

**Etude du choix d'acquisition d'automobiles : Application
du modèle probit multinomial bayésien à l'enquête “
Parc Auto ” (1997-2001)**

Roger Collet

► **To cite this version:**

Roger Collet. Etude du choix d'acquisition d'automobiles : Application du modèle probit multinomial bayésien à l'enquête “ Parc Auto ” (1997-2001). 2008. hal-00318691

HAL Id: hal-00318691

<https://hal.archives-ouvertes.fr/hal-00318691>

Submitted on 4 Sep 2008

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

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

Etude du choix d'acquisition d'automobiles :
Application du modèle probit multinomial bayésien
à l'enquête « Parc Auto » (1997-2001)

Roger Collet*

Institut National de Recherche sur les Transports et leur Sécurité (INRETS)

Département d'Economie et de Sociologie des Transports (DEST)

Doctorant à l'Université de PARIS 1 - Panthéon Sorbonne

Première version

(ne pas citer)

Résumé :

Dans ce papier, nous effectuons une analyse comportementale du choix individuel des agents lors d'une nouvelle acquisition automobile. En croisant les trois variables binaires du niveau de gamme (haut ou bas), du carburant (gasoil ou supers), et de l'état à l'achat (neuf ou occasion), huit types de voitures sont considérées comme alternatives concurrentes de choix. Pour caractériser le comportement des acquéreurs, nous ajustons la version bayésienne du Probit Multinomial (MNP) aux données françaises de « Parc Auto » (SOFRES). La méthode d'estimation utilisée repose sur la version hybride de l'échantillonnage de Gibbs proposée par Nobile (1998). Exploitant les résultats avec le simulateur de fréquence brute (CFS), nous effectuons une analyse *ceteris paribus* afin de déterminer l'impact d'un changement de caractéristiques de l'agent et l'impact de la hausse du prix des carburants sur la répartition des probabilités de choix (ou des parts des marchés) entre les différentes alternatives considérées.

Mots-clé :

Modèle Probit Multinomial, Analyse bayésienne, Echantillonnage de Gibbs, Motorisation, Automobile.

Classification JEL : C11 ; C25

* Correspondance à :

Roger Collet, INRETS/DEST, 2, avenue du général Malleret-Joinville, 94114 ARCUEIL Cedex, FRANCE.

E-mail : rcollet@inrets.fr

Introduction

Les préoccupations environnementales croissantes liées au réchauffement climatique mettent l'homme au banc des accusés. La croissance de ses activités s'est accompagnée d'un rejet croissant dans l'atmosphère des gaz à effet de serre qui sont, sous l'action des rayons UV solaires, responsables de la hausse du thermomètre. Les conséquences écologiques à terme prévoient un bouleversement irréversible de l'écosystème, dans lequel la place de l'homme est incertaine et controversée. Soucieuse de sa préservation, la ratification du protocole de Kyoto (1997) par certains pays comme la France marque une prise de conscience et une volonté communautaire de réduire l'émission de ces gaz, avec des objectifs ambitieux, pourtant jugés nécessaires.

Plus particulièrement, c'est l'activité de transport qui est visée comme étant une principale source génératrice de ces gaz, faisant notamment de l'automobile « l'instrument du délit ». La voiture est aussi impliquée dans bien d'autres affaires, plus locales cette fois, telles que la pollution atmosphérique, la sécurité, la dégradation du paysage, le bruit... mais la voiture permet aux agents une plus large maîtrise de l'espace et du temps, et donc un meilleur accès à l'emploi, aux services, aux loisirs, faisant de l'automobile un sujet de débat très controversé.

Au cœur de toutes ces préoccupations particulières se pose clairement la question du parc automobile. Les voitures récentes, moins bruyantes, moins gourmandes en carburants, moins émettrices en particules, mais aussi plus efficaces, plus sûres et plus écologiques que les anciens modèles, marquent un réel effort visant à limiter les effets

négatifs de l'automobile sur le confort de vie des automobilistes, mais aussi de la collectivité en générale. Pour faciliter le renouvellement du parc automobile en faveur de ces véhicules, ou orienter des politiques publiques visant à réduire les nuisances, il est nécessaire de connaître les fondements relatifs au choix d'acquisition des agents. Sommairement, ce document s'attache à répondre à la question suivante: « qui achète quoi ». Typiquement, le modèle économétrique approprié pour décrire ce comportement d'acquisition doit appartenir à la famille des modèles de choix discrets. Dans ce document, nous visitons plus précisément le modèle probit multinomial (MNP). Sa description, puis sa version bayésienne sont tout d'abord abordées dans une première section. La partie suivante traite de la résolution du modèle par des méthodes « MCMC » (Monte Carlo Markov Chain). Sur la base des données françaises issue de « Parc Auto » (section 3), nous avons choisi de regrouper les possibilités d'acquisition en 8 modalités, croisant les caractéristiques automobiles d'état à l'achat, de carburant utilisé, et de gamme. Enfin, les résultats de l'ajustement du MNP sont enfin décrits et exploités dans une dernière partie.

1) Modèle Probit Multinomial (MNP)

1.1) Présentation du modèle

Considérons la représentation compacte de la décision d'un agent i par un scalaire d_i qui prend la valeur j s'il choisi la j -ième alternative parmi un ensemble de p alternatives concurrentes et mutuellement exclusives : $D_i = j$ pour $1 \leq j \leq p$.

Le choix de l'agent est dans une version multinomiale représenté par un vecteur p -dimensionnel à composantes booléennes. Soit Y_i , un vecteur $(p \times 1)$, dont la j -ième composante, Y_{ij} , prend la valeur 1 si le choix de l'agent i se porte sur l'alternative j , et dont les autres composantes Y_{i-j} prennent la valeur 0 : $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{ip})'$ avec $Y_{ij} = 1$ et $Y_{i-j} = 0$ si $D_i = j$.

Classiquement, l'agent est supposé rationnel et choisira l'alternative qui lui procure le meilleur niveau d'utilité Y_i^* parmi l'ensemble des p choix possibles. Le vecteur multinomial observable Y_i résulte ainsi de la discrétisation des niveaux de satisfaction Y_i^* partiellement inobservables : $Y_{ij} = 1$ si $\max_k (Y_{ik}^*) = Y_{ij}^*$.

Dans les modèles de choix, on spécifie usuellement la fonction d'utilité par une relation linéaire de ses déterminants X_i et d'un terme d'erreur aléatoire v_i . On abouti à un système de régressions des variables latentes suivant : $Y_{ik}^* = \underline{X}_i \tilde{\beta}_k + v_{ik}$ pour $1 \leq k \leq p$.

Pour des facilités de manipulations ultérieures, on peut réécrire le modèle matriciel sous une forme plus habituelle :

$$\begin{pmatrix} Y_{i1}^* \\ Y_{i2}^* \\ \dots \\ Y_{ip}^* \end{pmatrix} = \begin{pmatrix} \underline{X}_i & 0 & \dots & 0 \\ 0 & \underline{X}_i & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \underline{X}_i \end{pmatrix} \times \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \dots \\ \tilde{\beta}_p \end{pmatrix} + \begin{pmatrix} v_{i1} \\ v_{i2} \\ \dots \\ v_{ip} \end{pmatrix}$$

soit :

$$Y_i^* = [I_p \otimes \underline{X}_i] \cdot \text{vec}(\tilde{\beta}_1, \dots, \tilde{\beta}_p) + v_i$$

ou encore :

$$Y_i^* = X_i \tilde{\beta} + v_i$$

Pour un individu i , sa probabilité d'avoir choisi l'alternative j , $\Pr(D_i = j)$ est égale à :

$$\Pr(D_i = j) = \Pr\left(\bigcap_{k \neq j} Y_{ik}^* \leq Y_{ij}^*\right).$$

Cette probabilité repose sur le type supposé de distribution suivie par les aléas v_i . En prenant v_{ik} indépendants entre agents, non corrélés entre alternatives, et suivant une loi valeur extrême, on abouti moyennant quelques transformations au modèle logit multinomial de McFadden (1973). L'hypothèse de non corrélation des aléas lui vaut de supposer l'indépendance des alternatives non pertinentes¹. Si certaines modalités retenues sont de proches substituts, cette hypothèse ne peut être crédible². Une façon de palier à cette défaillance est lever l'hypothèse d'indépendance des aléas entre alternatives et de donner une structure générale à la matrice des covariances des aléas v_i , par exemple en considérant que $v_i = (v_{i1}, \dots, v_{ip})'$ est normalement et indépendamment distribué : $v_i \sim \mathcal{N}_p(0, \Sigma^*)$, avec Σ^* une matrice de covariances $p \times p$ sans restriction d'indépendance des aléas entre alternatives de choix. Le modèle ainsi défini caractérise le modèle Probit Multinomial (Hausman et Wise, 1978).

- Identification

Le modèle spécifié n'est pas formellement identifiable. Notamment, l'ajout d'un scalaire quelconque dans chaque niveau d'utilité apporté par les alternatives laisse inchangé la structure des choix de l'agent. Ce problème de calage des paramètres (« location invariance ») est résolu en différenciant le système des régressions des variables latentes avec l'une d'entre elles, en choisissant une alternative de référence (ici l'alternative p) :

$$W_{ik} = Y_{ik}^* - Y_{ip}^* = \underline{X}_i (\tilde{\beta}_k - \tilde{\beta}_p) + (v_{ik} - v_{ip})$$

Ce passage aux différences permet d'observer que l'on ne pourra, à ce niveau, identifier que les vecteurs de paramètres $\beta_k = (\tilde{\beta}_k - \tilde{\beta}_p)$. Intuitivement, on comprend aisément que les p niveaux d'utilité n'apportent pas plus d'information que les $p-1$ différences

¹ Independence of irrelevant alternatives (IIA) hypothesis

² Pour l'illustrer, il est souvent fait référence au paradoxe bus bleu - bus rouge.

d'utilités relatives à celle résultant du choix p . Il est donc équivalent d'écrire le modèle MNP sous cette forme, pour $1 \leq k \leq p-1$:

$$\begin{aligned} W_{ik} &= \underline{X}_i \beta_k + \varepsilon_{ik} \\ W_{ip} &= 0 \\ D_i &= j \text{ si } W_{ij} = \max_k (W_{ik}) \text{ et } W_{ik} > 0 \\ D_i &= p \text{ si } \max_k (W_{ik}) < 0 \end{aligned}$$

Matriciellement, pour $1 \leq k \leq p-1$:

$$\begin{pmatrix} W_{i1} \\ W_{i2} \\ \dots \\ W_{ip-1} \end{pmatrix} = \begin{pmatrix} \underline{X}_i & 0 & \dots & 0 \\ 0 & \underline{X}_i & \dots & 0 \\ 0 & \dots & \underline{X}_i & 0 \\ 0 & \dots & 0 & \underline{X}_i \end{pmatrix} \times \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \dots \\ \varepsilon_{ip-1} \end{pmatrix}$$

soit :

$$W_i = (I_{p-1} \otimes \underline{X}_i) \times \text{vec}(\beta_1, \dots, \beta_{p-1}) + \varepsilon_i$$

ou encore :

$$W_i = X_i \beta + \varepsilon_i$$

avec $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{ip-1})'$, $\varepsilon_i \sim \mathcal{N}(0, \Omega)$, et $\Omega = [I_{p-1} \mid -e_{p-1}] \Sigma^* [I_{p-1} \mid -e_{p-1}]'$

Si le problème du calage des paramètres est résolu par la différenciation, l'identification des paramètres du modèle n'est pas encore acquise. Notamment, la multiplication par un scalaire positif non nul ($c > 0$) de l'ensemble des β_k laisse également inchangée la structure des préférences des agents entre les alternatives :

$$D(W_i) = D(cW_i) \rightarrow D(X_i \beta + \varepsilon_i) = D(X_i c\beta + c\varepsilon_i)$$

Ce problème d'échelle peut être surmonté en fixant l'une des composantes de β , ce qui demande la connaissance a priori de son signe. Plus habituellement, comme dans le modèle probit binaire classique, on fixe un élément de la matrice de covariance de

l'erreur Ω , par exemple $\Omega_{11} = 1$. L'identification usuelle du modèle achevée, on peut facilement écrire la fonction de vraisemblance du modèle MNP :

$$L(Y | \beta, \Omega, X) = \prod_{i=1}^N L_i(Y_i | \beta, \Omega, X_i),$$

$$L(Y_i | \beta, \Omega, X_i) = \int_{Z_i} (2\pi)^{-0.5(p-1)} |\Omega|^{-0.5} \exp(-0.5(W_i - X_i\beta)' \Omega^{-1} (W_i - X_i\beta)) dZ_i$$

$$\text{avec : } Z_i = \left\{ \begin{array}{l} Z_i : \left(\bigcap_{k \neq j} (W_{ij} > W_{ik}) \right) \cap (W_{ij} > 0) \text{ si } D_i = j \text{ et } D_i \neq p \\ Z_i : \bigcap_k (W_{ik} < 0) \text{ si } D_i = p \end{array} \right.$$

L'estimateur de β et Ω par maximum de vraisemblance implique l'optimisation de la fonction $L(Y|\beta,\Omega,X)$ et qui demande le calcul d'une intégrale multiple de dimension $p-1$ particulièrement difficile (voire impossible) à évaluer lorsque p est supérieur à trois. Pour ce motif, l'estimation du modèle MNP s'est très souvent effectuée à l'aide des méthodes de simulations. Lerman et Manski (1981) ont par exemple proposé une approche par maximum de vraisemblance simulée où les probabilités de choix entre alternatives sont obtenues via un simulateur de fréquences relatives. McFadden (1989) a introduit le concept des moments simulés pour estimer des probabilités de choix en utilisant une version lissée de l'estimateur de fréquence. Les deux méthodes précédentes ont largement bénéficié du simulateur GHK (Geweke, Hajivassiliou, Keane) pour simuler les probabilités. Dans un comparatif des différentes méthodes d'ajustement du modèle MNP, Geweke et al. (1994) ont notamment conclu aux bonnes propriétés de l'approche bayésienne suivante.

1.2) Analyse Bayésienne du MNP

Nous reprenons ici l'approche bayésienne du MNP de McCulloch et Rossi (1994). Dans celle-ci, l'analyse du modèle passe par la spécification de distributions *a priori* sur

l'ensemble des paramètres d'intérêt β et Ω . Pratiquement, on choisit des distributions conjuguées et indépendantes.

$$P(\beta) \sim \mathcal{N}(b, A) \propto |A|^{-0.5} \exp\left(-0.5(\beta - b)' A^{-1}(\beta - b)\right)$$

$$P(\Omega^{-1}) = P(G) \sim \mathcal{W}_{p-1}(v, V) \propto |G|^{(v-p)/2} \exp\left(-0.5 \text{tr}(GV^{-1})\right)$$

Où \mathcal{W} est le symbole conventionnel de la distribution de Wishart³. Le produit de la fonction de vraisemblance du modèle MNP avec ces distributions a priori est proportionnel à la distribution a posteriori des paramètres :

$$P(\beta, G | X, Y) \propto L(Y | \beta, G, X) \times P(\beta, G)$$

Le caractère plus ou moins précis des distributions *a priori* relativement à l'information apportée par la vraisemblance sur les paramètres détermine leur degré d'influence sur la distribution a posteriori. On peut directement comparer le passage des distributions *a priori* vers la distribution *a posteriori* à un processus de révision bayésien des croyances : le modélisateur reflète ses opinions et leur force en premier lieu dans les distributions *a priori* des paramètres $(P(\beta), P(G))$, puis sous l'éclairage de l'information contenue par les données, il révisé ses croyances initiales en formulant une distribution *a posteriori* $(P(\beta, G | X, Y))$. Pour notre étude, il convient d'imposer des distributions *a priori* très diffuses pour que la distribution *a posteriori* épouse⁴ largement l'allure de la fonction de vraisemblance $L(Y | \beta, G, X)$. Pour $P(\beta)$ notamment, il suffit de choisir A avec des éléments diagonaux très grands, et pour $P(G)$, il convient de choisir une valeur pour $v > p$ suffisamment petite par rapport à la taille N de l'échantillon (McCulloch et Rossi, 1994).

³ Nommée ainsi en l'honneur du statisticien John Wishart (1898-1956). Soit une variable aléatoire U qui suit une $\mathcal{N}_T(0, \Gamma)$ et U_i la i -ième réalisation de U , alors la matrice $\sum_{i=1}^n U_i U_i'$ est distribuée selon une loi de Wishart d'hyper paramètre Γ avec n degrés de liberté, notée $\mathcal{W}_p(n, \Gamma)$.

⁴ Voir Nobile (1998) pour une illustration graphique.

2) La méthode MCMC (Monte Carlo Markov Chain)

Comme nous l'avons évoqué, l'évaluation de la seule fonction de vraisemblance est très difficile. Ainsi, les méthodes numériques pour l'analyse bayésienne s'avèrent également très décourageantes. L'approche simulative consistant à faire des tirages dans une distribution permet de s'affranchir du calcul complexe d'intégrales via une simple exploitation statistique de ces tirages. Nous utilisons dans cette partie l'échantillonnage de Gibbs (Geman et Geman, 1984) avec augmentation de données (Tanner et Wong, 1987) pour caractériser $P(\beta, G | X, Y)$. L'intérêt de cet échantillonnage repose sur le remarquable résultat que des tirages itératifs et récursifs dans des lois conditionnelles des paramètres donne asymptotiquement lieu à une séquence de réalisations qui converge en distribution vers la loi jointe des paramètres. Rappelons brièvement son principe de fonctionnement avec un exemple.

2.1) Echantillonnage de Gibbs

Soit θ , le vecteur de paramètres décomposé en θ_1 et θ_2 .

1 - On initialise l'algorithme d'échantillonnage en θ_1^0 et θ_2^0 .

2 - Puis on effectue une boucle de tirages dans les distributions conditionnelles :

θ_1^{g+1} tiré dans $f(\theta_1 | \theta_2 = \theta_2^g)$, θ_2^{g+1} tiré dans $f(\theta_2 | \theta_1 = \theta_1^{g+1})$

3 - retour en 1

Il en résulte une séquence θ^k qui converge en distribution vers la loi jointe $f(\theta_1, \theta_2)$.

L'étape d'augmentation de données vient considérablement simplifier l'analyse bayésienne du modèle MNP. Celle-ci, en ajoutant l'ensemble des variables latentes à l'espace des paramètres dans l'échantillonneur de Gibbs, transforme le modèle MNP en modèle Bayésien linéaire. L'échantillonnage du vecteur d'utilités relatives W dans une distribution multivariée et tronquée peut être conduit avec la méthode d'acceptation-rejet décrite par Albert et Chib (1993). Le désavantage de cette méthode est qu'il faut parfois effectuer plusieurs tirages avant d'accepter celui qui tombera dans la bonne région impliquée par le choix D_i .

On retient la démarche de McCulloch et Rossi, en partitionnant le vecteur des variables latentes en variables univariées dans l'échantillonneur de Gibbs. Ainsi, l'ensemble des variables devant être simulées est : $(W_1, W_2, \dots, W_{p-1}, \beta, G)$.

2.2) Les distributions conditionnelles a posteriori

On s'intéresse tout d'abord aux tirages des variables latentes $W_{ij} | W_{i-j}, \beta, G, Y_i$ qui suivent une loi univariée normale tronquée dans la région conforme à la décision de l'agent i : $W_{ij} \sim \mathcal{N}(m_{ij}, r_{ij})$. Les moyennes et variances conditionnelles peuvent facilement être déduites après avoir découpé et réorganiser la matrice $\Omega = G^{-1}$ en 3 sous ensembles. Notons Ω_{-j-j} : la matrice Ω dont on a retiré la j -ième ligne et la j -ième colonne, Ω_{-jj} : le vecteur de la j -ième colonne de Ω sans Ω_{jj} , Ω_{jj} : la variance de ε_{ij} , située à l'intersection de la j -ième ligne et la j -ième colonne dans Ω . Schématiquement et après avoir placé par convenance l'alternative j en dernière position:

$$\Omega = \begin{pmatrix} \left(\begin{array}{c} \Omega_{-j-j} \end{array} \right) & \left(\begin{array}{c} \Omega_{-jj} \end{array} \right) \\ \left(\begin{array}{c} \Omega'_{-jj} \end{array} \right) & \left(\Omega_{jj} \right) \end{pmatrix}$$

On dérive la variance conditionnelle de $\varepsilon_{ij} | \varepsilon_{i-j}$ à l'aide des formules usuelles : $Var(\varepsilon_{ij} | \varepsilon_{i-j}) = r_{ij} = \Omega_{jj} - \Omega'_{-jj} (\Omega_{-j-j})^{-1} \Omega_{-jj}$. De même pour l'espérance de $\varepsilon_{ij} | \varepsilon_{i-j}$, notée S_{ij} : $E(\varepsilon_{ij} | \varepsilon_{i-j}) = S_{ij} = \Omega'_{-jj} (\Omega_{-j-j})^{-1} (W_{i-j} - X_{i-j}\beta)$, où W_{i-j} est le vecteur colonne W_i imputé de W_{ij} ; et X_{i-j} est la matrice des caractéristiques X_i imputée de la j -ième ligne. Enfin, on calcule l'espérance de $W_{ij} | W_{i-j}$: $E(W_{ij} | W_{i-j}) = mij = X_{ij}\beta + S_{ij}$, où X_{ij} est la j -ième ligne de X_i . Il convient ensuite d'effectuer l'échantillonnage de W_{ij} dans la région qui respecte le choix de l'agent i :

$$W_{ij} \in]\max(0, W_{i-j}); +\infty[\text{ si } D_i = j \text{ ou,}$$

$$W_{ij} \in]-\infty; \max(W_{i-j})[\text{ si } D_i \neq j \text{ et } D_i \neq p$$

$$W_{ij} \in]-\infty; 0[\text{ si } D_i = p$$

Pour effectuer directement des tirages dans la bonne zone d'échantillonnage, on peut opter pour la méthode par inversion de la fonction de répartition (Devroye, 1986).

L'étape suivante consiste en l'échantillonnage du vecteur de paramètre β dans sa distribution *a posteriori* conditionnellement à W , G , et X : $\beta | W, X, G$. Il est particulièrement arrangeant de transformer les vecteurs X et W avec la décomposition de Cholesky de G pour se ramener au cas d'une régression avec des erreurs *iid* entre les alternatives. Soit C cette transformation, telle que $G=CC'$. Le modèle transformé devient :

$$\begin{aligned} C'W_i &= C'X_i\beta + C'\varepsilon_i, \\ W_i^* &= X_i^*\beta + \varepsilon_i^* \\ \varepsilon_i^* &\sim N(0, I_{p-1}) \end{aligned}$$

On déduit la distribution *a posteriori* de β : $\beta | W, X, G \sim \mathcal{N}(b_1, B)$, avec : $B = (A^{-1} + X^*X^*)^{-1}$; $b_1 = B(A^{-1}b + X^*W^*)$. Enfin, la distribution conditionnelle *a posteriori* de G sachant W , X , β suit une loi de Wishart : $G | W, X, \beta \sim \mathcal{W}_{p-1}(v + N, V_1)$, avec : $V_1 = \left(V^{-1} + \sum_{i=1}^N (W_i - X_i\beta)(W_i - X_i\beta)' \right)^{-1}$.

Ce résultat est facilement obtenu en multipliant la fonction de densité *a priori* $\mathcal{W}_{p-1}(v, V)$ par la fonction de vraisemblance conditionnelle sachant W , X, β qui est proportionnelle à :

$$\exp \left(-\frac{1}{2} \text{tr} \left(\sum_{i=1}^N (W_i - X_i\beta)(W_i - X_i\beta)' \right) G \right)$$

La simulation par tirage d'une distribution de Wishart peut être aisément conduite grâce à la décomposition de Bartlett⁵.

⁵ Exemple : prenons S une matrice $(T \times T)$ triangulaire inférieure telle que l'on échantillonne ses éléments (non nuls) dans les distributions suivantes : $s_{ii} \sim \sqrt{\chi_{r-i+1}^2}$, $s_{ij} \sim \mathcal{N}(0, 1)$. Alors la matrice $P = SS'$ est un tirage aléatoire dans une $\mathcal{W}_T(r, I)$. Soit K , la décomposition de Cholesky de R , de dimension $(T \times T)$ telle que $R = KK'$, alors la matrice KPK' est un tirage aléatoire dans une $\mathcal{W}_T(r, R)$.

2.3) Version hybride de l'échantillonnage de Gibbs

Nobile (1995) a proposé d'intégrer une étape supplémentaire aux cycles d'échantillonnage de Gibbs qui peut améliorer sa capacité à explorer la distribution jointe *a posteriori* dans la version bayésienne du MNP. Ainsi, l'auteur propose de changer l'échelle des paramètres à la fin de chaque cycle dans une étape « Metropolis ». En effet, il n'est d'aucune importance entre deux échantillonnages de Gibbs de changer l'échelle du point de vue des paramètres identifiés. Mais cela peut amener un parcours plus rapide de l'espace des paramètres pour mieux caractériser la fonction de densité *a posteriori*. Par ailleurs, Nobile (1995) observe sur des données fictives que l'échantillonneur de Gibbs seul peine à rendre significatif les éléments de covariances de Ω , alors que l'ajout de l'étape Metropolis semble améliorer quelque peu sa capacité à les révéler. Utiliser conjointement dans une procédure d'échantillonnage les algorithmes de Gibbs et de Metropolis donne naissance à un algorithme MCMC⁶ dit « hybride ».

⁶ MCMC pour Monte Carlo Markov Chain.

3) Les données «Parc Auto »

L'enquête PARC AUTO (PA) est réalisée annuellement auprès d'un échantillon de 10000 ménages français. Le taux de renouvellement est d'environ un tiers par an. Cette enquête est :

- Cofinancée par l'Agence De l'Environnement et de la Maîtrise de l'Energie (ADEME), le Comité des Constructeurs Français d'Automobiles (CCFA), la Fédération Française des Sociétés d'Assurances (FFSA) et la Direction de la Sécurité Routière (DSCR),
- Réalisée par la SOFRES auprès de ses panélistes-METASCOPE⁷ , et
- Exploitée par l'Institut National de Recherche sur les Transports et leur Sécurité (INRETS).

L'ambition de l'enquête est la connaissance générale du parc automobile français à la disposition des ménages, notamment en termes d'équipement et d'usage. Sur les 10000 ménages interrogés par voie postale, on observe un taux de retour des questionnaires d'environ 65% (6872 en 1999, 6368 en 2000 et 6438 en 2001). «Parc Auto» constitue une source de données particulièrement riche tant l'éventail des variables descriptives des véhicules qu'il contient est large. On y trouve notamment renseignées à un degré très fin les caractéristiques physiques des voitures (niveau de gamme, carburant utilisé, état d'acquisition, kilométrage...), mais aussi les caractéristiques individuelles de leurs utilisateurs (sexe, age, CSP...). Pour ce document, nous avons sélectionné dans les vagues annuelles d'enquête 1997-2001 les automobiles nouvelles acquises par les agents et convenablement renseignées. Au final, nous disposons de 6633 observations de nouvelles acquisitions pour estimer le modèle MNP.

Nous avons retenu un total de huit modalités de choix pour l'acquisition d'un nouveau modèle automobile, recoupant deux caractéristiques de motorisation (essence ou diesel), deux caractéristiques de gamme (bas et haut), puis deux états possibles à l'achat du véhicule (neuf et occasion). La motorisation du véhicule acquis est renseignée dans PARC AUTO par le type de carburant principal utilisé pour alimenter son moteur. Sans problème, les voitures catégorisées « essence » sont celles qui utilisent un supercarburant pour rouler (super, super sans plomb 95, super sans plomb 98).

⁷ <http://www.tns-sofres.com/sofres/panels.htm>

Similairement, les voitures « diesel » sont celles ayant utilisé le plus souvent du gasoil comme carburant. L'état du véhicule à l'achat est sans équivoque puisqu'il est explicitement demandé dans le questionnaire dont l'enquête est issue et codé à l'aide d'une variable binaire. L'échelle de gamme retenue par les constructeurs est également celle adoptée dans l'enquête Parc Auto. Celle-ci décline les véhicules en quatre niveaux explicites (bas, moyen inférieur, moyen supérieur, haut de gamme), distingue les véhicules 4*4 ainsi que ceux n'ayant trouvé leur « famille » (ou imprécis). Afin d'éviter la multiplication des alternatives de choix, nous avons opter pour un découpage plus simple, entre gamme inférieure et gamme supérieure. Pour des convenances d'allègement de rédaction, nous utiliserons la signalétique de la table 1 pour désigner les modèles automobiles.

Table 1 : Codes de désignations des alternatives automobiles

Caractéristiques	Code	Véhicule
Gamme	P	gamme inférieure
	G	gamme supérieure
Moteur	E	Moteur essence (fonctionnant au super carburants)
	D	Moteur diesel (fonctionnant au gasoil)
Etat	O	Occasion (seconde main)
	N	Neuf (première main)

Exemples : "PDO" désigne une automobile de gamme inférieure, diesel et d'occasion.

"PD" désigne une automobile de gamme inférieure et diesel (neuve ou d'occasion).

"P" désigne une automobile de gamme inférieure (neuve ou d'occasion, essence ou diesel). Etc.

Les variables explicatives retenues pour modéliser ce choix à l'acquisition tiennent tout d'abord compte de la motorisation passée des individus. En effet, si l'on souhaite révéler l'existence de comportements d'habitude, ou la fidélité des agents pour certains modèles d'automobile, les caractéristiques du véhicule remplacé doivent être considérées. Plus particulièrement à notre cas d'étude, quatre variables indicatrices relatives à l'ancien véhicule ont été créées : celles du type de motorisation (D ou E)⁸, et celles de la gamme

⁸ Cf. Table 1

(P ou G). Pour les agents accédant à leur première voiture, la création d'une indicatrice supplémentaire (pas de véhicule précédent) a été nécessaire. Ainsi, l'introduction de ces variables comme éléments explicatifs confère au modèle la possibilité d'obtenir quelques conclusions intéressantes sur la dynamique comportementale à l'acquisition automobile. De plus, «Parc Auto» détaille convenablement la catégorie socio professionnelle (CSP) d'appartenance des utilisateurs principaux des voitures. Ces classes ont la vertu de regrouper des individus assez homogènes, notamment en caractéristiques de revenus. Ainsi, notre étude utilisera huit indicatrices de CSP et sera en mesure de révéler des singularités entre classe de profession en répondant à la question phare de ce document : «qui achète quoi ». L'équation fonctionnelle d'estimation intègre également des caractéristiques socio démographiques comme facteurs explicatifs pouvant affecter, semble-t-il, le choix du type de véhicule. Notamment, la différenciation du sexe de l'agent pourra mettre en lumière de possibles différences comportementales d'acquisition entre hommes et femmes. Par ailleurs, l'âge de l'utilisateur principal devra être considéré pour notamment tenir compte du changements de préférence des agents au cours du temps. Mais aussi, en reflétant leur position au cours du cycle de vie, l'âge des agents est étroitement corrélé avec d'autres déterminants potentiellement explicatifs, mais parfois inobservés ; tels que la structure familiale du ménage, la situation dans l'emploi, le degré d'endettement pour l'habitat. Pour prendre en compte d'éventuelles non linéarités, la variable d'âge mise au carré est également introduite comme régresseur dans la spécification. Enfin, compte tenu des fortes évolutions du prix des carburants depuis 1999, il convient de s'interroger sur leurs effets dans le choix d'acquisition. Sur la période d'observation (1997-2001), l'évolution du prix de ces énergies fossiles a principalement suivi celui du prix du pétrole brut sans être véritablement influencée par des mouvements de taxation. Cette évolution, presque de concert, pourrait rendre le modèle incapable de suffisamment bien distinguer les effets propres d'une variation du prix du gasoil et ceux d'une variation du prix des super carburants sur le choix des agents, tant leurs évolutions parallèles en font des variables extrêmement corrélées. A cet égard, il semble préférable d'insérer comme facteur explicatif une variable synthétique du coût des carburants tel que l'indice des prix, librement disponible auprès de l'INSEE⁹. L'idée sous jacente est de vérifier l'idée selon laquelle les acquéreurs d'automobiles porteront

⁹ Institut national (français) de la statistique et des études économiques

d'avantage leurs préférences sur des véhicules diesel, bénéficiant d'un carburant moins cher et en moyenne moins énergivores, que sur des véhicules essence lorsque le prix des deux carburants principaux évoluent conjointement à la hausse (et réciproquement).

Enfin, les données décrivent le lieu de résidence de chaque acquéreur, permettant au modèle de mesurer son impact sur le type de véhicule choisi. Plus particulièrement, l'ambition est de vérifier l'existence de comportements spécifiques aux franciliens. Le découpage géographique classique retenu crée quatre indicatrices résidentielles : Paris (département 75), la petite couronne franciliennes (dép. 92 93 94), la grande couronne francilienne (dép. 77 78 91 95), et la province (départements hors Ile-de-France). La table 2 présente le codage utilisé pour l'ensemble des variables explicatives sélectionnées.

Table 2 : codage et libellé des variables explicatives du MNP

Variable	Désignation
C	Constante
RX	Indicatrice : 1 si l'agent n'avait pas d'automobile avant d'acquérir son véhicule.
EX	Indicatrice : 1 si la précédente automobile était essence
DX	Indicatrice : 1 si la précédente automobile était diesel
PX	Indicatrice : 1 si la précédente automobile était de gamme inférieure (petite)
GX	Indicatrice : 1 si la précédente automobile était de gamme supérieure (grande)
AGE	Age de l'agent
AGE2	Age ² /100 de l'agent
HOM	Indicatrice : 1 si l'agent est un homme
FEM	Indicatrice : 1 si l'agent est une femme
CSP1	Indicatrice : 1 si l'agent appartient à la CSP des Agriculteurs
CSP2	Indicatrice : 1 si l'agent appartient à la CSP des Artisans
CSP3	Indicatrice : 1 si l'agent appartient à la CSP des Cadres (et professions intellectuelles supérieures)
CSP4	Indicatrice : 1 si l'agent appartient à la CSP des Professions Intermédiaires
CSP5	Indicatrice : 1 si l'agent appartient à la CSP des Employés
CSP6	Indicatrice : 1 si l'agent appartient à la CSP des Ouvriers
CSP7	Indicatrice : 1 si l'agent appartient à la CSP des Retraités
CSP8	Indicatrice : 1 si l'agent appartient à la CSP des sans activités professionnelles
PARIS	Indicatrice : 1 si l'agent réside à Paris
PC	Indicatrice : 1 si l'agent réside en petite couronne francilienne
GC	Indicatrice : 1 si l'agent réside en grande couronne francilienne
PROV	Indicatrice : 1 si l'agent réside en province (hors Ile-de-France)
IPC	Indice des prix des carburants

4) Résultats d'estimation

Nous avons fait concourir trois séries d'échantillonnage des paramètres sur 200000 boucles, indépendamment initialisées en des points dispersés dans leur (diffuse) distribution à priori. Les tirages conservés pour l'analyse sont les 100000 derniers avec une fréquence de 1 tous les 5 tirages (soit 20000 pour chaque série). Les tests (non reportés) de Gelman-Rubin (1992), Brooks-Gelman (1998) concluent simultanément à rejeter l'hypothèse de leur non convergence vers la distribution *a posteriori* pour chaque paramètre. Fort de ce résultat, l'analyse des estimations est effectuée à partir du cumul des trois chaînes d'échantillonnage.

La table 3 contient les statistiques classiques caractérisant les distributions marginales empiriques des paramètres échantillonnés. Pour chaque composante de $\hat{\beta}$, on remarque que l'intervalle de crédibilité à 90% est relativement symétrique autour de leurs valeurs médiane et moyenne. Leur analyse graphique individuelle montre que l'on peut confortablement leur ajuster une distribution normale. La table 4 contient les résultats d'estimation relatifs à la matrice des covariances des erreurs : globalement, la méthode d'estimation n'a pas permis de mettre en évidence des covariances significatives.

5) Exploitation des résultats

Si l'on peut effectivement extraire à partir des résultats d'estimation une mesure concernant l'impact *ceteris paribus* d'un changement de caractéristiques individuelles sur le niveau des utilités relatives W d'un agent avec les méthodes usuelles, cela ne présente qu'un intérêt particulièrement limité. En effet, une telle démarche ne permet pas de mesurer directement l'impact induit sur la répartition des probabilités de choix entre alternatives, qui dépend à la fois du niveau des espérance des variables latentes W et de la structure générale donnée par le modèle probit multinomial à la matrice des covariances des termes d'erreur.

Table 3 : Estimation MNP (Coefficients β)

Variables	PEO	PEN	PDO	PDN	GEO	GEN	GDO
Référence : PX & EX (PEX)							
RX	0.088 [-0.051 0.215]	-0.792 [-1.116 -0.534]	-0.098 [-0.266 0.047]	-0.115 [-0.276 0.033]	-0.115 [-0.282 0.029]	-0.289 [-0.474 -0.134]	0.150 [-0.023 0.348]
DX	-1.100 [-1.298 -0.916]	-0.775 [-1.081 -0.534]	-0.016 [-0.172 0.156]	0.419 [0.197 0.727]	-0.671 [-1.001 -0.424]	-0.643 [-0.901 -0.438]	0.183 [-0.028 0.428]
GX	-0.674 [-0.858 -0.480]	-0.780 [-1.086 -0.532]	-0.568 [-0.822 -0.354]	-0.685 [-0.972 -0.445]	0.132 [-0.032 0.310]	0.092 [-0.035 0.239]	0.418 [0.179 0.723]
AGE	-0.049 [-0.069 -0.029]	-0.011 [-0.029 0.005]	-0.047 [-0.070 -0.027]	-0.009 [-0.026 0.008]	-0.009 [-0.027 0.010]	-0.009 [-0.026 0.008]	0.023 [-0.002 0.052]
AGE2	0.040 [0.019 0.063]	0.027 [0.009 0.048]	0.037 [0.017 0.060]	0.011 [-0.007 0.030]	0.000 [-0.021 0.021]	0.014 [-0.004 0.033]	-0.039 [-0.074 -0.010]
Référence : HOM							
FEM	0.333 [0.239 0.430]	0.246 [0.147 0.364]	0.049 [-0.038 0.139]	-0.042 [-0.137 0.040]	-0.094 [-0.203 0.003]	-0.009 [-0.102 0.082]	-0.048 [-0.164 0.057]
Référence : CSP6							
CSP1	0.104 [-0.328 0.548]	-2.529 [-5.438 -0.657]	0.414 [0.040 0.848]	-0.263 [-0.790 0.205]	-0.166 [-0.596 0.237]	-1.745 [-4.375 -0.176]	-0.092 [-0.572 0.376]
CSP2	-0.249 [-0.516 0.014]	-0.066 [-0.314 0.174]	-0.329 [-0.596 -0.098]	-0.094 [-0.317 0.123]	-0.380 [-0.660 -0.145]	0.078 [-0.152 0.313]	-0.445 [-0.768 -0.174]
CSP3	-0.762 [-0.956 -0.576]	0.019 [-0.143 0.204]	-0.559 [-0.798 -0.342]	0.064 [-0.095 0.251]	-0.479 [-0.710 -0.290]	0.128 [-0.031 0.313]	-0.567 [-0.825 -0.342]
CSP4	-0.392 [-0.547 -0.244]	0.048 [-0.093 0.210]	-0.389 [-0.570 -0.230]	0.137 [0.000 0.310]	-0.445 [-0.653 -0.272]	0.105 [-0.041 0.274]	-0.393 [-0.595 -0.217]
CSP5	-0.092 [-0.234 0.045]	0.136 [-0.004 0.294]	-0.129 [-0.270 -0.003]	0.168 [0.035 0.331]	-0.092 [-0.228 0.032]	0.108 [-0.044 0.277]	-0.222 [-0.403 -0.067]
CSP7	-0.310 [-0.516 -0.111]	0.055 [-0.116 0.240]	-0.384 [-0.615 -0.187]	-0.009 [-0.171 0.163]	-0.260 [-0.468 -0.081]	0.076 [-0.098 0.264]	-0.333 [-0.586 -0.120]
CSP8	-0.080 [-0.247 0.085]	0.091 [-0.073 0.267]	-0.103 [-0.265 0.047]	0.007 [-0.153 0.170]	-0.116 [-0.284 0.036]	0.068 [-0.124 0.271]	-0.412 [-0.665 -0.198]
Référence : PROV							
PARIS	-0.047 [-0.320 0.223]	-0.062 [-0.303 0.172]	-0.344 [-0.680 -0.059]	-0.548 [-0.965 -0.226]	-0.067 [-0.316 0.170]	-0.014 [-0.229 0.197]	0.108 [-0.185 0.416]
PC	0.087 [-0.088 0.267]	0.282 [0.124 0.467]	-0.109 [-0.303 0.062]	0.007 [-0.150 0.161]	0.220 [0.062 0.406]	0.266 [0.123 0.437]	-0.088 [-0.318 0.116]
GC	0.075 [-0.063 0.214]	0.126 [0.003 0.266]	-0.196 [-0.364 -0.056]	-0.044 [-0.169 0.073]	0.020 [-0.104 0.145]	0.117 [0.007 0.239]	-0.084 [-0.250 0.067]
IPC	-0.335 [-0.737 0.064]	-0.745 [-1.215 -0.346]	0.253 [-0.100 0.652]	0.649 [0.273 1.127]	-0.234 [-0.626 0.119]	-0.684 [-1.133 -0.304]	0.150 [-0.280 0.610]
C	2.147 [1.418 2.897]	1.010 [0.376 1.705]	1.224 [0.606 1.911]	-0.746 [-1.731 -0.008]	0.772 [0.136 1.454]	0.617 [0.017 1.250]	-0.814 [-1.955 0.104]

Note : moyennes des échantillonnages. Intervalles de crédibilité à 90% entre parenthèses.

Table 4 : Estimation MNP (Covariances des erreurs Ω)

	PEO	PEN	PDO	PDN	GEO	GEN	GDO
PEO	1.000	-0.212	0.038	-0.081	0.096	0.202	-0.088
		[-0.506 0.073]	[-0.306 0.359]	[-0.442 0.226]	[-0.194 0.403]	[-0.010 0.385]	[-0.532 0.323]
PEN		0.642	0.024	0.145	0.015	0.063	-0.048
		[0.272 1.239]	[-0.206 0.291]	[-0.067 0.447]	[-0.214 0.257]	[-0.085 0.251]	[-0.334 0.231]
PDO			0.614	-0.035	-0.065	-0.045	-0.132
			[0.239 1.169]	[-0.272 0.179]	[-0.322 0.165]	[-0.198 0.102]	[-0.627 0.266]
PDN				0.455	-0.045	-0.032	0.036
				[0.175 0.928]	[-0.245 0.120]	[-0.163 0.081]	-[-0.205 0.335]
GEO					0.477	0.030	0.041
					[0.184 0.985]	[-0.113 0.181]	[-0.251 0.345]
GEN						0.278	0.009
						[0.128 0.529]	[-0.209 0.216]
GDO							0.877
							[0.337 1.618]

Note : moyennes des échantillonnages. Intervalles de crédibilité à 90% entre parenthèses.

Aux résultats des moyennes, la mesure du système de probabilités estimées d'un agent i réclame l'évaluation d'une intégrale multiple dès lors que le nombre d'alternatives est supérieur ou égal à trois. Sa résolution analytique étant très décourageante (voire impossible), l'usage des techniques de simulations est le recours le plus judicieux pour convenablement l'approximer. Parmi les simulateurs de probabilités les plus populaires dans la communauté statistique, on peut citer le simulateur CFS¹⁰ (Lerman et Manski, 1981), le simulateur de Stern (1992), et le simulateur GHK¹¹ (Hajivassiliou et al., 1996). On pourra lire Aris et Cox (1999) pour une introduction et une discussion de ces trois outils. Concernant notre étude, nous avons choisi d'utiliser le simulateur CFS, qui possède de bonnes propriétés usuelles: celui-ci est sans biais, de variance minimale (Lerman et Manski, 1981). Son principe de fonctionnement adapté à notre problème est intuitivement simple : le simulateur effectue T tirages dans la distribution des utilités relatives évaluée aux résultats d'estimation, puis la probabilité estimée du choix d'une

¹⁰ CFS pour « Crude Frequency Simulator ».

¹¹ GHK pour Geweke-Hajivassiliou-Keane, du nom des trois auteurs ayant contribué à l'élaboration du simulateur lisse récursif SRC (« Smooth Recursive Simulator »).

alternative j (pour $j \neq p$) est la fréquence de tirage avec laquelle W_{ij} est à la fois positive et maximale sur l'ensemble des utilités relatives W_{ik} . La probabilité estimée du choix de la modalité de référence p correspond à la fréquence de tirage avec laquelle toutes les variables latentes échantillonnées sont simultanément négatives. Schématiquement, les étapes de simulations des probabilités de choix sont les suivantes :

- 1) Soit Γ la transformation de Cholesky de $\hat{\Omega}$, telle que $\Gamma\Gamma' = \hat{\Omega}$,
 - 2) Soit A_s le s -ième tirage dans une $\mathcal{N}_{p-1}(0, I_{p-1})$, avec $s = [1 : T]$,
 - 3) Construire $\widehat{W}_{is} = \left(\text{vec}(X_i\hat{\beta}) \right) + \Gamma A_s$
 - 4) Calculer $\widehat{\text{Pr}}(d_i = j) = \frac{1}{T} \sum_{s=1}^T \mathbf{1}(\max_k (\widehat{W}_{iks}) = \widehat{W}_{ijs}) \mathbf{1}(\widehat{W}_{ijs} > 0)$ pour $j \neq p$,
- Calculer $\widehat{\text{Pr}}(d_i = p) = 1 - \sum_{j=1}^{p-1} \widehat{\text{Pr}}(d_i = j)$

Outre l'interprétation classique en termes de probabilités, les mesures $\widehat{\text{Pr}}$ peuvent de façon équivalente être décrite comme des « parts de marché ». Par exemple, si $\widehat{\text{Pr}}(d_i = j)$ mesure la probabilité estimée que l'individu i choisisse l'alternative j , c'est également dans ce contexte la part de marché estimée associée au modèle automobile j pour une « clientèle » ayant les caractéristiques de l'individu i . Rappelons que la distribution des probabilités entre les différentes alternatives de choix dépend du niveau de l'espérance des variables latentes, et donc des caractéristiques individuelles. De fait, l'analyse *ceteris paribus* des variations du système de probabilités estimées résultant d'un changement de caractéristiques ne peut être considérée qu'au cas par cas. Pour l'appliquer, il convient de déterminer un individu particulier, choisi conventionnellement à la moyenne des caractéristiques de l'échantillon. C'est donc sur la base des caractéristiques de l'individu moyen, parfois appelé individu « synthétique », que s'effectuera l'analyse « toutes autres caractéristiques égales avec l'agent moyen ».

Pour comparer les probabilités d'un même système (de l'agent i), un outil d'analyse particulièrement intuitif et largement utilisé dans la littérature des modèles de choix discrets est le « *odd* ». Celui-ci (noté O_{A-B}^i) définit le rapport de probabilités entre deux ensembles (A et B) d'alternatives mutuellement exclusifs, et s'interprète comme le rapport de « chances » de choisir l'un plutôt que l'autre .

Pour comparer les « *odd* » du choix d'un ensemble (*A*) d'alternatives sur un autre (*B*) entre deux individus *i* et *i'* différents, nous pouvons une interprétation en « *odds ratios* » (noté $OR_{A-B}^{i-i'}$)¹².

Effet de la motorisation passée (table 5)

Une propriété de la spécification est de pouvoir mettre en lumière un effet de dépendance du choix d'acquisition aux caractéristiques du véhicule précédent. L'estimation (table 3) montre une significative fidélité des agents au type de véhicule remplacé, tant au niveau de motorisation (E ou D) que sur le plan du niveau de gamme (P ou G).

Table 5 : répartition des probabilités d'acquisition
selon le type de véhicule précédent

Type acquis	Type de véhicule précédent ⁽¹⁾				
	RX	PEX	GEX	PDX	GDX
PEO	45.1%	34.2%	19.7%	10.9%	4.7%
PEN	10.3%	32.1%	13.9%	10.7%	3.3%
PDO	15.2%	12.5%	8.3%	23.3%	13.8%
PDN	9.1%	6.7%	2.7%	35.8%	15.1%
GEO	8.5%	7.5%	22.2%	3.3%	8.6%
GEN	1.2%	2.1%	11.5%	0.9%	3.6%
GDO	7.7%	4.0%	16.0%	10.1%	30.8%
GDN	2.9%	0.9%	5.7%	5.0%	20.1%

Notes : évaluation aux caractéristiques de l'agent synthétique, 25000 simulations du CFS. Types définis en table 1.

(1) : pour alléger les écritures : PEX équivaut à l'intersection de PX et de EX (une voiture précédente essence et de gamme inférieure), GEX équivaut à l'intersection de GX et de EX (une voiture précédente essence et de gamme supérieure), etc.

Toutes autres caractéristiques équivalentes avec celles de l'individu synthétique, un agent sans véhicule précédent possède près de 2 fois plus de chances de se reporter son

¹² $OR_{A-B}^{i-i'} = O_{A-B}^i / O_{A-B}^{i'}$

choix d'acquisition sur un véhicule essence que sur une voiture diesel : $\widehat{O}_{E-D}^{RX} = 1.9$ ($\widehat{O}_{E-D}^{PEX} = 3.1$, $\widehat{O}_{E-D}^{GEX} = 2.1$). Au contraire, si la voiture précédente de l'agent avait été PD, alors le modèle estime qu'il a près de 3 fois plus de chance de choisir un véhicule diesel qu'essence : $\widehat{O}_{D-E}^{PDX} = 2.9$ ($\widehat{O}_{D-E}^{GDX} = 4$). Ainsi, l'ancien propriétaire d'un véhicule PD a près de 9 fois plus de chances qu'un agent précédemment équipé d'un véhicule PE de choisir un véhicule diesel plutôt qu'essence : $\widehat{O}_{D-E}^{PDX-PEX} = 9.1$. Similairement, un agent dont l'ancien véhicule était GD a 8 fois plus de chances qu'un agent se séparant d'une voiture GE de choisir un véhicule diesel plutôt qu'essence : $\widehat{O}_{D-E}^{GDX-GEX} = 8.1$.

Lorsque l'individu synthétique n'avait pas de véhicule précédent, la part de marché la plus élevée à l'acquisition d'un véhicule est celle des voitures PE (55%) et plus généralement celle des voiture de gamme inférieures (79.7%). Dans des mesures comparables, lorsque l'agent moyen remplace son ancien véhicule PE, c'est pour une automobile de même gamme et utilisant le même carburant avec une probabilité estimée de 66.3%, alors que sa probabilité de choisir plus généralement un véhicule bas de gamme est de 85.5% : l'agent a donc près de 4 fois plus de chance de choisir un véhicule de gamme inférieure que supérieure. Le constat est beaucoup moins tranché pour l'agent ancien propriétaire d'un GE : en effet, celui ci a quasiment autant de chance d'acquérir un véhicule PE qu'un nouveau GE. Au contraire, si l'individu synthétique devait choisir de remplacer un véhicule GD, il choisirait dans 51% des cas une automobile équivalente,. Similairement, la part de marché des véhicules PD pour les agents détenant auparavant une voiture de même type est de 59.1%, et choisira près de 8 fois sur 10 un véhicule de bas de gamme.

Effet du sexe de l'acquéreur (table 6)

Le modèle met également en évidence un comportement d'acquisition significativement différent selon de sexe de l'acquéreur. Plus particulièrement, les véhicules PE sont la modalité privilégiée des femmes : toutes choses égales par ailleurs avec l'individu synthétique, la part de marché des véhicules PE est de 53.8% pour les femmes, contre 38.5% pour les hommes. De plus la probabilité d'acquérir une automobile de gamme inférieure pour les hommes est de 10 points inférieure à celle des femmes (78.8% contre 68.8% pour les hommes).

Table 6 : répartition des probabilités
d'acquisition selon le sexe

Type acquis	Sexe de l'agent	
	Femme	Homme
PEO	32.6%	23.0%
PEN	21.2%	15.3%
PDO	15.7%	17.3%
PDN	9.3%	13.2%
GEO	7.1%	11.5%
GEN	2.0%	3.5%
GDO	8.6%	11.1%
GDN	3.4%	5.0%

Note : évaluation aux caractéristiques de l'agent synthétique,
25000 simulations du CFS. Types définis en table 1.

Par ailleurs, on observe le même écart de points entre les parts de marché des véhicules essence entre les deux sexe (63% pour les femmes et 53.3% pour les hommes), les femmes présentant un $O_{E-D}^{fem} = 1.7$ contre $O_{E-D}^{hom} = 1.1$ pour les hommes. Les femmes ont ainsi 1.5 fois plus de chances que les hommes de choisir une automobile essence plutôt que diesel lors de l'acquisition d'une automobile.

Si l'on vient de vérifier la préférence relative des hommes pour les automobiles haut de gamme puis diesel, les résultats montrent que les deux sexes ont approximativement autant de chance de choisir un véhicule neuf qu'un véhicule d'occasion : la part de marché estimée des voitures d'occasion est de 64% pour les femmes et de 63% pour les hommes.

Effet de la catégorie socioprofessionnelle (CSP) (table 7)

L'un des comportements les plus atypiques dans l'acquisition d'une automobile reflété par les résultats du modèle est celui des agents issus de la catégorie des agriculteurs. A toutes autres caractéristiques égales avec l'agent moyen, un individu appartenant à la

CSP1 aura recours 9 fois sur 10 au marché de l’occasion, et choisira un véhicule diesel près de 6 fois sur 10. Plus spécifiquement, la part de marché associée aux véhicules DO est la plus forte au regard des autres CSP (50.7%). Cependant, il convient d’être prudent avec ces conclusions tant les observations appartenant à cette classe professionnelle sont peu représentées dans la base de données : celles-ci représentent moins de 1% des effectifs totaux (46 observations sur 6633). Il existe donc un risque plus élevé que dans l’ensemble des autres CSP d’avoir un échantillon insuffisamment représentatif du véritable comportement d’acquisition des agents agriculteurs.

Table 7 : répartition des probabilités d’acquisition
selon la CSP d’appartenance de l’agent

Type acquis	Catégorie Socioprofessionnelle de l’agent							
	CSP1	CSP2	CSP3	CSP4	CSP5	CSP6	CSP7	CSP8
PEO	31.6%	28.6%	15.9%	23.7%	28.3%	29.9%	26.0%	29.9%
PEN	1.0%	16.3%	21.4%	19.5%	17.3%	12.7%	19.4%	17.5%
PDO	38.9%	16.9%	12.8%	14.4%	17.2%	20.5%	13.9%	19.5%
PDN	4.5%	9.9%	16.8%	15.8%	12.1%	7.5%	10.9%	9.1%
GEO	8.3%	8.1%	7.7%	6.4%	10.5%	12.3%	10.0%	11.0%
GEN	1.0%	3.9%	7.4%	4.6%	2.3%	1.2%	3.6%	2.3%
GDO	11.8%	9.1%	8.0%	9.6%	10.0%	13.7%	10.6%	7.4%
GDN	2.9%	7.2%	9.9%	6.0%	2.2%	2.1%	5.5%	3.4%

Note : évaluation aux caractéristiques de l’agent synthétique, 25000 simulations du CFS. Types définis en table 1.

Les membres de la catégorie « cadres » (CSP3) montrent également des spécificités d’acquisition. *Ceteris paribus* avec l’agent synthétique, un individu issu de la CSP3 présente la propension la plus élevée à l’acquisition d’un modèle automobile neuf : selon les résultats du modèle, celui-ci n’a recours au marché de l’occasion qu’avec une probabilité de 44.5%, alors qu’elle est systématiquement supérieure à 50% dans l’ensemble des autres CSP. Notamment pour cet individu, la part de marché des véhicules DN est de loin la plus élevée (26.7%). Relativement aux autres catégories de profession, sa préférence pour les modèles haut de gamme est tout aussi remarquable. A titre de comparaison, cet agent a *ceteris paribus* près de 5 fois plus de chance qu’un

employé (CSP5) de s'équiper d'un véhicule GDN plutôt que d'un autre véhicule ($\widehat{OR}_{GDN-GDN}^{CSP3-CSP5} = 4.88$)¹³.

Dans une mesure moins tranchée que pour un agent issu de la CSP3, mais de façon comparable, un agent appartenant à la classe des professions intermédiaires (CSP4) présente une propension relativement élevée à recourir au marché des automobiles neuves, la part de marché associée étant de 45.8%. Pour comparaison et à toutes autres choses égales, l'agent issu de la CSP4 a près de 3 fois plus de chance qu'un ouvrier de choisir un véhicule neuf plutôt que d'occasion ($\widehat{OR}_{N-O}^{CSP4-CSP6} = 2.74$).

Hormis le cas spécifique de l'agent agriculteur, les agents cadre ou de professions intermédiaires présentent, à toutes autres caractéristiques égales, les plus fortes propensions à l'achat de véhicules diesel : respectivement 47.5% et 45.9%, alors que cette part de marché estimée est d'environ 6 points inférieure pour les agents employés ou inactifs.

Le comportement d'acquisition d'un ouvrier (CSP6) présente la singularité d'un fort recours au marché de l'occasion : toutes autres caractéristiques égales avec l'agent moyen, celui-ci a près de trois fois plus de chance d'acquérir une automobile de seconde main qu'un véhicule neuf ($\widehat{O}_{O-N}^{CSP6} = 3.24$). Pour cet agent, les parts de marché des véhicules PO (50.4%) et GO (26%) sont les plus élevées comparativement à celles des autres CSP (CSP1 exceptée). Outre l'explication par le niveau de revenu, cette préférence révélée des ouvriers pour les voitures de seconde main pourrait résulter en partie de leur capacité à pouvoir apprécier la véritable qualité des automobiles proposées sur le marché de l'occasion, ainsi que de leur aptitude à pouvoir réparer soi-même des automobiles plus âgées avec peu d'électronique embarquée.

Même s'il est moins important, on note également un recours assez élevé aux véhicules d'occasions chez les agents issus de la catégories des employés (CSP5) et des agents sans activité professionnelle (CSP8) : les probabilités respectives d'acquérir ces véhicules sont 66% et 67.8%. A toutes autres caractéristiques égales, un agent appartenant à l'une des catégories précédentes aura approximativement 2.5 fois moins de chance d'acquérir une automobile neuve qu'un agent cadre ($\widehat{OR}_{O-N}^{CSP5-CSP3} = 2.42$; $\widehat{OR}_{O-N}^{CSP8-CSP3} = 2.62$).

¹³ \overline{GDN} pour l'ensemble des alternatives, GDN exclue.

Effet de la zone résidentielle (table 8)

A toutes autres caractéristiques égales avec l'individu moyen, un agent parisien affiche une préférence plus marquée pour les véhicules de gamme supérieure que l'ensemble des autres résidents franciliens ou que les provinciaux. En effet, la probabilité estimée de choisir une voiture haut de gamme est de 37%, alors qu'elle est globalement inférieure de 10 points pour les autres individus. Ainsi, un agent résidant en grande couronne francilienne a 1.6 fois plus de chance qu'un agent résident à Paris de choisir un véhicule bas de gamme plutôt que de gamme supérieure. ($\widehat{OR}_{P-G}^{GC-Paris} = 1.60$).

Table 8 : répartition des probabilités d'acquisition
selon la zone de résidence

Type acquis	Zone de résidence de l'agent			
	PARIS	PC	GC	Province
PEO	28.7%	25.4%	29.0%	25.8%
PEN	19.2%	24.5%	21.7%	16.7%
PDO	11.3%	12.0%	11.7%	17.7%
PDN	3.8%	9.9%	10.8%	12.6%
GEO	10.2%	12.8%	9.6%	9.7%
GEN	3.6%	5.5%	3.9%	2.7%
GDO	14.7%	7.3%	8.7%	10.3%
GDN	8.5%	2.6%	4.6%	4.4%

Notes : évaluation aux caractéristiques de l'agent synthétique, 25000 simulations du CFS.

PC et GC respectivement pour petite et grande couronne francilienne.

Types définis en table 1.

Relativement à un provincial, les franciliens quelque soit leur zone de résidence présentent une préférence singulièrement plus élevée pour les véhicules à moteur essence. Selon les résultats du modèle, un agent de petite couronne à plus de 2 fois plus de chance d'acquérir un véhicule essence qu'une automobile diesel, contre 1.2 pour l'agent résidant en province ($\widehat{O}_{E-D}^{Paris} = 1.61$, $\widehat{O}_{E-D}^{GC} = 1.79$). Ce résultat occulte néanmoins un comportement tout à fait remarquable de l'agent résident à Paris : son recours aux

automobiles diesel s'effectue dans une large mesure plus fréquemment vers les voitures de gamme supérieure au détriment des voitures bas de gamme, alors qu'on observe un comportement opposé pour les résidents des autres zones. Ainsi, les probabilités d'acquisition des véhicules GD et PD sont respectivement évaluée à 23.3% et 15.1% pour l'agent parisien, alors qu'elles sont respectivement de 9.9% et 21.9% pour l'agent résident en petite couronne par exemple. Plus explicitement encore, l'agent parisien a 2.6 fois plus de chance que l'agent *ceteris paribus* résidant en grande couronne francilienne de choisir une voiture GD plutôt qu'une voiture PD ($\widehat{O}_{GD-PD}^{Paris-GC} = 2.61$).

Enfin, l'agent de petite couronne présente la propension la moins élevée à recourir au marché de l'occasion pour l'acquisition d'une automobile ($\widehat{\Pr}[d_{PC} = O] = 57.5\%$).

Effet des prix des carburants (table 9)

L'analyse *ceteris paribus* précédente reposait sur des comparaisons de systèmes de probabilités lorsque les caractéristiques des agents différaient en valeurs d'indicatrices. Au contraire, l'indice des prix agrégé des carburants étant une variable continue, sa variation permet d'évaluer les changements induits sur les parts de marchés des différentes alternatives en termes d'élasticités. Ainsi, les commentaires suivants relatent l'élasticité des parts de marché lorsque l'on simule une augmentation de 20% du niveau moyen de la variable IPC, aux caractéristiques de l'agent synthétique.

Les parts de marché les plus sensibles et négativement liées à la hausse du niveau des prix des carburants sont les véhicules EN, dont l'élasticité estimée pour les véhicules haut et bas de gamme sont respectivement de -1.98 et -1.33. En niveau, les variations induites correspondantes sont respectivement une diminution de 1.2 et 4.6 points, et plus globalement, la probabilité estimée d'acquérir un véhicule EN passe de 19.9% à 14.3% après l'augmentation de 20% de l'indice moyen des prix des carburants.

Les élasticités mesurée par le modèle des probabilités d'acquisition des voitures EO sont d'amplitude moins élevée que les précédentes, et très voisines : -0.35 pour les véhicules de gamme inférieure, et -0.34 pour les voitures haut de gamme. Dans l'ensemble, la part de marché des véhicules EO après une augmentation simulée de 20% du niveau moyen de la variable IPC passe de 36.2% à 33.7%.

Table 9 : effet de la hausse de 20% de l'indice
des prix moyen IPC

Type acquis	Répartition		Elasticité
	initiale	finale	
PEO	26.1%	24.3%	-0.35
PEN	17.1%	12.5%	-1.33
PDO	16.9%	19.2%	+0.70
PDN	12.4%	17.2%	+1.93
GEO	10.1%	9.4%	-0.34
GEN	2.9%	1.7%	-1.98
GDO	10.2%	10.8%	+0.29
GDN	4.3%	4.8%	+0.47

Notes : évaluation aux caractéristiques de l'agent synthétique, 25000 simulations du CFS.

Types définis en table 1. Les élasticités mesurent les sensibilités moyennes des probabilités rapportées à une augmentation de 1% de l'indice des prix des carburants moyen (intervalles de confiance associés non reportés). La répartition initiale (finale) correspond au système de probabilité d'acquisition de l'agent synthétique avant (après) l'augmentation de 20% de la moyenne de la variable IPC.

Ainsi, l'effet d'une hausse du prix moyen des carburants se traduit systématiquement par une diminution de l'attrait des agents pour les véhicules essence. Mécaniquement, elle marque au contraire un regain de préférence pour les voitures diesel.

Plus spécifiquement, la hausse de l'indice des prix augmente la probabilité pour l'agent synthétique de choisir un véhicule PDO (de 16.9% à 19.2%) avec une élasticité mesurée de +0.7. Par ailleurs, l'effet induit sur la part de marché des voitures GDO (de 10.2% à 10.8%) correspond à une élasticité de +0.29.

Mais ce sont surtout les alternatives en automobile DN qui rapportent les sensibilités positives au prix des carburants les plus élevées : leur part de marché passe de 16.8% à 22% sous l'impulsion de la hausse de 20% du niveau moyen de la variable IPC, établissant son élasticité à +1.55. Cette valeur particulièrement élevée tient surtout à l'augmentation induite de la probabilité pour l'agent moyen de choisir une voiture PDN, qui présente une élasticité estimée à +1.93 (contre +0.47 pour celle d'acquérir une automobile GDN). Au niveau agrégé, le modèle estime que l'agent moyen a, après la

hausse simulée de \overline{IPC} ¹⁴, près de 1.4 fois plus de chances de choisir un véhicule diesel plutôt qu'essence, qu'initialement ($\widehat{OR}_{D-E}^{1.2\overline{IPC}-\overline{IPC}} = 1.39$).

Conclusion

Dans ce document, nous avons ajusté la version bayésienne du modèle probit multinomial sur les données française de « Parc Auto », afin de décrire le comportement d'acquisition automobile des résidents français. Les résultats mettent en lumière le rôle prédominant des caractéristiques de l'automobile précédente sur le choix du véhicule, marquant la fidélité des agents pour le type de leur ancien modèle. Par ailleurs, l'effet du sexe montre une préférence plus élevée des femmes pour les automobiles essence de gamme inférieure. L'influence du revenu est appréhendée au regard de la différence des comportements d'acquisition, toutes choses égales par ailleurs, entre les différentes catégories professionnelles (CSP). Fort logiquement, les résultats *ceteris paribus* mesurent une préférence significative pour les véhicules neufs des agents issus des classes les plus riches (CSP3, CSP4), relativement à ceux appartenant aux classes plus modestes (CSP6, CSP8). La modélisation permet également de dégager un comportement atypique des franciliens qui vouent une préférence plus marquée que les provinciaux (spécialement les résidents des couronnes) pour les véhicules essences. Enfin, la progression du niveau des prix des carburants marque un report significatif des choix vers les véhicules diesels au détriment des voitures roulant aux « supercarburants ».

¹⁴ \overline{IPC} pour la moyenne de la variable IPC sur l'échantillon.

Bibliographie

- Arias, C. and Cox, T.L., 1998. Maximum Simulated Likelihood: A brief Introduction for Practitioners, University of Wisconsin-Madison, Department of Agricultural and Applied Economics, Staff Paper 421.
- Albert, J. and S. Chib, 1993. Bayesian analysis of binary and polychotomous data, *Journal of the American Statistical Association*, 88, 669-679.
- Brooks, S P. and Gelman, A., 1998. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*. 7. p434-455.
- Devroye, L.: 1986, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.
- Gelman, A., and Rubin, D.B., 1992, Inference From Iterative Simulation Using Multiple Sequences (with discussion), *Statistical Science*, 7, 457-472.
- Geman, S. and D. Geman, 1984, Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images, *IEEE Transactions on pattern analysis and machine intelligence*, 6, 721-741
- Geweke, J., Keane, M. and D. Runkle, 1994. Alternative Computational Approaches to Statistical Inference in the Multinomial Probit Model. *Review of Economics and Statistics*, 76:4, 609-32.
- Hajivassiliou, V., McFadden, D. & Ruud, P., 1996. Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and Computational results. *Journal of Econometrics*, 72 (1/2), 85-134
- Hausman, J., and D. Wise, 1978. A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences, *Econometrica*, Vol. 48, No. 2, March, pp. 403-426.
- Lerman, S. and C. Manski, 1981, "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with econometric Applications*, MIT Press.
- McCulloch R. and P. Rossi, 1994. An exact likelihood analysis of the multinomial probit model, *Journal of Econometrics*, 64, 207-240.
- Mc Fadden, D., 1973. "Conditional logit analysis of qualitative choice behavior", in P. Zarembka, ed., *Frontiers in econometrics*, New York : Academic Press.
- McFadden, D. L., 1989. A method of simulated moments for estimation of discrete response models without numerical integration, *Econometrica* 57, 995-1026.
- Nobile, A., 1998. A Hybrid Markov Chain for the Bayesian Analysis of the Multinomial Probit Model, *Statistics and Computing*, 8, 229 - 242.
- Stern, S., 1992. A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models. *Econometrica*, 60(1992): 943-52.
- Tanner, M.A. and W.H Wong, 1987. The calculation of posterior distributions by data augmentation, *Journal of American Statistical Association*, 82, 528-540.