

Intention de fidélité et hétérogénéité du comportement des touristes visitant la Tunisie

Moez JEDIDI

Assistant Contractuel ESSEC Tunis
Rue des Jasmins Hammamet 8050
Tél : 98 356 534
jedidime@yahoo.fr

&

Mohamed Ayadi

Professeur ISG Tunis
45, Rue de la liberté Bardo
Tél : 71 261 854
ayadimed2001@yahoo.fr

Résumé :

Nous analysons la relation entre l'intention de fidélité et le degré de satisfaction des consommateurs que nous illustrons grâce à l'analyse des déterminants de l'intention de retour des touristes visitant la Tunisie.

Nous utilisons un modèle à structure latente généralisée appliquée au modèle Probit avec auto – sélection, ce qui nous a permis de traiter le problème de l'hétérogénéité des consommateurs qui sera considéré comme une variable endogène.

Mots Clé : Hétérogénéité, inobservabilité des frontières, Intention de Fidélité, Satisfaction, Modèle Probit avec auto –sélection, structure latente généralisée, segmentation endogène.

Abstract:

We are analysing the relation between the loyalty intension and the consumer's degree of satisfaction which indicates the tourist intension of coming back to Tunisia.

We are adopting a model of a generalized latent structure applied to a sample –selection Probit model. This allows us to treat the consumer's problem of heterogeneity which will be considered as an endogen variable.

Keywords: Heterogeneity, not observable boundary, Loyalty Intension, satisfaction, a sample –selection Probit model, generalized latent structure, endogeneous segmentation.

INTRODUCTION :

Les comportements des consommateurs sont assez imprévisibles puisqu'ils sont assez complexes et surtout très hétérogènes. Afin d'éliminer les biais et conserver des propriétés économétriques adéquates à ces estimateurs toute étude comportementale doit tenir compte du problème de l'hétérogénéité caractérisant son échantillon.

L'une des analyses de comportement les plus délicates concerne l'analyse de la relation entre le degré de satisfaction des consommateurs et leur intention de fidélité. Ce type d'analyse est d'autant plus important s'il touche un secteur neurologique de l'économie à savoir le tourisme. En effet, le secteur hôtelier a été pris comme domaine d'application de notre étude. Le choix de ce secteur de service n'a pas été le fruit du hasard. En réalité, le secteur hôtelier a acquis au cours de ces dernières années, un poids prépondérant dans l'économie tunisienne.

En effet, l'hôtellerie joue un rôle d'animateur dans un pays tel que la Tunisie, pays en plein effort de développement économique et social. Il s'agit d'une industrie clé dont la prospérité se répercute sur toutes les branches de l'activité nationale. Cette prospérité est largement tributaire de la qualité des services offerts d'autant plus que l'hôtellerie tunisienne subit une concurrence mondiale et surtout méditerranéenne.

La résolution du problème de l'hétérogénéité des échantillons exige une analyse assez poussée des caractéristiques de la population ciblée. Une première étape de cette analyse passe obligatoirement par la segmentation du marché étudié. Dans notre article, nous allons essayer de dissocier les clients fidèles, des clients opportunistes qui ne reviennent en Tunisie que sous certaines conditions. En nous basant sur les deux articles pionniers de Jedidi et alii (1993), Desarbo et Choi (1999), nous montrons qu'il existe deux types de consommateurs distincts, ce qui nécessiterait un traitement particulier de l'hétérogénéité. Toutefois, le point clé reste l'inobservabilité des frontières qui séparent ces deux sous ensembles de consommateurs. Nous pouvons « endogénéiser » cette hétérogénéité grâce à l'usage de l'algorithme E – M. Nous procéderons entre autre à un ensemble de tests asymptotique pour comparer les différentes spécifications retenues et surtout pour juger de la segmentation adéquate des touristes. Nous retenons ainsi un modèle Probit avec auto – sélection et endogénéisation de l'hétérogénéité. Nous montrerons qu'un ensemble de biais sera éliminé, ce qui a donné lieu à une analyse plus conforme avec la réalité du secteur étudié.

Dans cette étude nous débuterons par une présentation des fondements théorique de la relation entre la satisfaction et l'intention de fidélité. Nous passerons, par la suite, à l'étude empirique de l'impact du degré de satisfaction des consommateurs sur leur intention de fidélité. Puis, nous présenterons notre base de données et notre spécification économétrique pour l'analyse et ainsi que les tests de segmentation adéquate. Dans la dernière partie de notre article nous présenterons, les résultats des différents modèles retenus.

LA RELATION ENTRE L'INTENTION DE FIDELITE ET LE DEGRE DE SATISFACTION :

La fidélité est depuis plusieurs années un domaine de réflexion important pour les universitaires. Progressivement, au cours des années 80- 90, elle est devenue une préoccupation majeure pour les entreprises sous forme d'abord de programme de fidélisation puis de stratégie de fidélisation de plus en plus élaborée (démarche rigoureuse, techniques variées).

La littérature sur la fidélité des individus aux marques est assez abondante; elle a abouti depuis plusieurs années à une définition et une décomposition de la fidélité en deux dimensions : comportementale et attitudinale.

En effet, Laban (1979) précédé par Jacoby et Chestnut (1978) ont réalisé une synthèse de ces définitions. Dans les deux cas les auteurs ont dénombré trois types de définitions de la fidélité :

- Les définitions comportementales comme celle de Lawrence (1969) qui dit :
"un client sera qualifié de fidèle s'il achète trois ou quatre fois de suite la même marque. "
- Les définitions attitudinales qui s'expriment à travers le développement d'une attitude favorable envers le produit.
- Les définitions composites mêlant attitudes et comportements sont déjà à cette époque considérées comme les plus réalistes et les plus complètes pour définir la fidélité des clients.

La grande majorité des définitions données dans les deux dernières décennies s'est ensuite reliée aux approches composites en définissant la fidélité par des attitudes et des comportements. Ces définitions ont parallèlement insisté sur le fait que l'attitude de fidélité est un déterminant du comportement de fidélité. Nous admettons ainsi d'après Trinqucoste (1996) que :

" La fidélité s'exprime par des comportements de consommation et s'explique par les attitudes favorables des consommateurs."

La fidélité du client ainsi définie permet d'augmenter le chiffre d'affaires de toute entreprise et ceci soit par la communication interpersonnelle faite par les clients fidèles ou par le renforcement de l'image de l'entreprise ou encore par les informations qu'ils procurent à la firme.

De ce fait, l'étude des facteurs qui influencent la fidélité est devenue primordiale; l'un des plus importants entre ces facteurs est la satisfaction.

En effet, plusieurs chercheurs ont étudié les relations entre la satisfaction et la fidélisation dans ses deux formes (comportementale et attitudinale).

Fornell et alii (1996) ont montré que la satisfaction cumulée est un antécédent majeur de la fidélité. Westbrook (1987), Bitner (1990), Reicheld et Sasser (1990), ont montré l'influence des communications formelles et informelles dues au niveau de satisfaction sur la fidélisation. Yi (1990) explique ces communications comme étant le résultat de l'écart entre les attentes et la performance réelle du service. De leur côté, Anderson (1995), Zeithaml et alii (1996), ont montré que l'augmentation de la satisfaction entraîne une baisse de la sensibilité aux variations des prix, cause importante de l'infidélité. Oliver (1980), Cronin et Taylor (1992), Trinqucoste (1996) ont montré les effets positifs de la satisfaction sur la forme attitudinale de la fidélisation (les intentions de ré-achat); puisque les intentions impliquent davantage le sujet et sont plus prédictives de son comportement future. De même, McAlexander et alii (1994) affirment dans leur étude conduite dans le secteur médical que la satisfaction du patient et la qualité de service ont un effet significatif sur l'intention d'achat future. Getty et Thomson (1994) ont confirmé cette relation dans leur étude dans le secteur des logements.

Bearden et Teel (1983) ont essayé d'étendre l'étude faite par Oliver par l'incorporation du comportement de réclamations dans le modèle théorique. Ces chercheurs ont remarqué que les intentions de se procurer le service sont fortement dépendantes de la satisfaction et des attentes.

Dufer et Moulins (1989) ont étudié la relation entre la satisfaction du consommateur et sa fidélité à la marque des produits comme le café, le shampoing et la lessive. Ils ont montré clairement que si la satisfaction peut expliquer la dimension psychologique de la fidélité à la marque (intention d'achat), c'est un très mauvais prédicteur du comportement d'achat effectif.

McDougall et Levesque (2000) ont étudié aussi la relation entre la satisfaction et la fidélité à l'aide de quatre services : service dentaire, service d'entretien de voiture, restauration et coiffure. Dans cet article deux intentions futures ont été mesurées : intention de fidélité (intention de retour et intention de recommander le service à des amis) et intention de changement.

Concernant la relation entre l'intention et la satisfaction, les chercheurs ont remarqué que la satisfaction du client prédit fortement les intentions de fidélités et faiblement les intentions de changements.

MODELISATION DE LA RELATION ENTRE L'INTENTION DE FIDELITE ET LE DEGRE DE SATISFACTION :

L'étude de la relation entre la satisfaction et l'intention de fidélité a été expliquée à travers de nombreux modèles comme il a été présenté précédemment.

La validation empirique de cette relation a été faite par plusieurs outils d'analyse quantitatives parmi les quels : la méthode des chemins récursifs proposé par Oliver (1980) et Bearton et Teel (1983), l'analyse de la variance utilisé par Dufer et Moulins (1989) ou encore les équations structurelles appliqué par Bitner (1990) et McDougall et Levesque (2000) et aussi la régression MCO employé par Hallowell 1996. Dans notre étude nous avons adopté les modèles à choix discret.

Une première façon de poser le problème serait d'estimer deux relations : une première explique le degré de satisfaction et la deuxième l'intention de fidélité et ceux par deux modèles Probit indépendants. Cependant l'interdépendance entre le sentiment de satisfaction et l'attitude de fidélité nous pousse à utiliser un modèle Probit avec auto-sélection. Ce modèle se caractérise par le fait que le passage du modèle de satisfaction à celui de l'intention de fidélité n'est pas automatique, il est conditionné par le degré de satisfaction.

Le modèle Probit avec auto - sélection :

Dans la formalisation de la relation entre l'attitude de fidélité et le degré de satisfaction, nous utilisons les variables observables : satisfaction et fidélité (noté : y_i et S_i respectivement) comme indicateur (ou signal) de l'intention de fidélité et du degré de satisfaction.

Ces variables seront traitées comme des signaux. Nous supposons que ces signaux résultent d'un calcul implicite mesurant les niveaux de satisfaction et d'intention de fidélité effectifs non observables et représentés par les variables latentes y_i^* et S_i^* respectivement.

La formulation du modèle Probit avec auto - sélection peut être présenté comme suit :

$$\begin{aligned}
 & y_i^* = X_i' b + \epsilon_i \\
 & S_i^* = Z_i' a + \eta_i \\
 & \text{Avec :} \\
 & y_i = 1 \text{ si } y_i^* \geq 0 \\
 & y_i = 0 \text{ si } y_i^* < 0 \\
 & \text{Et} \\
 & S_i = 1 \text{ si } y_i^* \geq 0, S_i^* \geq 0 \\
 & S_i = 0 \text{ si } y_i^* < 0, S_i^* < 0
 \end{aligned} \tag{1}$$

Avec : y_i^* et S_i^* les deux variables latentes.

X_i' : (X_{i1}, \dots, X_{iH}) : le vecteur des H variables exogènes.

b : (b_1, \dots, b_H) : le vecteur contenant les coefficients de régression associées au modèle.

Z_i' : (Z_{i1}, \dots, Z_{iH}) : Le deuxième ensemble de variables exogènes.

a : (a_1, \dots, a_H) : Le deuxième vecteur des paramètres de régression.

Nous supposons que les erreurs ϵ_i et η_i suivent la loi normale multivariée d'espérance nulle et de variance unitaire et ayant un coefficient de corrélation ρ différent de zéro.

Les probabilités de passage dans la première étape sont :

$$\begin{aligned}
 & p(y_i = 1) = \Phi(X_i' b) \\
 & p(y_i = 0) = 1 - \Phi(X_i' b)
 \end{aligned} \tag{2}$$

Où Φ est la fonction de répartition de la loi normale standardisée.

La probabilité que nous passons à la deuxième étape est :

$$p(y_i^* \geq 0, S_i^* \geq 0) = \int \frac{X_i' b + \rho (S_i - Z_i' a)}{\sqrt{1 - \rho^2}} \phi\left(\frac{X_i' b + \rho (S_i - Z_i' a)}{\sqrt{1 - \rho^2}}\right) \phi\left(\frac{S_i - Z_i' a}{\sqrt{1 - \rho^2}}\right) dS_i \tag{3}$$

Où ϕ est la fonction de densité de la loi normale standardisée.

Cette probabilité est obtenue grâce à la loi conditionnelle d'une variable normale dont l'espérance et la variance sont :

$$\begin{aligned} E(y_i^* / S_i, S_i^*) &= \mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} (S_i^* - \mu_2) \\ V(y_i^* / S_i, S_i^*) &= \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \end{aligned} \quad (4)$$

Avec μ_1 et μ_2 les moyennes respectives de y et S. En se basant sur ces résultats la vraisemblance s'écrit finalement :

$$L = \prod_{S_i \neq 0} \left[1 - \frac{X_i b}{\sigma_1} \right] \prod_{S_i = 0} \left[\frac{X_i b + (\mu_1 / \mu_2)(S_i - Z_i a)}{\sigma_1 (1 + \sigma_2^2)^{0.5}} \right] (\mu_2)^{0.5} \prod_{S_i = 0} \left[\frac{S_i - Z_i a}{\sigma_2} \right] \quad (5)$$

L'estimation des coefficients de régressions a et b est basée sur l'hypothèse de l'homogénéité de l'échantillon; alors que comme nous savons les consommateurs viennent de milieux très diversifiés. Certainement, ils se distinguent par leurs `` critères socio – démographique (âge, revenu, catégorie socio - professionnelle, niveau d'éducation, race, nationalité, classe sociale) ou leurs critères psychographiques (le style de vie, la personnalité) ", ce qui influence, par ailleurs, leurs décisions et comportements. De ce fait, les consommateurs ont des comportements assez variés, puisqu'ils sont soumis à un très grand nombre de stimuli.

Donc, nous pouvons saisir l'intérêt d'une segmentation de notre échantillon afin de minimiser le biais produit suite à cette hétérogénéité.

Cette segmentation peut se faire d'une façon exogène au modèle, ce qui nécessite des informations qui permettent de séparer un échantillon en un petit nombre de groupes relativement homogènes et mutuellement exclusifs. Une des approches de modélisation exogène de l'hétérogénéité est traduite par les modèles à changements structurels ; qui nécessitent une spécification du mécanisme de changement.

Le problème d'hétérogénéité peut être aussi traité de façon endogène au modèle et ceci dans le cas où l'information sur l'appartenance des individus aux segments peut ne pas être disponible ou ces informations peuvent ne pas être pertinentes comme pour les données démographiques qui ne peuvent décrire convenablement les différences comportementales et donc ils ne peuvent être de bons critères de segmentation.

Desarbo et Choi (1999) ont endogénéisé l'hétérogénéité en construisant une nouvelle procédure d'estimation basé sur la méthode de maximum de vraisemblance sans avoir besoin de base de segmentation pré-spécifiée. Ils traitent l'hétérogénéité en se basant sur le comportement de la population mère évalué grâce aux paramètres estimés par le modèle. Dans la construction de la fonction de maximum de vraisemblance, ils se sont basés sur la probabilité à posteriori d'appartenance et qui est inconnue, les π_j sont prises comme une variable aléatoire suivant l'approche bayésienne. C'est cette dernière méthodologie que nous avons choisie dans la modélisation de notre étude.

Le modèle Probit avec auto - sélection et endogénéisation de l'hétérogénéité :

A la lumière de ce problème, nous proposons de classifier les sujets de l'échantillon dans un petit nombre de classes relativement homogènes. Cette supposition permet au sujet dans le même segment d'être traité comme une réplique l'un de l'autre dans le cadre de cette formulation à structure latente. L'échantillon sera partagé dans notre cas en J segments mutuellement exclusifs.

Un consommateur i ne peut appartenir qu'à un segment spécifique j tel que :

$$(1 \text{ } j \text{ } J) \text{ avec } 0 \leq \pi_j \leq 1 \text{ et } \sum_{j=1}^J \pi_j = 1. \text{ Nous notons le vecteur } \pi = (\pi_1, \dots, \pi_J).$$

En se basant sur les études faite par Jedidi et alii (1993) et par Desarbo et Choi (1999)

la distribution jointe de y_i et S_i dans ce cas est :

$$g(y_i, S_i / A, B, \pi) = \sum_{j=1}^J \pi_j f(y_i, S_i / a_j, b_j, \pi_{1j}, \pi_{2j}, \pi_j) \quad (6)$$

$$\sum_{j=1}^J \pi_j [1 - \left(\frac{X_i' b_j}{\pi_{1j}}\right)^{1/2} y_i] \left[\frac{X_i' b_j - \pi_j (\pi_{1j} / \pi_{2j}) (S_i - Z_i' a_j)}{\pi_{1j} (1 - \pi_j^2)^{0.5}} \right] (\pi_{2j})^{0.5} \left[\frac{S_i - Z_i' a_j}{\pi_{2j}} \right]^{y_i}$$

Avec π_j (b_1, \dots, b_J) et π_j (a_1, \dots, a_J) les coefficients de régressions pour les différents groupes de l'échantillon.

$\pi_{1j}, \pi_{2j}, \pi_j$ Forment la matrice de variance – covariance du groupe j tel que :

$$\pi_j = \begin{pmatrix} \pi_{1j} & \pi_{12j} & \pi_j \\ \pi_{2j} & \pi_{22j} & \pi_j \\ \pi_j & \pi_j & \pi_j \end{pmatrix} \quad (7)$$

On note Σ la matrice de variance – covariance globale qui est une matrice bloc diagonale ;
chaque bloc correspond à un Σ_j différent.

Enfin $f(y_i, S_i / a_j, b_j, \Sigma_{1j}, \Sigma_{2j}, \Sigma_j)$ est la fonction de distribution jointe conditionnelle de y_i et S_i ; définie sous l'hypothèse d'un processus de décision Bernoullien et supposent l'indépendance d'un individu i à un autre; elle s'écrit comme suit :

$$f(y_i, S_i / a_j, b_j, \Sigma_{1j}, \Sigma_{2j}, \Sigma_j) = \left[\frac{1}{\Sigma_{1j}} \right]^{1/2} \left[\frac{X_i' b_j - \Sigma_j^{-1} (\Sigma_{1j} / \Sigma_{2j}) (S_i - Z_i' a_j)}{\Sigma_{1j} (1 - \Sigma_j^{-2})^{0.5}} \right] (\Sigma_{2j})^{0.5} \left[\frac{S_i - Z_i' a_j}{\Sigma_{2j}} \right]^{y_i} \quad (8)$$

La probabilité a posteriori qu'un consommateur i appartient au segment j est noté par :

$h_{ij}(A, B, \Sigma, \Sigma_j)$. Elle est calculée en utilisant le théorème de Bayes et en prenant les paramètres estimés de A, B, Σ et Σ_j à chaque itération et les données observées : $y_i (y_1 \dots y_I)$ et $S_i (S_1 \dots S_I)$.

Ainsi nous avons :

$$h_{ij}(A, B, \Sigma, \Sigma_j) = \frac{p(j / y_i, S_i, A, B, \Sigma, \Sigma_j) p(j / y_i, S_i) / p(y_i, S_i)}{p(j) p(y_i, S_i / j) / p(y_i, S_i)} = \frac{\Sigma_j^{-1} \left[\frac{X_i' b_j - \Sigma_j^{-1} (\Sigma_{1j} / \Sigma_{2j}) (S_i - Z_i' a_j)}{\Sigma_{1j} (1 - \Sigma_j^{-2})^{0.5}} \right] (\Sigma_{2j})^{0.5} \left[\frac{S_i - Z_i' a_j}{\Sigma_{2j}} \right]^{y_i}}{\sum_{j=1}^J \Sigma_j^{-1} \left[\frac{X_i' b_j - \Sigma_j^{-1} (\Sigma_{1j} / \Sigma_{2j}) (S_i - Z_i' a_j)}{\Sigma_{1j} (1 - \Sigma_j^{-2})^{0.5}} \right] (\Sigma_{2j})^{0.5} \left[\frac{S_i - Z_i' a_j}{\Sigma_{2j}} \right]^{y_i}} \quad (9)$$

Le log de vraisemblance s'écrit alors, comme suit :

$$\log L_c(A, B, \Sigma, \Sigma_j / y, S, X, Z) = \log \prod_{i=1}^I \prod_{j=1}^J f(y_i, S_i, A, B, \Sigma, \Sigma_j) \quad (10)$$

$$= \sum_{i=1}^I \log \prod_{j=1}^J \left[\frac{1}{\Sigma_{1j}} \right]^{1/2} \left[\frac{X_i' b_j - \Sigma_j^{-1} (\Sigma_{1j} / \Sigma_{2j}) (S_i - Z_i' a_j)}{\Sigma_{1j} (1 - \Sigma_j^{-2})^{0.5}} \right] (\Sigma_{2j})^{0.5} \left[\frac{S_i - Z_i' a_j}{\Sigma_{2j}} \right]^{y_i}$$

Où $X = (X_1 \dots X_I)$; $Z = (Z_1 \dots Z_I)$ et $y = (y_1 \dots y_I)$; et aussi $S = (S_1 \dots S_I)$.

Afin d'estimer les paramètres de cette fonction de log vraisemblance L_c (où nous ignorons le terme constant sans perte de généralité), nous allons utiliser une version modifiée de l'algorithme

E-M (estimation-maximisation) tout en respectant les restrictions : $0 \leq \pi_j \leq 1$ et $\sum_{j=1}^J \pi_j = 1$

Nous utilisons l'algorithme E – M dans la maximisation de la fonction de log vraisemblance (10) par des itérations alternatives entre ses deux étapes : l'étape E d'estimation et l'étape M de maximisation. Cet algorithme a été introduit par Demster, Lair et Rubin (1977) afin de chercher l'estimateur du maximum de vraisemblance dans le cas où quelques variables sont non observables.

Dans notre cas l'algorithme E – M, présente les étapes suivantes :

Dans l'étape E de l'estimation, nous utilisons la fonction h_{ij} dans chaque itération et nous calculons ainsi la probabilité a posteriori pour les consommateurs. Par la suite, nous calculons la proportion associée au groupe j : π_j comme une fonction des probabilités a posteriori des consommateurs. Ensuite, nous passons à la deuxième étape M où la fonction log de vraisemblance est maximisée par rapport aux paramètres de la structure latente B, A, S . L'algorithme passe d'une étape à l'autre suivant une règle de mise à jour, jusqu'à la convergence.

Pour plus de détails sur les deux étapes de l'algorithme E – M voir Annexe 1.

Détermination du nombre de segment :

Dans l'étape précédente nous avons pu former des groupements homogènes. Cependant, il nous faut choisir le nombre de segments adéquat. Pour ce fait nous utilisons les heuristiques AIC, BIC et CAIC.

En effet, la fixation du nombre de segment (valeur de J) est faite en utilisant le critère d'information d'Akaike AIC (1974) modifié proposé par Bozdogon et Sclove (1984) et qui consiste à choisir le nombre de segments J qui minimise l'équation :

$$AIC_j = -2 \log L - 2N_j \quad \text{Tel que : } N_j = JH - 2J - 1 \quad (11)$$

Où N_j est le nombre de paramètres libres dans le modèle total.

Nous pouvons utiliser de même le critère d'information de Schwartz (1978) (BIC) :

$$BIC_j = -2 \log L - N_j (\log I) \quad (12)$$

Où celui de Bozdogan (1987) proposé pour les grands échantillons qui corrige le biais de surestimation de AIC et BIC en pénalisant le sur paramétrage. Cette heuristique est défini comme suit :

$$CAIC_j = -2 \log L + N_j (\log I + 1) \quad (13)$$

Avec : $\log L$ est la fonction de log vraisemblance, H est le nombre de variable à estimer et I est la taille de l'échantillon.

APPLICATION :

Présentation de l'enquête et Choix des variables de l'étude :

Dans notre travail nous nous basons sur une enquête originale élaborée en Août 2000 par un bureau d'étude spécialisé, supervisé par l'ONTT (Office Nationale De Touriste Tunisien). Cette enquête a été faite auprès d'un échantillon de touristes réparties sur tout le territoire tunisien, le découpage de l'échantillon s'est fait en fonction de la capacité des lits par catégorie d'hôtel et par région. Le choix de la population à enquêter est fait au hasard, en utilisant une table de choix au hasard en fonction des numéros de chambres occupées. La taille requise de l'échantillon a été de 1032 personnes. Les questions de cette enquête sont du genre fermé à choix multiples.

Les rubriques du questionnaire utilisé sont comme suit :

- Fiche signalétique 1 : Région, Hôtel, catégorie de l'hôtel.
- Fiche signalétique 2 : Profil du touriste (âge, profession, nationalité, sexe, nombre de visites antérieures pour la Tunisie).
- Motivation du séjour.
- Satisfaction par rapport à l'environnement hôtelier.
- Satisfaction par rapport à l'environnement extra hôtelier.
- Synthèse de satisfaction par rapport au séjour.
- Opinion retenue sur la Tunisie (l'intention de retour).

Nous avons expliqué, dans notre modèle à deux étapes, comment la satisfaction du consommateur conditionne son intention de fidélité tout en tenant compte de l'importance de l'hétérogénéité de l'échantillon.

La variable endogène de la première partie du modèle est le degré de satisfaction (SA).

Les variables exogènes de cette première partie sont :

La catégorie de l'hôtel (CH) qui a été choisie nous a permis d'avoir une indication sur les capacités financières du touriste qui pour leurs parts conditionnent souvent ses critères de satisfaction.

La deuxième variable exogène est l'appréciation générale de l'hôtel (AH); elle nous a procuré une information sur le degré de satisfaction du touriste dans l'hôtel ce qui lui permet d'avoir une place prépondérante dans la spécification du degré de satisfaction générale du touriste.

L'accueil dans l'hôtel (ACC) a été choisi comme variable exogène puisqu'elle nous informe sur le côté relationnel dans ce secteur de service.

La dernière variable exogène dans cette première partie du modèle est la fréquence de visite en Tunisie (FV) qui nous a donné une idée sur l'expérience du touriste.

La variable endogène de la deuxième partie du modèle est l'intention de retourner en Tunisie (RE).

La fréquence de visite en Tunisie (FV) est utilisée aussi dans cette étape comme variable exogène puisqu'elle nous procure une information sur le degré de fidélité du touriste à la Tunisie.

Nous avons choisi les variables exogènes prestation de service réception, qualité de service dans les chambres et qualité de service restauration afin d'avoir une idée en détail sur l'importance et l'influence de chaque service dans l'hôtel sur la décision de retour du touriste.

Test d'hétérogénéité :

Les résultats obtenus ont été calculés suite à l'application du modèle Probit avec auto – sélection et à la segmentation de l'échantillon comme suit :

Pour $J = 1$ nous avons pris l'échantillon total.

Pour $J = 2$ nous avons segmenté notre échantillon en deux classes suivant que les touristes sont très bien satisfaits ou non.

Pour $J = 3$ nous avons segmenté notre échantillon en trois classes suivant que les touristes sont très bien satisfaits, bien satisfaits ou non.

Une segmentation en deux groupes est de loin préférable au deux autres; puisqu'elle a la plus grande Log L et la plus faible valeur du AIC, du CAIC et du BIC et par suite elle incarne la meilleure spécification.

Autrement dit nous rejetons l'hypothèse d'homogénéité en faveur de celle hétérogénéité. Néanmoins cette d'hétérogénéité ne concerne que deux groupes et pas trois comme le suggère la troisième spécification (voir Annexe 2).

Arguant de ces conclusions, nous mènerons des estimations agrégées et des estimations différenciées en deux groupes d'estimation, nous montrerons dans ce qui suit que la prise en compte de l'hétérogénéité des groupes est indispensable.

Interprétation des résultats :

Les développements théoriques nous ont permis de conceptualiser un modèle économétrique de données qualitatives; qui nous permet de remédier au problème de l'hétérogénéité de l'échantillon, donnant à une spécification plus adéquate à la relation entre le degré de satisfaction des touristes et leurs intentions de fidélités. D'après l'étude préliminaire des données, 98.8% des touristes qui sont très bien satisfaits, déclarent qu'ils vont retourner en Tunisie. Afin, d'estimer les variables des modèles : Probit avec auto - sélection dans les deux cas agrégé et désagrégé, nous avons utilisé le logiciel Limdep 7. Cependant, pour estimer le modèle Probit avec auto - sélection désagrégé avec endogénéisation de l'hétérogénéité nous avons opté pour le logiciel de programmation Gauss 3.2.27; puisque la forme de la fonction de vraisemblance n'est pas courante.

Estimation du Probit avec auto - sélection :

Les estimations du Probit avec auto - sélection ont été effectuées sur notre échantillon total et sur les deux sous groupes, nous avons utilisé le modèle Probit à deux étapes, dont la première conditionne le passage à la deuxième étape suivant que le touriste est très bien satisfait ou non.

Nous nous sommes basés sur l'équation de vraisemblance définie par l'équation (5) pour l'ensemble de l'échantillon ; composé de 1032 individus.

Les résultats de l'estimation sont présents dans le tableau 2. Nous remarquons que tous les coefficients de régression sont statistiquement significatifs. Cependant, ils ne sont pas tous de mêmes signes. Les variables : `` l'accueil dans l'hôtel " (ACC), `` L'appréciation générale de l'hôtel " (AH) augmentent avec l'augmentation du degré de la satisfaction du touriste, leurs coefficients associés sont de signes négatifs puis que la variable endogène est codifiée sous la forme : 0 = très bien apprécier le séjour ; 1 = Autre.

Par contre, `` la fréquence de visite " (FV) et `` la catégorie d'hôtel " (CH) évoluent en sens opposés avec le degré de satisfaction (leurs coefficients associés sont de signes positifs).

Nous pouvons dire que le touriste qui a une grande expérience de la Tunisie et qui a des moyens financiers importants, est plus exigeant dans ses critères de satisfaction.

L'estimation du Probit de la deuxième étape, du modèle agrégé, n'est réalisée que sur les 625 touristes qui ne sont pas très bien satisfaits. Les coefficients de régression associés aux variables exogènes : `` Prestation de service réception " (PR), `` prestation de service restauration " (SR) et `` qualité de service chambre " (SC) sont de signe positifs. En d'autre terme, plus l'appréciation du touriste de ces trois départements (restauration, réception et hébergement) est grande, plus il aura une intention de retour plus favorable.

La non significativité statistique de la fréquence de visite, nous indique que les touristes qui ne sont pas très bien satisfaits ne prennent pas en compte le nombre de leurs visites et expériences touristiques en Tunisie lors de leurs décisions de retours.

Estimation différenciée par groupe :

Nous avons estimé le modèle Probit avec auto-sélection pour deux groupes différents de touristes : le premier correspond aux touristes qui ont visité la Tunisie trois fois ou plus et le deuxième correspond aux touristes qui ont visité la Tunisie une ou deux fois

(Voir A₄ : Tableau 3).

En nous basant sur les résultats du premier groupe (les 776 touristes qui ont visité la Tunisie trois fois ou plus), à la première étape, nous remarquons que toutes les variables ont des effets statistiquement significatifs et seuls les coefficients de régression associés à `` la fréquence de visite " (FV) et `` la catégorie de l'hôtel " (CH) sont positifs comme pour le cas du modèle agrégé. Cependant, les coefficients de régression de la deuxième partie du modèle, appliquée à ce groupe, sont tous positifs. Donc `` la prestation de service réception " (PR), `` la fréquence de visite " (FV), `` la prestation de service restauration " (SR) et `` la qualité de service chambre " (SC) augmentent au fur et à mesure que l'intention de retour du touriste augmente. Le coefficient de régression associé à `` la fréquence de visite " (FV) est positif tout en restant non significatif. Ce changement de signe est dû au fait que le premier groupe n'est composé que de touristes qui ont une large expérience de la Tunisie mais qui ne sont pas très bien satisfaits (487 touristes). Néanmoins, le coefficient de régression associé à `` la qualité de service chambre " (SC) est

devenue statistiquement non significatif. Nous pouvons conclure alors que les touristes, qui ont une grande expérience de la Tunisie mais qui ne sont pas très bien satisfaits, n'accordent pas d'importance aux services dans les chambres.

Dans l'analyse des résultats de la première étape, concernant le deuxième groupe, les coefficients de régression gardent les mêmes signes que le premier groupe.

En effet, les coefficients de régression associés à `` la fréquence de visite " (FV) et

`` la catégorie de l'hôtel " (CH) sont positifs et ceux associés à ``l'accueil dans l'hôtel " (ACC) et `` l'appréciation générale de l'hôtel" (AH) sont négatifs. La seule différence par rapport au premier groupe est que le coefficient de régression associé à la variable accueil dans l'hôtel n'est plus significatif. Ce résultat s'explique, entre autre, par le fait que ces touristes n'ont pas encore construit de relation interpersonnelle dans l'hôtel. L'application de ce modèle à deux étapes, au deuxième groupe de touriste, nous permet de confirmer que les coefficients de régression ont conservé le même signe (positif) que ceux du premier groupe dans cette deuxième étape. Le coefficient de régression associé à `` la fréquence de visite " (FV) est non significatif, pour les 138 touristes qui ne sont pas très bien satisfaits de leur séjour en Tunisie. Le coefficient de régression associé à `` la prestation de service réception " (PR) est aussi non significatif. Nous remarquons enfin, que seul le coefficient de corrélation pour le deuxième groupe est statistiquement non significatif et aussi que les estimations des écarts types des erreurs dans les deux classes sont moins élevés que ceux du modèle agrégé ($s_1 = 0.3480804$, $s_2 = 0.2604025$ et $s = 0.4164531$), ce qui nous permet de comprendre l'importance de la segmentation dans l'explication de l'influence de chaque variable explicative sur le comportement des touristes.

Estimation du Probit avec auto - sélection à structure latente généralisée :

Lors de l'estimation de l'impact du degré de la satisfaction des touristes sur leurs intentions de retour en Tunisie par le modèle Probit avec auto –sélection à structure latente généralisé. Nous avons utilisé le logiciel Gauss 3.2.27 pour la programmation de l'algorithme E –M. Nous avons pris comme valeurs initiaux les résultats des estimations du modèle Probit avec auto –sélection pour les deux groupes de touristes dans l'étape E. Nous nous sommes basés sur ces valeurs initiales dans le calcul de la probabilité à posteriori de chaque individu dans chaque segment à l'itération (r) sans savoir auparavant à quelle classe le touriste appartient. Par la suite, nous avons estimé les proportions de l'échantillon π_j , pour l'itération (r+1), en utilisant les probabilités à

posteriori. Ces variables représentent le pourcentage d'individu dans chaque groupe; lors de la première itération elles sont égales à : $\gamma_1 = 0.23636763$ et $\gamma_2 = 0.76363237$. A l'étape M, nous maximisons la fonction de log vraisemblance en prenant la valeur de γ calculée à l'étape E et en se basant sur la méthode Berndt –Hall – Hall –Hausman (BHHH). ¹Suite à l'utilisation de notre algorithme E – M les nouvelles valeurs obtenues de γ sont $\gamma_1 = 0.38177719$ et $\gamma_2 = 0.61822281$ dont les résultats sont données par le tableau 4.

L'étude des résultats du premier groupe comportant 244 touristes; nous a permis de constater, qu'à la première étape, les coefficients de régression associés aux variables explicatives gardent les mêmes signes que le modèle précédent. En d'autres termes, les coefficients de régression associés aux variables `` accueil dans l'hôtel " (ACC) et `` appréciation de l'hôtel " (AH) augmentent avec l'augmentation du degré de la satisfaction du touriste (leurs coefficients associés sont de signes négatifs). Par contre, `` la fréquence de visite " (FV) et `` la catégorie d'hôtel " (CH) évoluent en sens opposés avec le degré de satisfaction (Leurs coefficients associés sont de signes positifs). A cette même étape, tous les coefficients de régression ont présenté des tests statistiquement non significatifs.

Donc pour tous les membres de ce groupe, il y a d'autres variables qui influent sur leur degré de satisfaction comme les variables de l'environnement extra hôtelier. Les coefficients de régression associés aux variables explicatives de la deuxième étape sont tous statistiquement significatifs et de signes positifs, ils évoluent dans le même sens que l'intention de retour.

($\beta_{PR} = 0.0316$, $\beta_{FV} = 0.2060$, $\beta_{SC} = 0.0213$ et $\beta_{SR} = 0.0209$).

En ce qui concerne le deuxième groupe des 788 touristes (donnés par les valeurs des γ_j trouvés) tous les coefficients de régression associés aux variables explicatives de la première étape sont significatifs. Nous remarquons aussi le changement de signe survenu au coefficient de régression associé à la variable `` accueil dans l'hôtel " (ACC) (Devenu positif); cette variable évolue donc en sens opposé avec le degré de satisfaction. Le deuxième coefficient de régression qui a changé de signe est celui associé à la variable `` fréquence de visite " (FV) (devenu négatif). Ceci se traduit par le fait que toute augmentation dans la fréquence de visite entraîne une augmentation

¹ Le choix de cet algorithme est survenu après un grand nombre d'essais suivant plusieurs algorithmes de linéarisation (la méthode du gradient conjugué, Newton-Raphson, steepest descent (SD), l'algorithme de Broyden, Fletcher, Goldfarb et Shanno (BFGS), l'algorithme de Davidon, Fletcher et Powell (DFP) ...), puisque nous avons rencontré un grand nombre de difficultés; dues essentiellement à l'inexistence du maximum de notre vraisemblance, ce qui a engendré une matrice Hessienne non inversible.

dans le degré de satisfaction. Les résultats de la deuxième étape dans ce groupe sont similaires à celle du premier groupe. En d'autre terme, tous les coefficients de régression associés aux variables explicatives sont significatifs et de signe positifs ($Z_{PR} = 7.384$, $Z_{FV} = -26.736$, $Z_{SC} = 9.227$ et $Z_{SR} = 7.343$). Le seul changement est que le coefficient de régression associé à `` la variable fréquence de visite " (FV) est devenu négatif; c'est à dire les touristes qui ne sont pas très bien satisfaits de leurs visites en Tunisie, ont des intentions de retour de plus en plus faible que leurs préférences de visite est importante. Nous remarquons enfin, que les estimations des écarts types des erreurs dans les deux classes pour ce modèle sont moins élevés que ceux du modèle précédent avec choix exogène des groupes ($s_1 = 0.0680$, $s_2 = 0.2131$), nous concluons donc, que ce modèle nous permet de réduire le biais produit par l'hétérogénéité de l'échantillon.

CONCLUSION :

Les modèles d'économétrie qualitatifs permettent de modéliser un grand nombre de problèmes liés au domaine Marketing. Cependant, leurs résultats peuvent être biaisés par l'hétérogénéité de l'échantillon. C'est ainsi, que nous avons essayé de remédier à ce problème en dégagant les comportements de différents types de ménage.

Nous concluons suite aux estimations du modèle Probit avec auto – sélection dans le cas agrégée et dans le cas de deux classes de l'échantillon totale que ce modèle est plus adéquat pour l'explication de l'intention de fidélité des touristes, car il incorpore le degré de satisfaction comme variable à expliquer pour étudier la décision de retour du touriste d'une façon séquentielle. Ainsi, l'estimation de l'échantillon suivant deux groupes est meilleure que l'estimation du modèle agrégé.

L'endogénéisation de l'hétérogénéité de l'échantillon dans le modèle Probit avec auto – sélection nous a permis de mieux saisir l'influence de chaque variable explicative sur l'intention de retour des touristes.

Ce modèle de recherche à structure latente généralisée, en plus de ses avantages économétriques, peut être utilisé comme une nouvelle méthode de segmentation du marché, puisque par l'emploi de cette méthode, les managers des entreprises peuvent identifier les groupes de consommateurs qui se différencient par leurs degrés de satisfaction.

Il serait intéressant dans les recherches futures d'appliquer la méthode de maximum de vraisemblance dans le cas de l'endogénéisation de l'hétérogénéité des échantillons à d'autres modèles économétriques, afin d'éliminer des biais liés à l'omission de l'hétérogénéité et des biais relatifs à la segmentation ad hoc des ménages.

REFERENCES:

- Bearden William O., Teel Jesse E., (February 1983): ``Selected determinants of consumer satisfaction and complaint reports: *Journal of marketing research*, 20, 21 - 28.
- bon J., Tissier- Desbordes E., (Janvier- février 2000) : *Fidéliser les clients ? oui mais...*, Revue Française de Gestion, 52- 60.
- Bozzo C., (1999): Fidélisé, rétention et inertie des clients..., *Centre d'Etudes et Recherches sur les Organisations et la Gestion*, WP n° 547.
- Desarbo W., Choi J., (1999): A latent structure double hurdle regression model for exploring heterogeneity in consumer search patterns: *Journal of Econometrics*, 89, 423 - 455.
- Dufer J. et Moulins J –L., (1989): La relation entre la satisfaction du consommateur et sa fidélité à la marque : un examen critique, *Recherche et Application en Marketing*, 2, 21 - 36.
- Gourieroux C., (1989): *Econométrie des variables qualitatives*, Economica, 2^{ème} édition.
- Greene William H., (2000): *Econometric analysis*, International Edition, pp 908.
- Jedidi K., Ramaswamy V., Desarbo W., (2000): A maximum likelihood method for latent class regression involving a censored dependent variable, *Psychometrika*, 58, 3, 375 - 394.
- kandampully J., suhartanto D., (2000): Customer loyalty in the hôtel industry : the role of customer satisfaction and image, *International Journal of Contemporary Hospitality Management*, 12.
- Kotler P., Dubois B., (1983) : *Marketing management*, 4^{ème} édition.
- Office nationale du tourisme (2000), *Tunisien Le Tourisme Tunisien en Chiffres*.
- Maddala G. S., (1983) : *Limited - dependent and qualitative variables in econometrics*, Cambridge University Press.
- McDougall Gordon H.G. et Levesque T., (2000): Customer satisfaction with services: putting perceived value into the equation", *Journal of Services Marketing*, 14.

Colloque ATM 2004

Meurgey B., bouchet., (novembre 2000) : Intérêt et limite du label qualité dans la développement du loisir sportif , *www.infosport.org*.

Neal Radford M., Hinton Geoffrey E. (2002): A view of the EM algorithm that justifies incremental, sparse, and other variants, *www.cs.toronto.edu*.

Mottelet S.,Elbagdouri MD., (Automne 2000) : Optimisation non-linéaire, Faculté des Sciences Marrakech, Maroc.

Oliver Richard L., (Novembre 1980): A cognitive model of the antecedents and consequences of satisfaction decisions ', *Jounal of Marketing Research*, XVII, 460- 469.

Thomas A., (2000) : Econométrie des variables qualitatives, Editions Dunod, 140- 142.

Annexe 1 :

L'algorithme E – M :

L'étape – E :

Dans cette étape de l'algorithme, nous traitons la probabilité a posteriori des individus comme donnée manquante puisqu'il n'est pas connu à l'avance à quelle classe spécifique les consommateurs peuvent appartenir. Cette probabilité à l'itération r : $h_{ij}^{(r)}$ est calculée comme suit :

$$h_{ij}^{(r)}(A, B, ?, ?) = \frac{?^{(r)}_j f(y_i, S_i / a, b, ?_{1j}, ?_{2j}, ?)}{g(y_i, S_i / A, B, ?, ?)}$$

Ensuite, on estime $?_j$ par la différenciation de la fonction log de vraisemblance :

$$a = (a_1, \dots, a_H)'$$

Donc la valeur de la proportion des groupes $?^{(r+1)}$ peut être connue à l'itération r sur la base de $A^{(r)}, B^{(r)}, ?^{(r)}$ après l'obtention de $h_{ij}^{(r)}$. Après cette étape de prévision, nous passons à l'étape de maximisation pour obtenir une nouvelle prévision des paramètres estimés : $A^{(r+1)}, B^{(r+1)}, ?^{(r+1)}$; en se basant sur l'équation de log de vraisemblance construite comme suit :

$$\log L_c(A, B, ?, ? / y, S, X, Z) = \sum_{i=1}^I \log[?^{(r+1)}_j f(y_i, S_i / a, b, ?_{1j}, ?_{2j}, ?) / g(y_i, S_i / A, B, ?, ?)]$$

Et Avec $H = (h_{ij})$ considérée comme donnée manquante.

L'étape M :

Nous avons choisi, pour la maximisation de la vraisemblance dans cette étape, l'algorithme Berndt - Hall - Hall – Hausman (BHHH). Ce processus itératif continu jusqu'à ce que l'amélioration de la fonction log vraisemblance serait impossible suivant le critère de convergence $?_{telque} < 0.001$.

La méthode d'estimation, exposée ci-dessus, nous donne simultanément une estimation des paramètres de régression $\hat{\beta}$ et $\hat{\sigma}^2$ la matrice des covariances $\hat{\Sigma}$ les proportions de l'échantillon \hat{p}_j et les probabilités a posteriori des consommateurs \hat{H} de chaque segment J.

Annexe 2 :

Tableau 1 : Choix du nombre de segment :

J	Log L	AIC	CAIC	BIC
1	654,1892	1319,3784	13480744	1343,1348
2	615,8918*	1253,7836*	1319,1148*	1308,1148*
3	620,4142	1274,8284	1375,7948	1358,7948

? Se limiter à une segmentation a deux groupes.

Annexe 3 :

Tableau 2 : Estimation par le modèle Probit avec auto –sélection

Cas de l'échantillon total

Variable	Coefficient (β)	T –Student (Z)
Première étape		
FV	0.2603758	7.290
ACC	-0.1412651	-2.331
AH	-0.3888744	-6.259
CH	0.4049163	8.257
Deuxième étape		
PR	0.0982404	4.756
FV	-0.0175506	-1.015
SC	0.0672766	3.459
SR	0.089021	3.940
s	0.4164531	14.648
?	0.4996884	4.225

Annexe 4 :

Tableau 3 : Estimation par le modèle Probit avec auto –sélection

Cas de deux groupes

Estimation du premier groupe

Variable	Coefficient (β)	T –Student (Z)
Première étape		
FV	0.610455	9.03
ACC	-0.3184751	-4.177
AH	-0.5377702	-6.81
CH	0.251318	4.184
Deuxième étape		
PR	0.098117	2.998
FV	0.021224	0.72
SC	0.0320043	1.294
SR	0.039508	2.332
s	0.3480804	8.781
?	0.4709547	3.572

Estimation du deuxième groupe

Variable	Coefficient (β)	T –Student (Z)
Première étape		
FV	0.7487993	4.317
ACC	-0.1420941	-1.136
AH	-0.4575308	-3.070
CH	0.3817209	3.891
Deuxième étape		
PR	0.002432515	0.08
FV	0.10446336	1.557
SC	0.1258881	4.094
SR	0.1196864	3.624
s	0.2604025	9.0522
?	0.4750066	1.768

Annexe 5 :

Tableau 4 : Estimation par le modèle Probit avec auto –sélection
et endogènèisation de l'hétérogénéité

Cas de deux groupes

Estimation du premier groupe

Variable	Coefficient (β)	T –Student (Z)
Première étape		
FV	21.8897	0.227
ACC	-0.9155	-1.034
AH	-16.8035	-0.232
CH	1.2600	0.973
Deuxième étape		
PR	0.0316	5.732
FV	0.2060	38.355
SC	0.0213	3.488
SR	0.0209	3.594
s	0.0680	31.492
?	0.7064	3.351

Estimation du deuxième groupe

Variable	Coefficient (β)	T –Student (Z)
Première étape		
FV	-0.2377	-4.814
ACC	0.3468	5.542
AH	-0.6254	-9.878
CH	0.5456	7.562
Deuxième étape		
PR	0.1134	7.384
FV	-0.2235	-26.736
SC	0.1541	9.227
SR	0.1311	7.343
s	0.2131	28.803
?	0.0067	0.076