Estimation d'un modèle de choix discret/continu dans un contexte de prix non linéaires

N-D Yameogo [1] J-T Bernard [2] et D. Bolduc[3]
Département d'économique, Université Laval

Avril 2006

(1) Étudiante au doctorat au département d'économique, GREEN, Université Laval
(2) Professeur au département d'économique, GREEN, CDAT, Université
(3) Professeur au département d'économique, GREEN, CDAT, Université Laval

1 Resumé

nous nous intéressons à l'estimation simultanée de trois choix : choix du mode de chauffage, choix de la tranche de consommation et choix de la quantité d'énergie. La plus part des travaux sur la demande d'énergie se limite à l'estimation simultanée de deux choix (mode de chauffage et quantité d'énergie ou tranche de consommation et quantité d'énergie). La demande d'énergie résidentielle est une démande dérivée. Le ménage a en effet besoin d'énergie pour satisfaire ses besoins. Pour cela, il doit choisir un stock de capital (mode ou alternative) qui lui fournit des services (chauffage de l'eau et de l'espace) pour maximiser son utilité. Le choi de la quantité d'énergie est réliée au choix du mode de chauffage et vis-versa. Par ailleurs, le prix marginal de l'énergie est souvent défini par tranche de consommation comme c'est le cas avec la tarification utilisée par Hydro-Québec. Le ménage est appelé à choisir une tranche de consommation donnée en même temps qu'il choisit la quantité d'énergie. Toute méthode d'estimation basée sur des prix marginaux non linaires et qui ne prend pas en compte la simultanéité entre ces trois choix produirait des estimateurs biaisés et non convergents. Les modèles à classes latentes semblent être un outils approprié pour résoudre un tel problème. Les tranches de consommation peuvent être considérées comme des classes non observables directement par le chercheur. Ces classes traduisent une hétérogénéité groupée au sein de l'échantillon. Un ménage donné appartiendra à une classe donnée avec une certaine probabilité. Nous prenons donc en compte les biais de sélection provenant du choix du mode de chauffage et le biais d'endogénéité dû au choix de la tranche de consommation. Les paramètres du modèle sont estimés par la méthode du maximum de vraisemblance.

2 Motivation

Nous sommes parfois confronté, dans le cadre de l'analyse de la demande, à des situations où les choix continus sont reliés aux choix discrets. Dans un tel context, toute estimation qui ignore cette simultanéité produirait des estimateurs biaisés et non convergents. Dubin et McFadden (1984) furent les premiers à prendre réellement en compte cette simultanéité dans les 2 choix en estimant un modèle de demande d'électricité et de choix des appareils de chauffage.
De plus, suite à la tarifcation par tranche, les ménages doivent choisir le mode de chauffage ainsi que la tranche de consommation d'énergie à laquelle ils appartiennent. Dans cette étude, nous voulons estimer un modèle de choix discret/continu dans un contexte de prix non linéaire. En effet, Hydro-Québec fixe un prix marginal d'électricité qui dépend de la tranche de consommation dans laquelle le ménage se situe. Il utilise un tarif croissant par tranche : un premier prix (p_1) lorsque la consom-

mation journalière ne dépasse pas les 30 kwh et un autre prix ($p_2 > p_1$) si elle dépasse ce niveau. Le prix (marginal) dépend donc de la demande, mais il constitue aussi une variable explicative de la fonction de demande. Il se pose alors un problème d'endogénéité du prix. Toute estimation du modèle de choix discret/continu (avec prix non linéaire) qui ignore ce problème produirait des estimateurs biaisés et non convergents, même si on corrigeait pour le problème de sélection associé au mode de chauffage. Un modèle à classe latente semble mieux approprié pour estimer la demande d'énergie tout en corrigeant le biais de sélection dû au mode et le biais d'endogénéité dû au choix de la tranche. La méthode du maximum de vraisemblance est utilisée pour estimer les paramètres du modèle.

3 brève revue des écrits antérieurs

3.1 Le problème de la tarification non linéaire

On sait depuis le papier d'Houthakker (1951) que les prix par tranche de consommation ont des implications économétriques importantes. Taylor (1975) soulignait le biais potentiel qui résulterait de l'estimation des élasticités de la demande lorsque la structure des prix est non linéaire. Selon Taylor (1975), la non linéarité de la contrainte budgétaire dont fait face le consommateur a des conséquences sur la fonction de demande, les courbes d'Engel et l'équilibre du consommateur.

Plusieurs auteurs ont cherché à résoudre le biais provenant de la non linéarité des prix. On peut citer pour la demande d'électricité : Acton, Mitchell et Mowill (1976), Barnes, Gillingham et Hageman (1981), Dubin (1985a et b), McFadden, Puig et Kirshner (1977), Taylor (1975), Herrigues et al. (1994), Reiss et White (2005) ; pour l'eau, on a : Billings et Agthe (1980, 1981), Foster et Beattie (1985a,b) ; pour le gaz naturel : Barnes, Gillingham et Hageman (1982), Polzin (1984).

Dans les écrits antérieurs, il existe une variété de méthodes d'estimation des modèles de demande avec endogénéité des prix. Herriges et al. (1994) en donnent une synthèse. Lorsqu'il y a un problème de variable explicative endogène, les estimateurs des moindres carrés ordinaires sont biaisés vers le haut. L'estimation par la forme réduite est une alternative. Elle consiste à utiliser un sous-ensemble du tarif de prix ou une combinaison d'éléments du tarif dans la forme réduite du modèle de demande. Taylor (1975) suggère d'utiliser à la fois les prix marginal et moyen. McFadden (1977) utilise 3 prix différents pour capter tous les effets de la tarification. Mais, l'estimation par la forme réduite a deux limites importantes :

- un problème d'identification des paramètres peut se poser lorsqu'il n'y a pas assez de variabilité dans les données (le seuil de changement de prix est constant comme c'est le cas pour les données qui nous concernent)

– la forme réduite est souvent ad hoc ce qui donne peu de justification quant aux variables à inclure dans le modèle.

L'estimation par variables instrumentales permet aussi de résoudre le biais dû à l'endogénéité du prix. La démarche consiste à trouver un instrument pour le prix moyen dans la fonction de la demande. Cette méthode a été utilisée par : Barnes, Gillingham et Hagemann (1981), Hausman, Kinnucan et McFadden (1979), Hausman et Trimble (1984). Cette méthode est supérieure à la forme réduite parce qu'elle permet de résoudre non seulement le biais d'endogénéité du prix, mais aussi elle permet de spécifier l'équation de demande en cohérence avec la théorie néoclassique de maximisation de l'utilité. Cependant, les estimateurs de la méthode à variables instrumentales ne sont pas les meilleurs estmateurs.

Une autre méthode est celle du maximum de vraisemblance structurelle initialement proposée par Burtless et Hausman (1978). Elle fut utilisée en autres par Dubin (1985b) pour estimer la demande d'électricité. Elle consiste à résoudre le problème de maximisation de l'utilité du consommateur sous sa contrainte budgétaire non linéaire. Sa fonction d'utilité directe contient un terme aléatoire qui traduirait soit de l'hétérogénéité aléatoire dans les préférences et/ou des erreurs de mesures dans les variables explicatives. La solution au problème de maximisation est une fonction de demande définie de façon spécifique pour chaque bloc. La loi du terme d'erreur permet de construire la fonction de vraisemblance à maximiser. Cependant, il se peut que les probabilités qui entrent dans la fonction de vraisemblance soient négatives. Alors, Herriges et al. (1994) proposent la méthode du maximum de vraisemblance structurelle modifiée pour résoudre ce problème. Leur méthode consiste à contrôler les probabilités de sorte à éviter des valeurs négatives. Un problème avec cette méthode est que la vraisemblance peut être assez complexe à maximiser, alors, Reiss et White (2005) proposent une approche basée sur la méthode des moments généralisés qui utilisent les deux premiers moments de la différence entre la consommation observée et la consommation espérée.

3.2 Le problème de la simultanéité entre choix discret et choix continu

Lorsqu'on a un modèle de choix discret/continu, estimer le modèle continu sans prendre en compte le choix discret produit des estimations biaisées et non convergentes. King (1980) a soulevé le problème et a tenté de le résoudre, mais c'est Dubin et McFadden (1984) qui furent les premiers à proposer vraiment une solution à ce problème. Leur démarche se résume en 2 grandes étapes. La première étape consiste à estimer le modèle de choix et à constituer ensuite les correcteurs du biais de sélection. Dans la deuxième étape, ils estiment la fonction de demande mais il s'agit d'une demande conditionnelle à l'alternative choisie. Ils appliquent ensuite les procédures classiques d'estimation : méthode des variables instrumentales, forme réduite. Ils font les comparent ensuite avec les estimateurs par moindres carrés ordinaires. Leurs estimateurs convergents mais moins efficaces que ceux du maximum de vraisemblance.

Haneman (1984) estime également un modèle de choix discret/continu (choix de la marque d'un produit et choix de la quantité de ce produit) en suivant la même démarche que Dubin et McFadden (1984). Il propose également l'estimation par le maximum de vraisemblance qui donne des estimateurs plus efficaces que ceux en deux étapes.

Bernard, Bolduc et Bélanger (1996) estiment un modèle de choix discret/continu appliquée à la consommation d'électricité des ménages Québécois. Le modèle est estimés en deux étapes. Ils utilisent un logit mixte pour le modèle discret afin de corriger le biais de sélection.

Sanga (1999) estime un modèle de choix discret/continu par la méthode du maximum de vraisemblance simulé et par la méthode de Heckman en deux étapes. Même si l'estimation par le maximum de vraisemblance simulé est plus efficace, elle est plus complexe et prend plus de temps de calcul que l'estimation en deux étapes.

Vaage (2000) utilise l'approche en deux étapes pour estimer la demande d'électricité des ménages Norvégiens. Nesbakken (2001) utilise la méthode du maximum de vraisemblance pour estimer la consommation d'énergie pour le chauffage des ménages norvégiens. Elle estime simultanément les paramètres du modèle de choix discret et ceux du modèle continu.

Comme nous pouvons le constater, il n'existe pas encore de travaux qui ait pris en compte à la fois le problème de la simultanéité entre choix discret et choix continu et celui de l'endogénéité du prix. La structure des données dont nous disposons implique qui si nous voulons estimer la demande d'électricité des ménages avec la struture tarifaire d'Hydro-Québec, nous devons résoudre simultanément ces deux problèmes pour obtenir des estimateurs convergents. Notre modèle serait ainsi beaucoup plus adapté à la réalité.

4 Modèle proposé : Modèle à classe latente

4.0.1 Le modèle de choix du mode de chauffage : le logit multinomial

Nous proposons un modèle logit multinomial dans un premier temps. Compte tenu de la corrélation potentielle qui existe entre les alternatives, nous comptons utiliser plus tard un modèle logit mixte.

Pour qu'un ménage choisisse un mode de chauffage donné, il faut que ce mode lui rapporte la plus grande utilité comparée aux autres modes. Nous supposons une fonction d'utilité indirecte qui satisfasse les propriétés classiques de maximisation de l'utilité. Après avoir choisi un mode de chauffage (ou portefeuille d'appareils), le ménage les utilise pour produire des biens ou services de consommation dont le processus de production nécessite de l'énergie.

Supposons l'hypothèse de maximisation de l'utilité que nous notons par U_{jn}. Étant donné le mode de chauffage choisi, le ménage a une fonction d'utilité indirecte conditionnelle de la forme :

$$U\left(j, rev - r_j, X, Pe, Pg, Pm, Pbie, \epsilon_j, \eta\right) \tag{1}$$

où $Pe, Pg, Pm, Pbie$ sont les prix de l'électricité, du gaz naturel, du mazout et de la biénergie respectivement ; ϵ_j est l'ensemble des caractéristiques non observables du mode de chauffage, η est l'ensemble des caractéristiques non observables du ménage, r_j est le coût total du mode j et rev le revenu du ménage, X est un vecteur de variables exogènes.

En utilisant l'identité de Roy, la demande d'électricité pour les gens qui sont tout à l'électricité est donnée par :

$$y_{j=7} = \frac{\partial U\left(j, rev - r_j, X, Pe, Pg, Pm, Pbie, \epsilon_j, \eta\right)/\partial P_e}{\partial U\left(j, rev - r_j, X, Pe, Pg, Pm, Pbie, \epsilon_j, \eta\right)/\partial rev} \tag{2}$$

Le ménage choisit l'alternative j (le mode de chauffage) si et seulement si : $U_{jn} > U_{in} \quad \forall i \neq j$. Définissons la variable δ_{jn} comme suit :

$$\delta_{jn}(\epsilon) = \begin{cases} 1 & si \quad U_{jn} > U_{in} \quad \forall i \neq j \\ 0 & sinon \end{cases}$$

La fonction d'utilité U_{jn} se décompose en 2 parties : une déterministe V_{jn} et l'autre aléatoire ε_{jn} :

$$U_{jn} = V_{jn} + \varepsilon_{jn}$$

La probabilité que le ménage n choisisse l'alternative j est :

$$\begin{aligned} P_n\left(j\right) &= P\left(U_{jn} > U_{in} \quad \forall \quad i \neq j\right) \\ &= P\left(V_{jn} + \varepsilon_{jn} > V_{in} + \varepsilon_{in} \qquad \forall \quad i \neq j\right) \\ & \quad P\left(\varepsilon_{in} - \varepsilon_{jn} < V_{in} - V_{jn} \qquad \forall \quad i \neq j\right) \end{aligned} \tag{3}$$

Supposons le que le terme d'erreur de la fonction d'utilité indrecte suit une distribution $i.i.d$ Gumbel.

$$P_n\left(j\right) = \frac{\exp\left(V_{jn}\right)}{\sum_{i \in J} \exp\left(V_{in}\right)}$$

La partie déterministe est définie comme suit :

$$V_{jn} = X_{jn}\beta + \kappa\left(rev_n - r_j\right)$$

où X est un vecteur de toutes les variables exogènes sauf le revenu net des coûts de l'alternative choisie.

4.0.2 La fonction de demande d'énergie

Après avoir choisi le mode de chauffage, le ménage est appelé à choisir la quantité d'énergie nécessaire pour satisfaire ses besoins. Il choisit en même temps la tranche de consommation à laquelle il appartient.

La fonction de demande dérivée à partir de l'identité de Roy peut s'écrire de la façon suivante :

$$\begin{aligned} y_{jn} &= X_{jn}\beta + Q_{jn}\lambda + \kappa\left(rev_n - r_j\right) + \eta_{jn} \\ \eta_{jn} &\sim N\left(0, \sigma_\eta^2\right) \end{aligned}$$

où Q est un vecteur de variables exogènes spécifiques à la demande.

Le modèle sous forme matricelle est :

$$y = W\theta + \eta \qquad \eta \sim N\left(0, \Sigma_\eta\right)$$

où :

$$\begin{aligned} W &= [X, Q, (rev - r)] \\ \theta &= [\beta, \lambda, \kappa] \end{aligned}$$

La variable X est de dimension $(NJ \times K_1)$, le vecteur β est de dimension $K_1 \times 1$, la variable Q est de dimension $(NJ \times K_2)$, la variable $(rev_n - r)$ est de dimension $(NJ \times 1)$, le paramètre λ est de dimension $K_2 \times 1$, le paramètre κ est un scalaire. W de dimension $NJ \times K$ où $K = K_1 + K_2 + 1$, θ de dimension $K \times 1$ et enfin le vecteur y est de dimension $(NJ \times 1)$.

Comme souligné auparavant, Hydro-Québec a une structure de tarification croissante par tranche de consommation. Il existe deux tranches et donc deux prix différents. La consommation de la deuxième tranche est plus chère que celle de la première tranche. Définissons la variable $y^* = consommation \quad journalière - 30kwh$

Si $y^* \leq 0$, le ménage se situe dans la première tranche. Si $y^* > 0$, le ménage se situe dans la deuxième tranche. Cependant, le ménage peut ne pas savoir qu'il est dans la première ou la deuxième tranche ; la consommation journalière n'est pas directement observable. Ceci nous ramène à introduire la notion de classe latente. Les deux classes ne sont pas directement observables.

Les modèles à classes latentes sont souvent utilisés pour prendre en compte l'hétérogénéité qui existe entre différents groupes d'observations. La notion de classe latente fut introduite en 1950 par Lazarsheld pour un modèle dichotomique. 20 ans plus tard, Goodman (1974) rend son modèle applicable en développant un algorithme pour obtenir les estimateurs du maximum de vraisemblance des paramètres du modèle. Beaucoup d'autres auteurs se sont intéressés à ce type de modèle et il existe des logiciels appropriés pour son estimation. Les modèles à classe latente peuvent s'appliquer à des variables discrètes ou continues. Ces modèles sont assez proches du modèle logit mixte qui prend en compte l'hétérogénéité entre les observations. Nous nous inspirons des travaux de Gopinath (1995) et Vermunt et Magidson (2000).

Nous supposons que le ménage sera dans une des deux tranches avec une certaine probabilité. Notons $s = 1, 2$ les classes possibles d'appartenance d'un ménage. Ainsi, la probabilité que le ménage appartienne à la première tranche s'écrit :

$$P\left(y^* \leq 0\right) = \pi_{s=1}$$

La probabilité que le ménage appartienne à la deuxième classe est :

$$P\left(y^* > 0\right) = \pi_{s=2} = 1 - \pi_{s=1}$$

Supposons que y^* est fonction d'un certain nombre de variables explicatives permettant de connaître la classe du ménage :

$$\begin{aligned}
y^* &= M'\alpha + u \qquad u \sim N\left(0, \sigma_u^2\right) \\
P\left(y^* \leq 0\right) &= \pi_{s=1} = P\left(\frac{y^* - M'\alpha}{\sigma_u} \leq -\frac{M'\alpha}{\sigma_u}\right) \\
&= \Phi\left(-\frac{M'\alpha}{\sigma_u}\right) = 1 - \Phi\left(\frac{M'\alpha}{\sigma_u}\right) \\
P\left(y^* > 0\right) &= \pi_{s=2} = 1 - P\left(y^* \leq 0\right) \\
&= \Phi\left(\frac{M'\alpha}{\sigma_u}\right)
\end{aligned}$$

Nous fixons $\sigma_u = 1$ pour que le modèle soit identifiable.

4.0.3 Estimation par le maximum de vraisemblance

Nous voulons estimer le modèle de demande par la méthode du maximum de vraisemblance. Écrivons la fonction de densité conjointe associée au choix de la quantité d'énergie, de la classe et de l'alternative :

$$\begin{aligned}
f(y_{jn}, j, s, W; \theta) &= f(y_{jn} | j, s, W ; \theta) \, g(j, s) \\
&= f(y_{jn} | j, s, W ; \theta) \, P(j) \, \pi_s
\end{aligned}$$

où : π_s est la probabilité que le ménage appartienne à la classe s, $P(j)$ est la probabilité que le ménage choisisse l'alternative j, et enfin $f(y_j | j, s, W ; \theta)$ est la quantité d'énergie demandée conditionnellement à la classe et à l'alternative choisies.

Étant donné qu'il s'agit d'un modèle à classe latente, la fonction de vraisemblance s'écrit :

$$\begin{aligned}
L(y, W, \theta) &= \prod_{n=1}^{N} \prod_{j=1}^{J} \left[\sum_{s=1}^{S} f(y_{jn}, j, s, W; \theta) \right]^{\delta_{jn}} \\
&= \prod_{n=1}^{N} \prod_{j=1}^{J} \left[\sum_{s=1}^{S} f(y_{jn} | j, s, W ; \theta) \, P(j) \, \pi_s \right]^{\delta_{jn}} \\
&= \prod_{n=1}^{N} \prod_{j=1}^{J} \begin{pmatrix} (f(y_{jn} | j, s=1, W ; \theta_{s=1}) \, P(j) \, \pi_{s=1})^{\delta_{jn}} + \\ (f(y_{jn} | j, s=2, W ; \theta_{s=2}) \, P(j) \, \pi_{s=2})^{\delta_{jn}} \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
\ln L(y, W, \theta) &= \ln \begin{pmatrix} \prod_{n=1}^{N} \prod_{j=1}^{J} (f(y_{jn} | j, s=1, W ; \theta_s) \, P(j) \, \pi_{s=1})^{\delta_{jn}} + \\ (f(y_{jn} | j, s=2, W ; \theta_s) \, P(j) \, \pi_{s=2})^{\delta_{jn}} \end{pmatrix} \\
&= \ln \begin{pmatrix} \prod_{n=1}^{N} \prod_{j=1}^{J} \left(f(y_{jn} | j, s=1, W ; \theta_s) \, P(j) \times \left(1 - \Phi\left(\frac{M'\alpha}{\sigma_u}\right)\right)\right)^{\delta_{jn}} + \\ \left(f(y_{jn} | j, s=2, W ; \theta_s) \, P(j) \times \Phi\left(\frac{M'\alpha}{\sigma_u}\right)\right)^{\delta_{jn}} \end{pmatrix}
\end{aligned}$$

Nous déterminons ensuite les densités marginales (normale tronquées) $f(y_{jn} | j, s=1, W ; \theta)$ et $f(y_{jn} | j, s=2, W ; \theta)$.

$$\begin{aligned}
f(y_{jn} | j, s=1, W ; \theta_s) &= f(y_{jn} | y^* \leq 0, j, W ; \theta_1) \\
f(y_{jn} | j, s=2, W ; \theta_s) &= f(y_{jn} | y^* > 0, j, W ; \theta_2)
\end{aligned}$$

Simplifions cette écriture en omettant les autres éléments (j, W, θ) pour avoir des formes plus faciles à manipuler :

$$\begin{aligned}
f(y_{jn} | j, s=1, W ; \theta_s) &= f(y_{jn} | y^* \leq 0) \\
f(y_{jn} | j, s=2, W ; \theta_s) &= f(y_{jn} | y^* > 0)
\end{aligned}$$

Nous supposons que les fonctions de demande conditionnelle sont de la forme :

$$\begin{aligned}
y_{j1} &= W_j'\theta_1 + \eta_{j1} \\
y_{j2} &= W_j'\theta_2 + \eta_{j2} \\
\begin{pmatrix} \eta_{j1} \\ \eta_{j2} \end{pmatrix} &\sim N\left(0, \Sigma_\eta\right) \\
\Sigma_\eta &= \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \\
\Sigma_{\eta u} &= \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho_1\sigma_1\sigma_u \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho_2\sigma_1\sigma_u \\ \rho_1\sigma_1\sigma_u & \rho_2\sigma_1\sigma_u & \sigma_u^2 \end{pmatrix}
\end{aligned}$$

Les moments conditionnels (espérance et variance conditionnelles) sont définis ci-dessous :

$$\begin{aligned}
E\left(y_{j1} \left| y^* \leq 0\right.\right) &= W_j'\theta_1 - \rho_1\sigma_1 \frac{\phi\left(-\frac{M'\alpha}{\sigma_u}\right)}{\Phi\left(-\frac{M'\alpha}{\sigma_u}\right)} \\
&= W_j'\theta_1 - \rho_1\sigma_1 \frac{\phi\left(\frac{M'\alpha}{\sigma_u}\right)}{1-\Phi\left(\frac{M'\alpha}{\sigma_u}\right)}
\end{aligned}$$

$$\begin{aligned}
V\left(y_{j1} \left| y^* \leq 0\right.\right) &= \sigma_1^2 - \rho_1^2\sigma_1^2 \left[-\frac{M'\alpha}{\sigma_u} \times \frac{\phi\left(-\frac{M'\alpha}{\sigma_u}\right)}{\Phi\left(-\frac{M'\alpha}{\sigma_u}\right)} + \left[\frac{\phi\left(-\frac{M'\alpha}{\sigma_u}\right)}{\Phi\left(-\frac{M'\alpha}{\sigma_u}\right)}\right]^2 \right] \\
&= \sigma_1^2 \left[1 - \rho_1^2 \left(-\frac{M'\alpha}{\sigma_u} \times \frac{\phi\left(\frac{M'\alpha}{\sigma_u}\right)}{1-\Phi\left(\frac{M'\alpha}{\sigma_u}\right)} + \left[\frac{\phi\left(\frac{M'\alpha}{\sigma_u}\right)}{1-\Phi\left(\frac{M'\alpha}{\sigma_u}\right)}\right] \right) \right]^2 \Bigg]
\end{aligned}$$

Pour la deuxième classe, nous aurons :

$$\begin{aligned}
E\left(y_{j2} \left| y^* > 0\right.\right) &= W_j'\theta_2 + \rho_2\sigma_2 \frac{\phi\left(\frac{M'\alpha}{\sigma_u}\right)}{\Phi\left(\frac{M'\alpha}{\sigma_u}\right)} \\
V\left(y_{j1} \left| y^* > 0\right.\right) &= \sigma_1^2 \left[1 - \rho_2^2 \left(\frac{M'\alpha}{\sigma_u} \times \frac{\phi\left(\frac{M'\alpha}{\sigma_u}\right)}{\Phi\left(-\frac{M'\alpha}{\sigma_u}\right)} + \left[\frac{\phi\left(\frac{M'\alpha}{\sigma_u}\right)}{\Phi\left(\frac{M'\alpha}{\sigma_u}\right)}\right]^2 \right) \right]
\end{aligned}$$

Après avoir déterminé la loi conjointe de

$$\begin{pmatrix} y_{j1} & |y^* \\ y_{j2} & \end{pmatrix}$$

qui est une normale tronquée. Il reste alors à remplacer ces fonctions de densités conditionnelles par leurs expressions dans la fonction log-vraisemblance.

5 Les données

Nous utilisons les données provenant d'une enquête postale ménée par Hydro-Québec en 1989 portant sur la consommation d'énergie résidentielle. Le taux de réponse au questionnaire était de 44.9 pourcent avec 45833 ménages ayant répondu sur un total de 101977 questionnaires expédiés. Le sous-échantillon de 3090 observations est constitué de ménages ayant une maison unifamiliale (détachée, semi-détachée ou en rang avec entrées séparées). Les tableaux ci-dessous présentent les principales informations disponibles dans cette base de données.

Quelques variables

Variables	Nom	Moyenne échantillonnale
Options (Locaux-eau)		
Gaz/gaz	1	0.010
Gaz/électricité	2	0.004
Bi-énergie/mazout	3	0.024
Bi-énergie/électricité	4	0.072
Mazout/mazout	5	0.005
Mazout/électricité	6	0.007
Électricité/Électricité	7	0.801
Bois/électricité	8	0.050
Bois-électricité/électricité	9	0.028
Secteur	sect	
Rural		0.254
Peu urbain		0.103
Urbain		0.031
Haute densité		0.612
Degrés-jours de chauffage	hdd	4844.0
Année de conversion (1986-1989)	datconv	1987.1
Année de construction (1920-1989)	datcon	1977.4
Nombre de personnes par ménage	nbre_residents	3.09
Surface (pieds carrés)	surf	1656.5
Age du chef de famille	age	42.1
Revenu du ménage ($)	rev	42536.4
Coût d'opération annuel ($)	cout_operation	1394.9
Coût fixe annuel ($)	cout_fixe	392.2
Coût fixe annuels revenu ($10^3 \times \$^2$)	cfixy	1744.9
Consommation annuelle (kwh)		23136.35

source : Bernard, Bolduc et Bélanger (1996)

Répartition des observations par tranche

Mode	**Tranche 1**	**Tranche 2**	**total**	**Nom du mode**
1	20 (62, 5%)	12	32	**gaz/gaz**
2	7 (53, 85%)	6	13	**gaz/électricité**
3	8 (10, 96%)	65	73	**biénergie/mazout**
4	14 (6, 33%)	207	221	**biénergie/électricité**
5	10 (62, 5%)	6	16	**mazout/mazout**
6	6 (28, 57%)	15	21	**mazout/électricité**
7	161 (6.5%)	2313	2474	**électricité/électricité**
8	39 (25, 5%)	114	153	**bois/électricité**
9	12 (13, 8%)	75	87	**bois-électricité/électricité**
Total	277 (8, 96%)	2813	3090	
Prix maginal électricité	3.76	4.46		

6 Résultats empiriques

Références

[1] Acton, J. P., Mitchell, B., M., and Mowill, R., S. (1976) : *Residential Demand for Electricity in Los Angeles : An Econometric Study of Disaggregate Data,* Rand-Report R-1899-NSF, Rand Corporation, Santa Monica, CA.

[2] Barnes, R., Gillingham, R., and Hagemann, R. (1981) : *The Short-Run Residential Demand for Electricity,* The Review of Economics and Statistics, Vol. 62, p. 541-551.

[3] Burtless, G., and Hausman, J., A. (1978) : *The Effect of Taxation of Labor Supply : Evaluating the Gray Neagtive Income Taxe Experiment*, Journal of Political Economy, 86, 276-278.

[4] Amemiya, T. (1984) *Tobit models : a surveys*, Journal of Econometrics, Vol.24 , p. 3-61.

[5] Amemiya, T. (1985) *Advanced econometrics*, Havard Univertsity Press, Cambridge 1985

[6] Bernard, J-T, Bolduc, D. and Belanger, D. (1996), *Quebec Residential Electricity Demand : A Microeconometric Aproach.* The Canadian Journal of Economics, vol. 29, No.1 , p. 92-113.

[7] Billings, R., B. (1982) : *Specification of Bolck Rate Price Variable in Demand Models,* Land Economics, Vol. 56, p. 73-84.

[8] Dubin, J. A. (1985a) : *Consumer Durable Choice and the Demand for Electricity,* Amsterdam : North-Holland.

[9] Dubin, J. A. (1985b) : *Evidence of Block Switching in Demand Subject to Decling Block Rates-A New Approach*, Social Science Working Paper 583, California Institute of Technology, Dept. of Economics.

[10] Dublin, J., J. and McFadden, D. (1984) *An Econometric Analysis of Residential Electric Apliance Holdings and Consumption*, Econometrica, Vol. 52, No. 2, p. 345-362.

[11] Foster, H., S., Jr., and Beattie, B., R. (1981a) : *Urban Residential Demand for Water in the United Stated : Reply,* Land Economics, 57, 257-265.

[12] Foster, H., S., Jr., and Beattie, B., R. (1981b) : *On the Specification of Price in Studies of Consumer Demand Under Block Price Scheduling,* Land Economics, 57, 624-629.

[13] Gopinath D. A. (1995) : *Modeling Heterogeneity in Discrete Choice Processes : Application to Travel Demand* PhD MIT

[14] Greene W. H. and Hensher D. (2002) : *A Latent Class Model for Discrete Choice Analysis : Contrasts with Mixed Logit* Working Paper, Institute of Tansport Studies

[15] Hanemann, W. M. (1984), *Discrete/Continuous Models of Consumer Demand*, Econometrica, Vol. 52, No. 3, p. 541-562

[16] Hausman, J.A. (1979) *Individual Discount Rates and the Purchase and Utilisation of Energy-Using Durables*, The Bell Journal of Economics, Vol. 10, No. 1, p. 33-54.

[17] Hausman, J.A. (1985) *The Econometrics of Nonlinear Budget Sets*, Econometrica, Vol. 53, No. 6, p. 1225-1282

[18] Hausman, J.A., Kinnucan, M., and McFadden, D. (1979) *A Two-Level Electricity Demand Model*, Journal of Econometrics, Vol. 10, p. 263-289.

[19] Heckman, J. J (1978),*Dummy endogenous variables in a simultaneous equations system*, Econometric, Vol 46, No.4, p. 931-959

[20] Heckman, J. J (1979), *Sample selection bias as a specification error,* Econometrica, Vol. 47, No.1, p. 153-162

[21] Herriges, J., A. and King, K., K. (1994), *Residential Demand for Electricity under Inverted Block Rates : Evidence from a Controlled Experiment*, Journal of Business and Economic Statistics, Vol. 12, No. 4, p. 419-430.

[22] Lazarsfeld, P. F. (1950) : *The Logical and Mathematical Foundation of Latent Structure Analysis & The Interpretation and Mathematical Foundation of Latent Structure Analysis.* S. A. Stouffer et al. (eds.) *Measurement and Prediction* 362-472. Princeton, NJ : Princeton University Press.

[23] Larsen, B., M. and Nesbakken, R.(2004) *Household Electricity End-use Consumption : Results from Econometrics and Engineering Models*, Energy Economics, Vol.26, p. 179-200

[24] Lee, L. F.(1978), *Unionism and Wage Rates : A Simultaneous Equations Models with Qualitative and Limited Dependent Variables.*, International Economic Review, Vol. 19, p. 415-433.

[25] McFadden, D., Puig, C., and Kirshner, D. (1977), *Determinants of the Long-Run Demand for Electricity*, In the Proceeding of the Business and Economic Statistics Section, American Statistical Association, p. 109-119.

[26] Olsen, R.J. *Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model*. Econometrica, Vol. 46, No.5, p. 1211-1215

[27] Polzin, P. E. (1984) : *The Specification of Price in Studies of Consumer Demand Under Block Price Scheduling,*in the Proceedings of the Business and Economic Statistics Section, American Statictical Association, p. 109-119.

[28] Reiss, P. and White M.W. (2005) *Household Electricity Demand Revisited.* Review of Economic Studies, Vol. 72, p. 853-883.

[29] Nesbakken, R.(2001) *Energy Consumption for Space Heating : a Discrete-Continuous Aproach.*, Scandinavian Journal of Economics, Vol.103, No.1, p. 165-184

[30] Sanga, D. M. (1999), *Estimation des Modèles Économétrique de Choix Discrets/Continus avec Choix Polytomiques Interdépendants : une Aproche par Simulation.* Thèse (Ph.D), Université Laval

[31] Taylor, L., D., (1975), *The Demand for Electricity : A Survey*, The Bell Journal of Economics, Vol. 6, No. 1, p. 74-110.

[32] Train, K. (1993) *Continuous/Discrete Models. Qualitative Choice Analysis : Theory, Econometrics, and an Aplication Aplication to Automobile Demand.* Cambridge : The MIT Press

[33] Vaage K. (2000) : *Heating Technology and Energy Use : a Discret/Continuous Choice Aproach to Norwegian Household Energy Demand.* Energy Economics, Vol.22, No.6, p. 649-666

[34] F. Vekeman, F., Bolduc D. and Bernard J-T (2004) *Household Vehicles Choice and Use : What a more Disaggregated Aproach Aproach Reveals ?* Working Paper, (GREEN), Department of Economics, Université Laval.

[35] Vermunt J. K. and Magidson J. (2000) : *Latent Class Cluster Analysis* Statistical Innovations Inc. Tiburg University