pareto/NBD, qui intègre l'inactivité client, prend sa revanche. Mais nous nous situons dans des intervalles de temps lointains (presque deux ans) qui réduisent l'intérêt de ce mieux.

Au niveau individuel et non plus agrégé, les principaux résultats sont résumés dans le tableau suivant. Ils sont basés sur les 35 premières semaines de prévision afin de ne pas alourdir la présentation et conserver un horizon de prévision « raisonnable ».

Insérer tableau 5

Les écarts sont faibles mais systématiquement à l'avantage du modèle Pareto/NBD. Ce résultat est a priori paradoxal dans la mesure où, au niveau agrégé, le modèle Pareto/NBD était moins efficace. L'examen plus attentif des données individuelles montre que cela tient à la nature de l'erreur ; pour le modèle Pareto/NBD, elle est très souvent liée à une sur-estimation du nombre d'achats. Pour le modèle à classes latentes, les erreurs peuvent être liées autant à une sur-estimation qu'à une sous-estimation. A un niveau agrégé, cette dernière approche donne donc de meilleurs résultats ; à un niveau désagrégé, où il ne peut y avoir d'effets de compensation, elle offre des prévisions légèrement moins pertinentes.

Ce constat est corroboré par l'examen de l'espérance conditionnelle. Le modèle Pareto/NBD surestime quasi systématiquement le réalisé et en tout état de cause, les prévisions du modèle mixte fini.

Insérer figure 5

Interprétation

Au-delà de ses capacités prédictives, le modèle à classes latentes offre une vision claire des différentes populations en présence. Si l'on considère les principales variables, les segments se présentent de la manière suivante.

Insérer tableau 6

Sans vouloir se livrer à une analyse exhaustive de tous les segments, les situations apparaissent clairement contrastées. Au segment 1 dont les membres, acheteurs récents et distants du magasin n'ont effectué quasiment achat répété s'oppose le segment 5 composé d'acheteurs très fréquents, largement porteurs de la carte du magasin et fidèles (0,79 carte de fidélité autre que celle du magasin). Mais entre ces deux extrêmes, la palette est large. Le segment 6 ne compte pas de détenteur de carte de fidélité mais affiche une fréquence de consommation presqu'aussi élevée que le segment 7. Ce segment rassemble à la fois le plus fort taux de porteurs de carte et apparaît comme le plus versatile (nombre total de cartes de fidélités).

Un effort de fidélisation devrait spécifiquement être conduit auprès du segment 4, comptant une minorité de porteurs de la carte de fidélité (3%) et faiblement consommateurs (3,6 achats répétés).

Naturellement cette analyse doit être approfondie en fonction des variables explicatives pour aboutir à la définition d'actions différenciées mais ce simple tableau fait apparaître l'étendue et la finesse des possibles.

CONCLUSIONS ET VOIES DE RECHERCHE

A un niveau purement managérial, les modèles mixtes finis attestent de leur efficacité. Face à une formulation plus moderne, considérablement plus exigeante en termes mathématiques et en implémentation, ils font la preuve de leur robustesse et de leur flexibilité. Outre leur atout opérationnel du fait de la présence de variables explicatives, ils offrent concrètement des prévisions d'une qualité au moins comparable avec une grande facilité d'implémentation. Ils ne nécessitent pas l'identification exacte de la date de début d'activité des clients et offrent, dès la période d'estimation, des résultats cohérents. Cette qualité, l'aisance de son implémentation sans recours à des fonctions lourdes et sensibles et son aspect opérationnel sont autant d'atouts. Il faut y rajouter l'existence d'une vraie lecture des comportements permettant une allocation optimisée des ressources marketing. Sans cela, on ne peut envisager de gestion de portefeuille client.

La limite de la formulation retenue, à base de distribution Poisson, est l'absence de prise en compte de l'inactivité client. Si dans les premières semaines, cet aspect n'est pas problématique, il tend à le devenir avec le temps qui passe. Le modèle Pareto/NBD qui prend en compte cet aspect sous forme probabiliste, obtient logiquement de meilleurs résultats. Toutefois il apparaît délicat à adapter à un cadre non contractuel (importance de la date de début d'activité). Toutefois, une opérationnalisation des modèles à classes latentes peut tout à fait passer par une ré-estimation régulière de ceux-ci (par exemple à une fréquence annuelle). Cette ré-estimation permet une prise en compte de l'inactivité et lève l'hypothèque de la stationnarité.

En définitive, l'intérêt de chaque approche est lié à l'objectif poursuivi. S'il s'agit d'obtenir des prévisions d'un volume d'achats à court ou moyen terme ou de disposer d'une vision claire des facteurs explicatifs d'un comportement afin de développer une approche marketing aboutie, l'avantage va clairement aux modèles à classes latentes. A l'inverse, si l'objet est la détermination d'une valeur actualisée client ou d'un capital client à une échelle agrégée, le modèle Pareto/NBD reste le seul réaliste.

Ceci conduit tout naturellement à dégager deux axes de recherche en matière de modèle mixte fini, même s'ils sont d'importance inégale. Le premier axe, le moins crucial, concerne la nécessaire sélection des variables classe par classe. Il est toujours préjudiciable de laisser

figurer, au sein des modèles, des variables non significatives. La qualité de l'estimation s'en ressent sans nul doute, il convient donc de lever cette restriction.

Mais au-delà de cet aspect, il convient de réfléchir à la philosophie du modèle. La prise en compte d'une simple formulation de Poisson ne permet pas de rendre compte de l'inactivité client. La principale voie de recherche réside ainsi dans l'estimation non d'un modèle de Poisson mais d'une formulation incluant un paramètre d'inactivité. C'est donc au prix d'une réflexion véritable sur le comportement client que pourra être dégagée une formalisation mixte fini. A cet égard, la « philosophie » du modèle Pareto/NBD peut être intéressante à la double condition qu'elle soit développée dans un cadre de classes latentes et qu'elle intègre un corpus de variables explicatives.

En tout état de cause, l'investissement de recherche dans les modèles à classes latentes se justifie pleinement au vu des résultats obtenus avec une formalisation basée uniquement sur une distribution de Poisson. Le caractère très restrictif de ce choix n'empêche en effet pas des prévisions pertinentes ; une formulation plus élaborée en termes marketing doit conduire à des progrès déterminants sur du long terme.

BIBLIOGRAPHIE

Ayache A. et alii (2006), Stochastic approach to customer Equity and Lifetime Value calculations with applications to customer retention models and some extensions, *35th Conference of EMAC,* Athènes 2006.

Baltagi, B. H. (Editor), 2003, *A Companion to Theoretical Econometrics*, Willinston, USA, Blackwell Publishing

Castéran H., Meyer-Waarden L. et Bénavent C. (2007a), Incorporation of covariates in the Pareto/NBD model: first formulations and comparison with others models in the retailing sector, *Third German French Austrian Conference,* Paris, ESSEC.

Castéran H., Meyer-Waarden L. et Bénavent C. (2007b), Une évaluation empirique des modèles NBD pour le calcul de la Valeur Actualisée Client dans le domaine de la grande distribution, *Congrès International de l'AFM,* Aix-les-Bains 2007.

Crié, D. (1999), Les produits fidélisants dans la relation client-fournisseur, Thèse pour le doctorat de Sciences de Gestion, IAE de Lille

Dempster A., Laird N. et Rubin D. (1977), Maximum Likelihood from Incomplete Data via the EM-algorithm, *Journal of the Royal Statistical Society, Series B ,* 56, 363-375.

Ehrenberg, A. S. C. (1959), The Pattern of Consumer Purchases, *Applied Statistics*, 8(1), p. 26-41.

Ehrenberg A.S.C. (1988), *Repeat Buying, Facts, Theory and Applications*. London, C. Griffin and Co. Ltd, Oxford University Press New York.

Fader P., Hardie B. et Lok Lee K. (2005), « Counting your Customers » the Easy Way: An Alternative to the Pareto/NBD Model, *Marketing Science*, 24, 2, p. 275-284.

Grün B. et Leisch F. (2007), Applications of finite mixtures of regression models, *white paper*

Hausman J. et McFadden D. (1984), Specification Tests for the Multinomial Logit Model, *Econometrica*, 52, 5, p. 1219-1240

Heckman J. et Singer B. (1984), A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data., *Econometrica,* Mar1984, Vol. 52 Issue 2, p271-320.

Ho T.-H., Park Y.-H. et Zhou Y.-P. (2006), Incorporating Satisfaction into Customer Value Analysis: Optimal Investment in Life-Time Value, *Marketing Science*, 25, 3, p. 260-277

Jackson B.B. (1985), Build customer relationship that last, *Harvard Business Review*, 63, p.120-128.

Laird N. (1978), Nonparametric Maximum Likelihood Estimation of a Mixing Distribution., *Journal of the American Statistical Association,* Dec78, Vol. 73 Issue 364, p805.

Leisch F. (2007), FlexMix : A General Framework for Finite Mixture Models and Latent Class Regression in R, *white paper*

Newcomb S. (1886), A Generalized Theory of the Combination of Observations so as to Obtain the Best Result, *American Journal of Mathematics,* 8: 343-366.

Oliveira-Brochado A. et Martins F. (2005), Assessing the Number of Components in Mixture Models: a Review, *Working Papers (FEP) -- Universidade do Porto,* Nov2005 Issue 86, p1-31.

Pearson K. (1894), Contributions to the Mathematical Theory of Evolution, *Philosophical Transactions,* A185: 71-110.

R Development Core Team (2007), R: A Language and Environment for Statistical Computing v. 2.5.1, *R Foundation for Statistical Computing,* Vienne, Autriche.

Schmittlein D., Morrison D. et Colombo R. (1987), Counting your Customers: Who are They and What will They do Next? *Management Science,* 33, 1, p. 1-24

Schmittlein D. et Peterson R.A. (1994), Customer Base Analysis: An Industrial Purchase Process Application, *Marketing Science*, 13, 1, p. 41-67

Villanueva J. et Hansens D. (2007), Customer equity: Measurement, Management and Research Opportunities, *Foundations and Trends in Marketing,* Vol. 1 N° 1 p. 1-95.

Wedel M. et alii (1993), A Latent Class Poisson Regression Model For Heterogeneous Count Data, *Journal of Applied Econometrics,* Oct-Dec93, Vol. 8 Issue 4, p397-411.

Tableau 1 :

Indicateurs d'ajustement en fonction de la présence de variables explicatives dans les expressions

		Probabilités a priori		
		sans variables explicatives		**avec variables explicatives**
λ constant par classe (pas de variables explicatives)	Nombre de classes	7		7
	Log-vraisemblance	-6 691		-5 540
	BIC	13 511		12 062
λ variable par classe (avec variables explicatives)	Nombre de classes	5		7
	Log-vraisemblance	-5 238		-4 736
	BIC	11 078		11 040
Modèle de Poisson standard (avec variables explicatives)	Nombre de classes			1
	Log-vraisemblance			-7 589
	BIC			15 338

Tableau 2 :

Probabilités a priori et taille des classes

	prior	size	post>0	ratio
Comp.1	0,1788	363	364	0,9973
Comp.2	0,1009	220	1008	0,2183
Comp.3	0,0899	150	1377	0,1089
Comp.4	0,4023	905	1676	0,5400
Comp.5	0,0632	121	962	0,1258
Comp.6	0,0576	59	1198	0,0492
Comp.7	0,1074	213	1276	0,1669

'logLik.' -4735.863 (df=206)

AIC: 9883.726 BIC: 11040.68

Tableau 3 :

Coefficients du modèle Pareto/NBD

r	0,712
α	2,1269
β	194,584
s	1

Tableau 4 :

Comparaison des indicateurs statistiques

	Modèle Pareto/NBD	Modèle mixte fini (7 classes)
Log-vraisemblance	-36 873	-4 736
BIC	73 778	11 041

Tableau 5 :

Principaux résultats empiriques sur les 35 premières semaines de la période de validation

	Modèle Pareto/NBD	**Modèle mixte fini (7 classes)**
Corrélation réalisé/ prévu	0,911	0,902
Moyenne de la valeur absolue des écarts prévu / réalisé	3,97	4,1

Tableau 6 :

Principales caractéristiques en fonction de la classe d'affectation

Classe d'affectation	Nombre d'achats répétés	Nombre total de cartes de fidélité possédées	Distance moyenne au magasin (km)	Taux de chef famille cadre	Taux de chef de famille inactif	Taux de possession de la carte du magasin	Nombre de semaines entre 1er et dernier achat observés	Revenu moyen estimé
1	0,48	1,05	3,12	14%	28%	17%	17,18	1 876,86 €
2	24,54	0,91	1,97	19%	29%	10%	33,30	2 063,18 €
3	14,95	1,02	2,32	18%	29%	17%	33,31	2 030,00 €
4	3,62	0,93	2,87	14%	26%	3%	25,29	1 841,88 €
5	31,61	1,09	1,94	17%	34%	30%	33,68	1 920,66 €
6	20,95	0,71	2,81	24%	24%	0%	28,89	2 249,15 €
7	22,68	1,24	2,19	22%	30%	38%	31,08	2 136,62 €
Total	10,33	0,99	2,65	16%	28%	13%	26,51	1 933,43 €

Figure 1 :

Distribution des achats répétés sur la période d'estimation

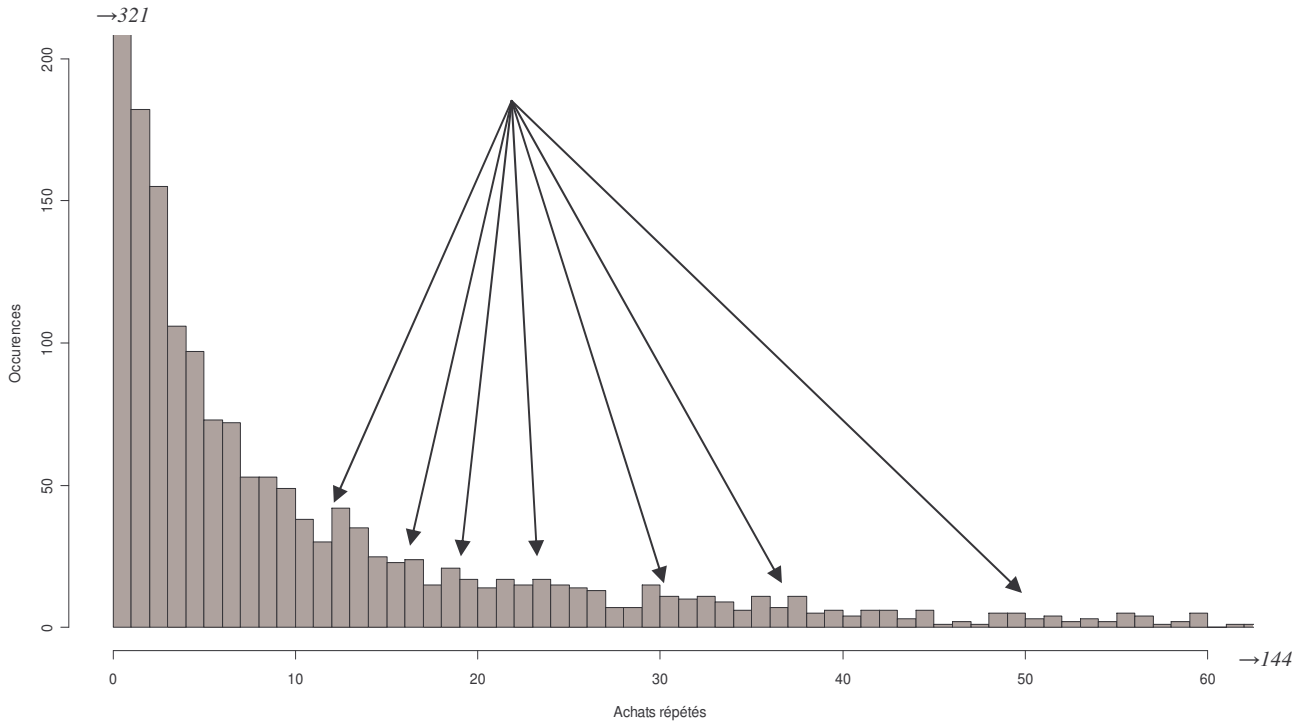


Figure 2 :

Rotogramme des classes

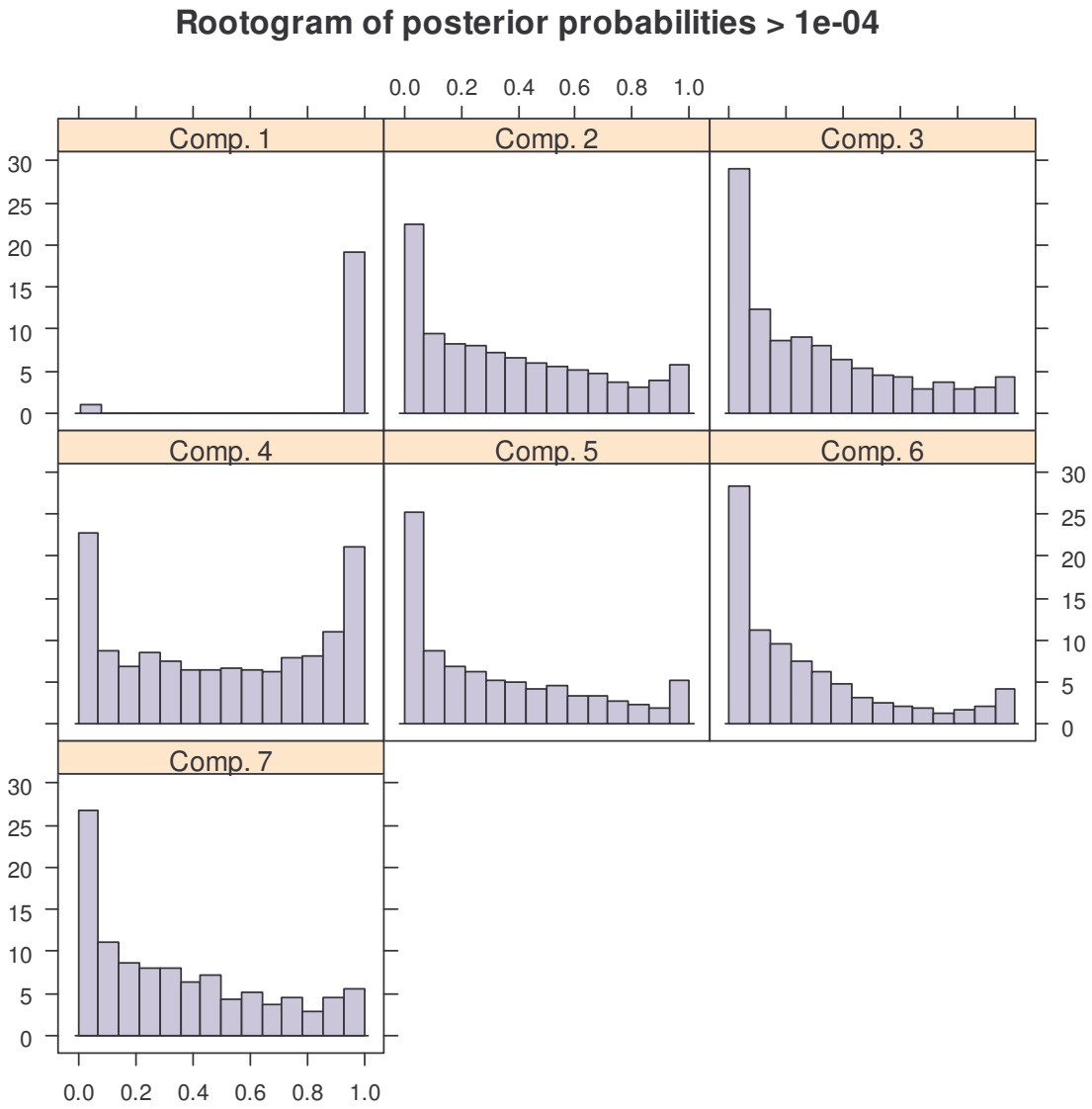


Figure 3 :

Evolution des probabilités postérieures (affectation par classe) en fonction du nombre d'achats répétés

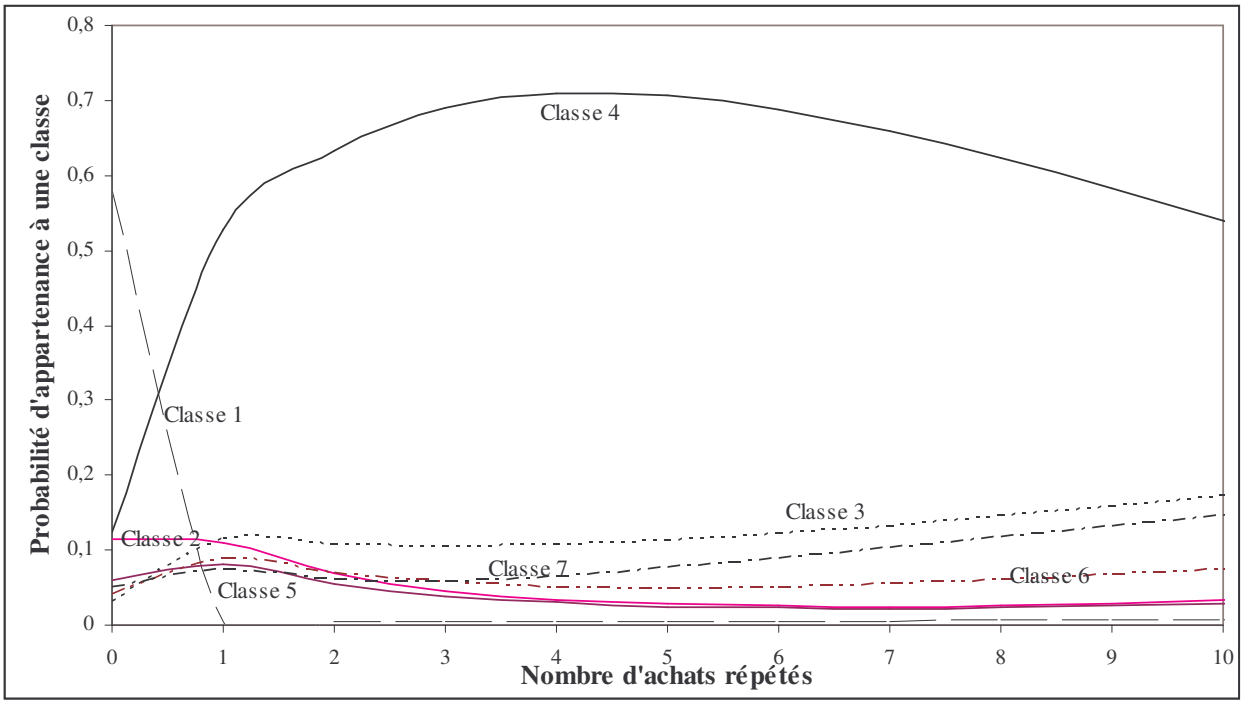


Figure 4 :

Comparatif valeurs prédites / valeurs réelles sur la période de validation

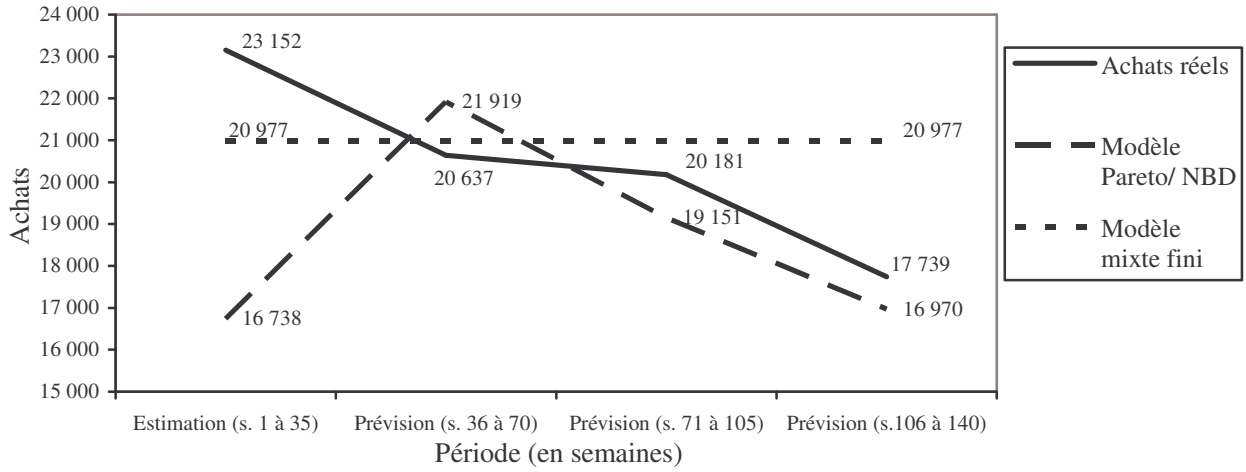
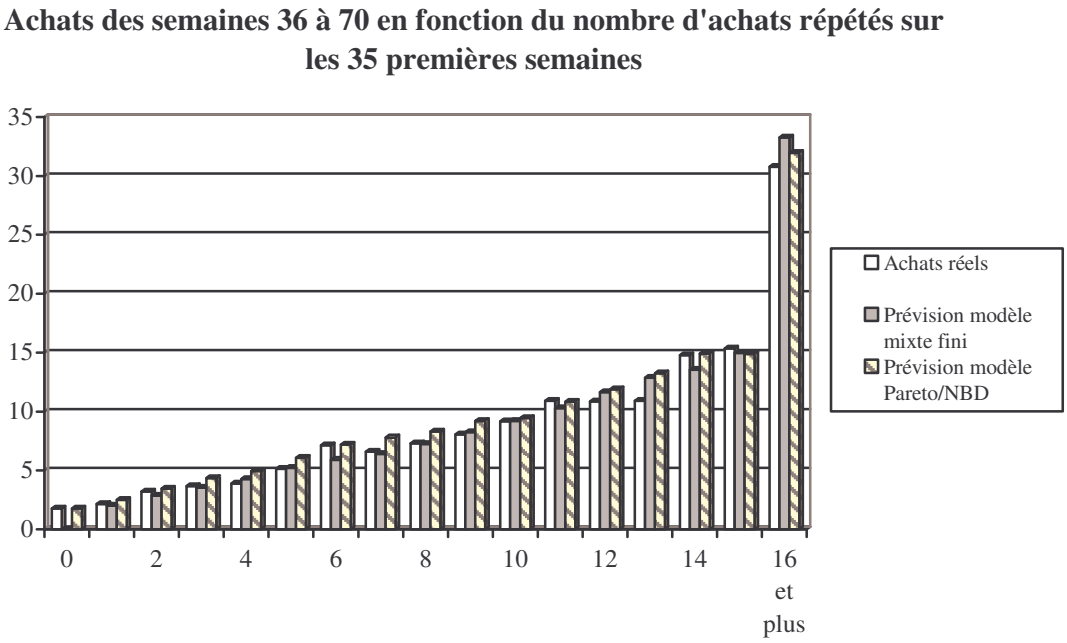


Figure 5 :

Comparatif achats réalisés/ achats prévus sur les 35 premières semaines de validation en fonction des achats de la période d'estimation



Annexe 1

Log-vraisemblance du modèle Pareto/NBD

$$LL(r,\alpha,s,\beta) = \ln\left[\frac{\Gamma(r+y)\alpha^r\beta^s}{\Gamma(r)}\right] + \ln\left[\frac{1}{(\alpha+T)^{r+y}(\beta+T)^s} + \left(\frac{s}{r+s+y}\right)B_1\right]$$

avec B_1 égale :

- si $\alpha \geq \beta$

$$B_1 = \frac{{}_2F_1\left(r+s+y, s+1, r+s+y+1, \frac{\alpha-\beta}{\alpha+t_y}\right)}{(\alpha+t_y)^{r+s+y}} - \frac{{}_2F_1\left(r+s+y, s+1, r+s+y+1, \frac{\alpha-\beta}{\alpha+T}\right)}{(\alpha+T)^{r+s+y}}$$

- si $\alpha \leq \beta$

$$B_1 = \frac{{}_2F_1\left(r+s+y, r+y, r+s+y+1, \frac{\beta-\alpha}{\beta+t_y}\right)}{(\beta+t_y)^{r+s+y}} - \frac{{}_2F_1\left(r+s+y, r+y, r+s+y+1, \frac{\beta-\alpha}{\beta+T}\right)}{(\beta+T)^{r+s+y}}$$

Annexe 2

Probabilité d'activité à une date t pour un consommateur donné (modèle Pareto/NBD)

$$P(\tau > T \,|\, r,\alpha,s,\beta,Y=y,t,T) = \left[1+\left(\frac{s}{r+s+y}\right)(\alpha+T)^{r+y}\,(\beta+T)^{s}\,A_1\right]^{-1}$$

avec

$$A_1 = \begin{cases} \dfrac{{}_2F_1\left(r+s+y, s+1; r+s+y+1; \dfrac{\alpha-\beta}{\alpha+t_y}\right)}{(\alpha+t_y)^{r+s+y}} - \dfrac{{}_2F_1\left(r+s+y, s+1; r+s+y+1; \dfrac{\alpha-\beta}{\alpha+T}\right)}{(\alpha+T)^{r+s+x}} & si\ \alpha \geq \beta \\ \dfrac{{}_2F_1\left(r+s+y, r+y; r+s+y+1; \dfrac{\beta-\alpha}{\beta+t_y}\right)}{(\beta+t_y)^{r+s}} - \dfrac{{}_2F_1\left(r+s+y, r+y; r+s+y+1; \dfrac{\beta-\alpha}{\beta+T}\right)}{(\beta+T)^{r+s}} & si\ \alpha \leq \beta \end{cases}$$

Annexe 3

Signification des noms de variables

T : ancienneté du premier achat (en semaines)
T_X : récence du dernier achat (en semaines)
CARTE_GE : carte de fidélité du magasin (indicatrice)
HH_SIZE : taille du foyer
TOTCARTE : nombre de cartes de fidélité possédées
DIS_M1 : distance au magasin
DUREE1 : durée de trajet
CADRE_C : profession cadre
INTERM_C : profession intermédiaire
EMPLOY_C : profession employé
OUVRIER_C : profession ouvrier
INACT_C : profession inactif
INTERM_P : profession du conjoint (profession intermédiaire)
EMPLOY_P : profession du conjoint (employé)
INACT_P : profession du conjoint (inactif)
R_M1 : revenu de moins de 1 000 € mensuels (
R_1.2 : revenu entre 1 000 et 1 200 € mensuels
R_1.8 : revenu entre 1 200 et 1 800 € mensuels
R_2.3 : revenu entre 1 800 et 2 300 € mensuels
R_2.7 : revenu entre 2 300 et 2 700 € mensuels
R_3.3 : revenu entre 2 700 et 3 300 € mensuels
A30_39_C : âge entre 30 et 39 ans.
A40_49_C : âge entre 40 et 49 ans
A50P_C : plus de 50 ans
A50P_P : conjoint de plus de 50 ans

Annexe 4 :

Coefficients estimés par classe

```
Number of components: 7

$Comp.1
             Estimate    Std. Error  z value  Pr(>|z|)
(Intercept) -2.4094e+01  3.5222e+03 -0.0068   0.99454
T           -2.9453e-02  2.2431e-02 -1.3131   0.18916
T_X          2.3296e-01  3.4074e-02  6.8368   8.097e-12 ***
CARTE_GE     2.1164e+01  3.5222e+03  0.0060   0.99521
HH_SIZE     -2.1075e-02  1.6006e-01 -0.1317   0.89524
TOTCARTE     1.6668e-01  1.8066e-01  0.9226   0.35619
DIS_M1      -8.2866e-02  1.2967e-01 -0.6390   0.52279
DUREE1       3.3866e-04  9.9718e-04  0.3396   0.73415
CADRE_C     -9.4056e-01  1.4308e+00 -0.6574   0.51095
INTERM_C     2.3039e-01  9.8580e-01  0.2337   0.81521
EMPLOY_C    -9.2645e-01  8.1290e-01 -1.1397   0.25442
OUVRIER_C   -6.5093e-01  8.3303e-01 -0.7814   0.43457
INACT_C      1.0586e-01  1.3392e+00  0.0790   0.93700
INTERM_P    -1.7789e-01  8.3837e-01 -0.2122   0.83196
EMPLOY_P     1.1246e+00  4.7068e-01  2.3894   0.01688 *
INACT_P     -1.3860e-01  1.0272e+00 -0.1349   0.89266
R_M1         1.3826e+00  1.3587e+00  1.0176   0.30889
R_1.2       -9.7012e-01  1.3727e+00 -0.7067   0.47972
R_1.8        6.5322e-01  1.3254e+00  0.4928   0.62213
R_2.3        2.0527e-01  1.3089e+00  0.1568   0.87538
R_2.7        1.3320e-01  1.4131e+00  0.0943   0.92490
R_3.3        1.0736e+00  1.2466e+00  0.8612   0.38914
A30_39_C     9.8876e-01  5.9327e-01  1.6666   0.09559 .
A40_49_C     2.9603e-01  5.3339e-01  0.5550   0.57890
A50P_C     8.2982e-01  7.8180e-01  1.0614   0.28850
A50P_P     4.2895e-01  4.9096e-01  0.8737   0.38228
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$Comp.2
             Estimate     Std. Error  z value  Pr(>|z|)
(Intercept) -4.0344e+00  4.0364e-01  -9.9950 < 2.2e-16 ***
T           -8.3677e-02  2.1448e-02  -3.9014 9.562e-05 ***
T_X          3.2486e-01  1.7425e-02  18.6437 < 2.2e-16 ***
CARTE_GE     7.9516e-02  4.2758e-02   1.8597 0.0629313 .
HH_SIZE     -1.0175e-01  1.4906e-02  -6.8265 8.698e-12 ***
TOTCARTE     2.5355e-02  1.8682e-02   1.3572 0.1747118
DIS_M1      -3.4105e-02  9.8965e-03  -3.4462 0.0005686 ***
DUREE1       2.1378e-04  9.4253e-05   2.2682 0.0233183 *
CADRE_C     -2.0604e-01  9.9773e-02  -2.0650 0.0389193 *
INTERM_C     5.8008e-01  1.1807e-01   4.9129 8.973e-07 ***
EMPLOY_C     8.1152e-01  1.1230e-01   7.2264 4.961e-13 ***
OUVRIER_C    3.6770e-01  1.0658e-01   3.4500 0.0005607 ***
INACT_C      5.8282e-01  1.2981e-01   4.4900 7.124e-06 ***
INTERM_P    -4.8296e-03  8.5497e-02  -0.0565 0.9549531
EMPLOY_P    -4.5766e-01  6.3456e-02  -7.2123 5.502e-13 ***
INACT_P     -3.6771e-01  7.9834e-02  -4.6059 4.107e-06 ***
R_M1        -1.5383e+00  8.2523e-02 -18.6413 < 2.2e-16 ***
R_1.2       -1.3161e+00  6.5933e-02 -19.9609 < 2.2e-16 ***
R_1.8       -8.2895e-01  5.2425e-02 -15.8123 < 2.2e-16 ***
R_2.3       -6.0361e-01  5.4219e-02 -11.1328 < 2.2e-16 ***
R_2.7       -6.3370e-01  5.9209e-02 -10.7029 < 2.2e-16 ***
R_3.3       -5.5680e-01  5.5505e-02 -10.0316 < 2.2e-16 ***
A30_39_C     1.6889e-01  6.1935e-02   2.7269 0.0063937 **
A40_49_C     2.2218e-01  5.8643e-02   3.7886 0.0001515 ***
A50P_C       3.5517e-02  6.5535e-02   0.5420 0.5878477
A50P_P      -1.1445e-01  5.1564e-02  -2.2195 0.0264498 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
$Comp.3
             Estimate    Std. Error  z value Pr(>|z|)
(Intercept) -0.32625013  0.37700391  -0.8654 0.3868325
T            0.03676467  0.01285691   2.8595 0.0042427 **
T_X          0.07865284  0.00720916  10.9101 < 2.2e-16 ***
CARTE_GE     1.10611134  0.05565275  19.8752 < 2.2e-16 ***
HH_SIZE     -0.03115199  0.01936298  -1.6088 0.1076506
TOTCARTE    -0.34033757  0.02832562 -12.0152 < 2.2e-16 ***
DIS_M1       0.00461359  0.01040899   0.4432 0.6575981
DUREE1       0.00033865  0.00012177   2.7810 0.0054190 **
CADRE_C     -0.34863380  0.11351553  -3.0712 0.0021317 **
INTERM_C    -0.59300833  0.13176478  -4.5005 6.779e-06 ***
EMPLOY_C    -0.30294585  0.11890796  -2.5477 0.0108425 *
OUVRIER_C   -0.90343129  0.12313167  -7.3371 2.182e-13 ***
INACT_C     -1.46941569  0.14433551 -10.1806 < 2.2e-16 ***
INTERM_P     0.40896461  0.09221615   4.4348 9.214e-06 ***
EMPLOY_P     0.62421101  0.07309383   8.5399 < 2.2e-16 ***
INACT_P      0.47163738  0.08148125   5.7883 7.111e-09 ***
R_M1         0.34488394  0.10154010   3.3965 0.0006825 ***
R_1.2       -0.50836423  0.09743207  -5.2176 1.812e-07 ***
R_1.8       -0.26142626  0.07618870  -3.4313 0.0006007 ***
R_2.3       -0.56581383  0.07888221  -7.1729 7.343e-13 ***
R_2.7       -0.07212251  0.07884482  -0.9147 0.3603281
R_3.3       -0.66203167  0.08364388  -7.9149 2.475e-15 ***
A30_39_C    -0.24598843  0.07801851  -3.1529 0.0016163 **
A40_49_C    -0.43139801  0.07943676  -5.4307 5.613e-08 ***
A50P_C       0.44916215  0.09839675   4.5648 5.000e-06 ***
A50P_P      -0.42657942  0.08799046  -4.8480 1.247e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$Comp.4
             Estimate    Std. Error z value Pr(>|z|)
(Intercept) -1.0226e-01  1.8905e-01 -0.5409 0.5885533
T           -9.9002e-03  4.0350e-03 -2.4536 0.0141442 *
T_X          6.2980e-02  3.3372e-03 18.8720 < 2.2e-16 ***
CARTE_GE     8.2851e-02  8.0466e-02  1.0296 0.3031816
HH_SIZE      8.8552e-02  1.7013e-02  5.2051 1.939e-07 ***
TOTCARTE     2.3713e-02  2.2498e-02  1.0540 0.2918845
DIS_M1      -6.6046e-02  1.1094e-02 -5.9534 2.627e-09 ***
DUREE1       2.9713e-04  8.2733e-05  3.5914 0.0003289 ***
CADRE_C      1.2510e-01  1.3008e-01  0.9617 0.3362043
INTERM_C     1.1257e-01  1.3215e-01  0.8518 0.3943183
EMPLOY_C     9.2313e-02  1.3030e-01  0.7085 0.4786525
OUVRIER_C    1.3345e-01  1.2908e-01  1.0338 0.3012224
INACT_C      1.8560e-01  1.5637e-01  1.1869 0.2352582
INTERM_P    -2.6360e-02  8.8923e-02 -0.2964 0.7668993
EMPLOY_P    -4.8355e-02  7.0405e-02 -0.6868 0.4922020
INACT_P     -1.5274e-01  9.4771e-02 -1.6117 0.1070230
R_M1         3.5687e-01  9.8703e-02  3.6156 0.0002997 ***
R_1.2        4.5901e-01  9.4312e-02  4.8669 1.133e-06 ***
R_1.8        3.3655e-01  8.9759e-02  3.7495 0.0001772 ***
R_2.3        2.7521e-01  8.9244e-02  3.0838 0.0020437 **
R_2.7        1.4453e-01  9.6899e-02  1.4916 0.1358156
R_3.3        2.4077e-01  9.7084e-02  2.4800 0.0131397 *
A30_39_C    -2.5133e-01  7.0278e-02 -3.5762 0.0003486 ***
A40_49_C    -4.5067e-01  7.4064e-02 -6.0849 1.165e-09 ***
A50P_C      -6.5329e-02  9.8049e-02 -0.6663 0.5052257
A50P_P      -2.7661e-01  8.9033e-02 -3.1069 0.0018909 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
$Comp.5
             Estimate    Std. Error   z value Pr(>|z|)
(Intercept)  0.48513390  0.42936469   1.1299 0.2585235
T           -0.16037769  0.02801484  -5.7247 1.036e-08 ***
T_X          0.27889848  0.02428365  11.4850 < 2.2e-16 ***
CARTE_GE     0.16323376  0.04818182   3.3879 0.0007044 ***
HH_SIZE      0.11794227  0.01909999   6.1750 6.617e-10 ***
TOTCARTE    -0.15685408  0.02377455  -6.5976 4.180e-11 ***
DIS_M1      -0.31841354  0.01964657 -16.2071 < 2.2e-16 ***
DUREE1      -0.00025888  0.00011883  -2.1785 0.0293667 *
CADRE_C     -0.25749717  0.11943906  -2.1559 0.0310924 *
INTERM_C    -0.88156711  0.11858752  -7.4339 1.054e-13 ***
EMPLOY_C    -0.62836868  0.10115327  -6.2120 5.230e-10 ***
OUVRIER_C   -1.49517520  0.12020432 -12.4386 < 2.2e-16 ***
INACT_C     -1.12512617  0.12015809  -9.3637 < 2.2e-16 ***
INTERM_P    -0.38364248  0.12052674  -3.1830 0.0014573 **
EMPLOY_P    -0.14043828  0.08640759  -1.6253 0.1040986
INACT_P      0.40136189  0.08597383   4.6684 3.035e-06 ***
R_M1         1.20819882  0.11003489  10.9801 < 2.2e-16 ***
R_1.2        0.99351253  0.10077371   9.8588 < 2.2e-16 ***
R_1.8        0.65737644  0.09165617   7.1722 7.380e-13 ***
R_2.3        1.06065411  0.09196818  11.5328 < 2.2e-16 ***
R_2.7        0.90330542  0.10116820   8.9287 < 2.2e-16 ***
R_3.3        1.43472508  0.09632404  14.8948 < 2.2e-16 ***
A30_39_C    -0.97469589  0.08134579 -11.9821 < 2.2e-16 ***
A40_49_C    -0.70922197  0.07833002  -9.0543 < 2.2e-16 ***
A50P_C      -0.58751077  0.07467904  -7.8671 3.628e-15 ***
A50P_P      -0.27447090  0.05611015  -4.8916 1.000e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$Comp.6
             Estimate    Std. Error  z value  Pr(>|z|)
(Intercept)  0.60741499  0.34858438   1.7425 0.0814176 .
T           -0.14324078  0.01695547 -8.4481 < 2.2e-16 ***
T_X          0.19106390  0.01603950 11.9121 < 2.2e-16 ***
CARTE_GE     0.43956103  0.21365526  2.0573 0.0396538 *
HH_SIZE     -0.24346449  0.02763120 -8.8112 < 2.2e-16 ***
TOTCARTE    -0.33689108  0.03638938 -9.2580 < 2.2e-16 ***
DIS_M1       0.04266559  0.01253259  3.4044 0.0006632 ***
DUREE1       0.00248350  0.00019626 12.6543 < 2.2e-16 ***
CADRE_C     -1.22295107  0.23186024 -5.2745 1.331e-07 ***
INTERM_C    -2.10616583  0.24709147 -8.5238 < 2.2e-16 ***
EMPLOY_C    -0.74505757  0.23261979 -3.2029 0.0013605 **
OUVRIER_C   -0.55892741  0.21882976 -2.5542 0.0106443 *
INACT_C     -2.76904162  0.28446310 -9.7343 < 2.2e-16 ***
INTERM_P     0.82718172  0.16571237  4.9917 5.986e-07 ***
EMPLOY_P     0.61341408  0.09327152  6.5766 4.812e-11 ***
INACT_P      2.07349334  0.11873119 17.4638 < 2.2e-16 ***
R_M1        -0.42479279  0.11430033 -3.7165 0.0002020 ***
R_1.2        0.07953681  0.09826955  0.8094 0.4183001
R_1.8        0.35898868  0.08921354  4.0239 5.724e-05 ***
R_2.3       -0.36298553  0.10181885 -3.5650 0.0003638 ***
R_2.7       -0.96520837  0.13919444 -6.9342 4.084e-12 ***
R_3.3       -0.46706482  0.11672106 -4.0015 6.293e-05 ***
A30_39_C     0.04332091  0.10293778  0.4208 0.6738679
A40_49_C     0.37618540  0.09499688  3.9600 7.496e-05 ***
A50P_C      -0.93032376  0.13610397 -6.8354 8.178e-12 ***
A50P_P       0.03489754  0.11102979  0.3143 0.7532872
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
$Comp.7
             Estimate    Std. Error  z value  Pr(>|z|)
(Intercept)  1.33407534  0.28696123   4.6490 3.336e-06 ***
T           -0.27735164  0.01403219 -19.7654 < 2.2e-16 ***
T_X          0.29572098  0.01365062  21.6636 < 2.2e-16 ***
CARTE_GE     0.43041992  0.04041727  10.6494 < 2.2e-16 ***
HH_SIZE      0.12271802  0.01545209   7.9418 1.992e-15 ***
TOTCARTE    -0.09786017  0.01893160  -5.1691 2.352e-07 ***
DIS_M1      -0.16048118  0.01204675 -13.3215 < 2.2e-16 ***
DUREE1      -0.00010523  0.00010189  -1.0327  0.301732
CADRE_C      1.50464472  0.23477117   6.4090 1.465e-10 ***
INTERM_C     1.26892134  0.23440830   5.4133 6.188e-08 ***
EMPLOY_C     0.70672703  0.23380721   3.0227  0.002505 **
OUVRIER_C    1.10140086  0.23246599   4.7379 2.159e-06 ***
INACT_C      1.40748059  0.24597505   5.7220 1.052e-08 ***
INTERM_P     0.69273080  0.07793350   8.8887 < 2.2e-16 ***
EMPLOY_P     0.34345776  0.06411671   5.3568 8.473e-08 ***
INACT_P      0.08054574  0.07851502   1.0259  0.304956
R_M1         0.18216710  0.08200499   2.2214  0.026323 *
R_1.2        0.32400696  0.07448850   4.3498 1.363e-05 ***
R_1.8        0.74825768  0.06407218  11.6784 < 2.2e-16 ***
R_2.3        0.26183443  0.06441329   4.0649 4.805e-05 ***
R_2.7        0.13426935  0.06825202   1.9673  0.049153 *
R_3.3        0.09206737  0.06928891   1.3287  0.183932
A30_39_C    -0.54373517  0.06719358  -8.0921 5.866e-16 ***
A40_49_C    -0.30855313  0.06433494  -4.7960 1.618e-06 ***
A50P_C      -0.06541331  0.08035177  -0.8141  0.415595
A50P_P      -0.38684784  0.06462520  -5.9860 2.150e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```