



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Projet de Semestre

**Analyse théorique et empirique
de modèles de choix discrets**

ANTILLE Nicolas

Responsable:

BIERLAIRE Michel

Chaire de Recherche Opérationnelle
du Professeur Th. M. LIEBLING

Département de Mathématiques

Hiver 1999 - 2000

Table des matières

1	Introduction	2
2	Objectifs du projet	3
2.1	Etude théorique	3
2.2	Etude empirique	4
3	Analyse théorique du modèle cross-nested étendu	4
3.1	Le modèle logit multinomial et le modèle nested	5
3.1.1	Le modèle logit multinomial	5
3.1.2	Le modèle nested	5
3.2	Introduction au modèle cross-nested	6
3.3	La famille GEV	8
3.4	La formule de probabilité	11
3.5	Les dérivées partielles	13
4	Etude d'un problème de choix de mode de transport	17
4.1	Présentation du problème	17
4.2	Calibration du modèle	18
4.2.1	Idée générale	18
4.2.2	Le jeu de données	19
4.2.3	Le codage des variables	21
4.2.4	Les modèles successifs	23
4.2.5	Le nettoyage du modèle	24
4.3	Résultats	25
4.3.1	Les fonctions d'utilité	25
4.3.2	La valeur des coefficients	27
4.3.3	L'estimation d'un modèle nested	29
4.4	Analyse des résultats	30
4.4.1	Les coefficients non significatifs	30
4.4.2	Les coefficients significatifs	31
4.4.3	La valeur du temps	32
4.4.4	Quelques applications	33
5	Conclusion	40
6	Bibliographie	41

1 Introduction

Durant toute sa vie, un individu est confronté à des choix. Il doit en effet effectuer des choix professionnels, des choix sentimentaux, mais surtout il doit effectuer chaque jour une multitude de choix moins fondamentaux, mais nécessaires au bon déroulement de sa vie. Par exemple, il doit choisir dans quel magasin il va faire ses achats, mais aussi à quel journal il veut s'abonner ou encore quel mode de transport il va utiliser pour un déplacement. Un choix peut s'avérer difficile à effectuer ou pas, mais quelle que soit sa nature, on suppose que l'individu va presque toujours adopter la même démarche : il va évaluer les différentes possibilités qui s'offrent à lui en se basant sur des critères et choisit celle qui lui paraît la plus adéquate. Les critères de décision peuvent être très divers, mais lorsqu'ils sont quantifiables ou lorsqu'ils sont logiques(oui/non), alors il existe une théorie mathématique qui propose des modèles permettant de prédire le comportement d'individus confrontés à un choix. Cette théorie est la théorie des modèles de choix discrets (Ben-Akiva et Lerman, 1985 et Ben-Akiva et Bierlaire, 1999).

Comme toute modélisation mathématique, cette théorie repose sur plusieurs hypothèses fondamentales. Parmi celles-ci, on considère qu'un individu choisit l'alternative qui possède selon lui la plus grande utilité. En toute généralité, on suppose que l'utilité, qui est une variable aléatoire, se laisse écrire sous la forme suivante:

$$U = V + \varepsilon$$

où V est la partie déterministe et ε la partie aléatoire

Ainsi, les modèles vont devoir estimer la fonction d'utilité de chaque alternative à disposition de l'individu. Pour cela, on suppose que l'utilité est une variable aléatoire dont la partie déterministe V est linéaire en les coefficients inconnus. Cet aspect, ainsi que les méthodes utilisées pour estimer les coefficients de la fonction d'utilité, seront explicités ultérieurement.

Actuellement, deux modèles sont principalement utilisés pour l'analyse empirique de problèmes de choix discret. Le modèle logit multinomial est le plus simple, car il est basé sur une hypothèse d'indépendances des termes purement aléatoires des fonctions d'utilité. De ce fait, il ne tient pas compte des éventuelles corrélations entre les alternatives et aboutit parfois à des aberrations. Dans le but de remédier à cette faiblesse, un modèle un peu plus complexe a été développé : le modèle logit emboîté (nested logit model). Ce

modèle crée une partition de l'ensemble des alternatives qui tient partiellement compte des corrélations. Chaque élément de cette partition est appelé un nid. Ces deux modèles sont donc les modèles les plus communément utilisés aujourd'hui pour résoudre des problèmes du type "choix d'un mode de transport" ou "choix d'un chemin". Toutefois, on a constaté que même le modèle nested possède des limites dans ses applications, notamment lorsque les corrélations entre les alternatives sont multiples et complexes. Cela a conduit récemment au développement d'un nouveau modèle: le modèle logit emboîté croisé (cross-nested logit model) proposé par Vovsha(1997). Comme avant, l'ensemble des alternatives est partitionné en nids, mais de plus on introduit des coefficients supplémentaires qui traduisent pour chaque alternative son degré d'appartenance à chacun des nids. Ce modèle nécessite le calcul (estimation) de beaucoup de coefficients et aucune méthode d'estimation efficace n'a été proposée à ce jour. Toutefois, il est possible que dans le futur, il devienne le modèle de référence.

2 Objectifs du projet

Ce projet sur les modèles de choix discret se divise en deux parties distinctes, l'une à caractère théorique, l'autre à caractère empirique.

2.1 Etude théorique

Cette partie consiste en une étude théorique approfondie d'une version plus générale du modèle cross-nested. L'appartenance de ce modèle à la famille des modèles de valeurs extrêmes généralisés (GEV) sera tout d'abord démontrée dans cette section. Ensuite, on établira la formule permettant de calculer la probabilité de choisir une alternative quelconque dans l'ensemble de toutes les alternatives disponibles. Finalement, on calculera les dérivées partielles de la formule de probabilité par rapport à chacun des coefficients inconnus. Ces dérivées permettent par la suite d'implémenter un algorithme calculant les coefficients inconnus à l'aide d'une méthode de maximum de vraisemblance.

2.2 Etude empirique

Cette partie s'attache à la modélisation d'un problème concret de choix de mode de transport. Swissmetro est un projet qui consiste à relier les principales villes de Suisse par un réseau ferroviaire souterrain. Grâce à la sustentation magnétique, les trains pourraient atteindre des vitesses de près de 500 km/h. Dans le cadre d'une enquête demandée pour Swissmetro, des données ont été récoltées grâce à une méthode dite de préférences déclarées. Il s'agit de déclarations de préférences de voyageurs placés dans une situation hypothétique. Sous certaines conditions, les voyageurs interrogés devaient choisir entre la voiture, le train ou Swissmetro pour effectuer leur déplacement. L'étude consistera à traiter ces données et ensuite à calibrer un modèle à l'aide du logiciel HieLoW (Bierlaire, 1994), spécialisé pour étudier ce type de problèmes. Une fois le modèle calibré de façon satisfaisante, on pourra comparer les résultats obtenus grâce au modèle multinomial et au modèle nested, analyser les résultats et les interpréter dans le contexte du problème.

3 Analyse théorique du modèle cross-nested étendu

Avant de présenter les développements calculatoires concernant le modèle cross-nested, il convient pour une meilleure compréhension de définir certaines grandeurs et certains paramètres :

C : l'ensemble de toutes les alternatives disponibles

M : le nombre de nids dans la partition de C

V_i : la partie déterministe de la fonction d'utilité de l'alternative i

N : le nombre d'alternatives dans C

N.B. : Il faudrait normalement écrire C_n pour l'ensemble des alternatives disponibles pour l'individu n, car chaque individu peut avoir un ensemble de choix différent. En effet, il peut arriver que certains individus ne puissent pas choisir une ou plusieurs alternatives. Par exemple, une personne ne possédant pas de permis de conduire ne peut évidemment pas choisir la voiture comme mode de transport. De même, il faudrait noter V_{in} la partie déterministe de la fonction d'utilité de l'alternative i relativement à l'individu n. Toutefois, afin d'alléger l'écriture, cette notation ne sera introduite qu'à partir de la section 3.4.

Précisons encore qu'une des hypothèses principales concernant la modélisation

de problèmes de choix discret est que la partie déterministe de la fonction d'utilité dépend linéairement des attributs qui influencent le choix. Ainsi, on peut écrire:

$$V_k = \sum_t \beta_t x_{kt}$$

Ainsi, dans la pratique on cherchera la valeur des composantes du vecteur β et pas directement celle de la partie déterministe V_k de la fonction d'utilité. Toutefois, en connaissant l'expression de la dérivée partielle de $\ln P(i_n|C)$ par rapport à V_k , on trouve immédiatement la valeur de la dérivée partielle de $\ln P(i_n|C)$ par rapport à β_t . En effet:

$$\frac{\partial \ln P(i_n|C)}{\partial \beta_t} = \frac{\partial \ln P(i_n|C)}{\partial V_k} \frac{\partial V_k}{\partial \beta_t} = \frac{\partial \ln P(i_n|C)}{\partial V_k} x_{kt}$$

Cette remarque peut paraître prématurée, mais elle justifiera le calcul des dérivées de la section 3.5.

3.1 Le modèle logit multinomial et le modèle nested

3.1.1 Le modèle logit multinomial

Le modèle logit multinomial est un modèle extrêmement simple, car il ne tient pas compte des éventuelles corrélations entre les alternatives. On suppose de plus que la partie aléatoire de l'utilité ε est distribuée selon une loi Gumbel avec paramètre d'échelle μ . Ainsi, pour ce modèle, on écrira la probabilité de choisir l'alternative i dans C :

$$P(i|C) = \frac{e^{\mu V_i}}{\sum_{j \in C} e^{\mu V_j}}$$

3.1.2 Le modèle nested

Le modèle nested logit ou simplement modèle nested a été proposé pour la première fois par Ben-Akiva (1973) et est une extension du modèle logit multinomial. Cette extension vise à tenir compte des corrélations entre les alternatives. Pour cela, on partitionne l'ensemble C en M sous-ensembles que l'on appelle nids. Ces nids sont disjoints et leur réunion est l'ensemble C tout entier. Chaque alternative appartient à un seul nid. La fonction d'utilité de chaque alternative se compose d'un terme spécifique à l'alternative et d'un

terme associé au nid auquel appartient l'alternative. Ainsi, si l'alternative i appartient au nid C_m , on écrira:

$$U_i = \tilde{V}_i + \tilde{\varepsilon}_i + \tilde{V}_{C_m} + \tilde{\varepsilon}_{C_m}$$

Les termes d'erreur $\tilde{\varepsilon}_i$ et $\tilde{\varepsilon}_{C_m}$ sont supposés indépendants. Les termes d'erreur $\tilde{\varepsilon}_i$ sont indépendants et identiquement distribués selon une loi Gumbel avec paramètre d'échelle égal à μ_m . La distribution de $\tilde{\varepsilon}_{C_m}$ est telle que la variable aléatoire $\max_{j \in C_m} U_{jm}$ est distribué selon une loi Gumbel avec paramètre d'échelle μ . De plus, on associe à chaque nid une utilité composite :

$$V_{C_m} = \tilde{V}_{C_m} + \frac{1}{\mu_m} \ln \left(\sum_{j \in C_m} e^{\mu_m \tilde{V}_j} \right)$$

Ainsi, la probabilité pour un individu de choisir l'alternative i parmi l'ensemble C de toutes les alternatives est donnée par :

$$P(i|C) = P(C_m|C)P(i|C_m)$$

où

$$P(C_m|C) = \frac{e^{\mu V_{C_m}}}{\sum_{l=1}^M e^{\mu V_{C_l}}}$$

$$P(i|C_m) = \frac{e^{\mu_m \tilde{V}_i}}{\sum_{j \in C_m} e^{\mu_m \tilde{V}_j}}$$

Bien que ce modèle fournisse dans bien des cas de meilleurs résultats que le modèle logit multinomial, il ne permet pas de modéliser des situations lorsque les corrélations entre les alternatives sont multiples et complexes, c'est-à-dire lorsque deux alternatives appartenant à deux nids différents sont corrélées. C'est essentiellement pour remédier à cette faiblesse qu'a été développé le modèle cross-nested qui est présenté en détail dans les sections 3.2 à 3.5.

3.2 Introduction au modèle cross-nested

Ce deuxième paragraphe présente brièvement le modèle cross-nested. Le modèle cross-nested est une extension du modèle nested, car chaque alternative peut appartenir à plusieurs nids. Comme dans le modèle nested, l'ensemble C de toutes les alternatives est partitionné en M nids C_m . De plus,

pour chaque alternative i et pour chaque nid C_m , on introduit un paramètre α_{im} qui représente le degré d'appartenance de l'alternative i au nid C_m . Dans ce cas, l'utilité de l'alternative i est donnée par:

$$U_{im} = \tilde{V}_i + \tilde{\varepsilon}_i + \tilde{V}_{C_m} + \tilde{\varepsilon}_{C_m} + \ln \alpha_{im}$$

où $V_i = \tilde{V}_i + \tilde{V}_i$

Les termes d'erreur $\tilde{\varepsilon}_i$ et $\tilde{\varepsilon}_{C_m}$ sont supposés indépendants. Les termes d'erreur $\tilde{\varepsilon}_i$ sont indépendants et identiquement distribués selon une loi Gumbel avec paramètre d'échelle égal à μ_m . Dans la première version du cross-nested, Vovsha avait posé $\mu_m = 1$, ce qui restreignait la généralité du modèle. Ce paramètre d'échelle μ_m est une des raisons pour lesquelles on parle ici de modèle cross-nested étendu. La distribution de $\varepsilon_{\tilde{C}_m}$ est telle que la variable aléatoire $\max_{j \in C_m} U_{jm}$ est distribué selon une loi Gumbel avec paramètre d'échelle μ . Ainsi, la probabilité pour un individu de choisir l'alternative i parmi l'ensemble C de toutes les alternatives est donnée par :

$$P(i|C) = \sum_{m=1}^M P(C_m|C)P(i|C_m) \quad (1)$$

où

$$P(C_m|C) = \frac{e^{\frac{\mu}{\mu_m} V_{C_m}}}{\sum_{l=1}^M e^{\frac{\mu}{\mu_l} V_{C_l}}}$$

$$P(i|C_m) = \frac{\alpha_{im} e^{\tilde{V}_i \mu_m}}{\sum_{j \in C_m} \alpha_{jm} e^{\tilde{V}_j \mu_m}}$$

avec

$$V_{C_m} = \mu_m \tilde{V}_{C_m} + \ln \left(\sum_{j \in C_m} \alpha_{jm} e^{\tilde{V}_j \mu_m} \right)$$

3.3 La famille GEV

La famille des modèles de valeurs extrêmes généralisées est une famille de modèles qui inclus notamment le modèle multinomial, le modèle nested et le modèle cross-nested. A chacun de ces modèles on peut associer une fonction $G : \mathbb{R}_+^N \longrightarrow \mathbb{R}$ qui est différentiable et qui vérifie les propriétés suivantes (la relation entre la fonction et le modèle sera introduite dans la section suivante):

- 1) G est non-négative $\forall x \geq 0$
- 2) G est homogène de degré $\mu > 0$
- 3) $\lim_{x_i \rightarrow \infty} G(x_1, \dots, x_N) = \infty \quad \forall i = 1, 2, \dots, N \quad \forall x \geq 0$
- 4) la k^e dérivée partielle par rapport à k x_i distincts est non-négative si k est impair et non-positive si k est pair, donc :

$$\frac{\partial^k G(x)}{\partial x_{i_1} \dots \partial x_{i_k}} \begin{cases} \geq 0 & \text{si } k \text{ est impair} \\ \leq 0 & \text{si } k \text{ est pair} \end{cases} \quad \forall x \in \mathbb{R}_+^N$$

On va maintenant montrer que le modèle cross-nested étendu dérive de la fonction $G(x)$ suivante:

$$G(x) = \sum_{m=1}^M \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}$$

avec $\alpha_{jm} \geq 0$, $\mu > 0$ et $0 \geq \frac{\mu}{\mu_m} \geq 1, \mu_m \geq 0 \forall m$

- 1) Soit $x \in \mathbb{R}_+^N$, alors $\forall m$ et $\forall j$ on peut écrire:

$$\begin{aligned} & \alpha_{jm} x_j \geq 0 \\ \Rightarrow & \alpha_{jm} x_j^{\mu_m} \geq 0 \\ \Rightarrow & \sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \geq 0 \\ \Rightarrow & \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \geq 0 \\ \Rightarrow & \sum_{m=1}^M \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \geq 0 \\ \Rightarrow & G(x_1, \dots, x_N) \geq 0 \end{aligned}$$

donc $G(x)$ est non-négative et de plus $G(x)$ est clairement différentiable.

2) Soit $x \in \mathbb{R}_+^N$ et $\beta \geq 0$, on a alors:

$$\begin{aligned}
G(\beta x) &= \sum_{m=1}^M \left(\sum_{j \in C_m} \alpha_{jm} \beta^{\mu_m} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \\
&= \sum_{m=1}^M \left(\beta^{\mu_m} \sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \\
&= \sum_{m=1}^M \beta^{\mu} \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \\
&= \beta^{\mu} \sum_{m=1}^M \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \\
&= \beta^{\mu} G(x)
\end{aligned}$$

donc $G(x)$ est homogène de degré μ .

3) Vérifions la propriété relative aux limites :

$$\begin{aligned}
\lim_{x_i \rightarrow \infty} G(x_1, \dots, x_M) &= \lim_{x_i \rightarrow \infty} \sum_{m=1}^M \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \\
&= \sum_{m=1}^M \left(\lim_{x_i \rightarrow \infty} \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \right)
\end{aligned}$$

La condition est vérifiée, c'est-à-dire la limite vaut l'infini, si $\forall j, \exists m$ tel que $\alpha_{jm} > 0$. Cela revient à dire que la condition est vérifiée si on peut placer chaque alternative de C dans au moins un des nids C_m . Notons encore que la limite vaut $+\infty$ par positivité des coefficients α_{jm} , μ et μ_m . Remarquons au passage que l'hypothèse de Vovsha pour l'appartenance à la famille GEV était : $\sum_{m=1}^M \alpha_{jm}^{\mu} = 1 \forall j$.

4) Cette dernière propriété se démontre au travers d'une proposition :

Proposition:

$$\frac{\partial^k G(x)}{\partial x_{i_1} \dots \partial x_{i_k}} = \sum_{m=1}^M \left(\mu_m^k \prod_{n=i_1}^{i_k} (\alpha_{nm} x_n^{\mu_m-1}) \prod_{n=0}^{k-1} \left(\frac{\mu}{\mu_m} - n \right) \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu-k\mu_m}{\mu_m}} \right)$$

La démonstration se fera par récurrence:

Cas $k = 1$:

$$\begin{aligned} \frac{\partial^k G(x)}{\partial x_{i_1}} &= \sum_{m=1}^M \left(\frac{\mu}{\mu_m} \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu-\mu_m}{\mu_m}} \mu_m \alpha_{i_1 m} x_{i_1}^{\mu_m-1} \right) \\ &= \sum_{m=1}^M \left(\mu_m \alpha_{i_1 m} x_{i_1}^{\mu_m-1} \frac{\mu}{\mu_m} \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu-\mu_m}{\mu_m}} \right) \end{aligned}$$

proposition vérifiée pour $k = 1$.

Supposons la propriété vraie pour k et montrons qu'elle est vraie pour $k+1$:

Hypothèse :

$$\frac{\partial^k G(x)}{\partial x_{i_1} \dots \partial x_{i_k}} = \sum_{m=1}^M \left(\mu_m^k \prod_{n=i_1}^{i_k} (\alpha_{nm} x_n^{\mu_m-1}) \prod_{n=0}^{k-1} \left(\frac{\mu}{\mu_m} - n \right) \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu-k\mu_m}{\mu_m}} \right)$$

Conclusion :

$$\frac{\partial^{(k+1)} G(x)}{\partial x_{i_1} \dots \partial x_{i_{k+1}}} = \sum_{m=1}^M \left(\mu_m^{k+1} \prod_{n=i_1}^{i_{k+1}} (\alpha_{nm} x_n^{\mu_m-1}) \prod_{n=0}^k \left(\frac{\mu}{\mu_m} - n \right) \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu-(k+1)\mu_m}{\mu_m}} \right)$$

Démonstration :

$$\begin{aligned} \frac{\partial^{k+1} G(x)}{\partial x_{i_1} \dots \partial x_{i_{k+1}}} &= \frac{\partial}{\partial x_{i_{k+1}}} \frac{\partial^k G(x)}{\partial x_{i_1} \dots \partial x_{i_k}} \\ &= \frac{\partial}{\partial x_{i_{k+1}}} \sum_{m=1}^M \left(\mu_m^k \prod_{n=i_1}^{i_k} (\alpha_{nm} x_n^{\mu_m-1}) \prod_{n=0}^{k-1} \left(\frac{\mu}{\mu_m} - n \right) \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu-k\mu_m}{\mu_m}} \right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{m=1}^M \mu_m^k \prod_{n=i_1}^{i_k} (\alpha_{nm} x_n^{\mu_m-1}) \prod_{n=0}^{k-1} \left(\frac{\mu}{\mu_m} - n \right) \left(\frac{\mu}{\mu_m} - k \right) \\
&\quad \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu - k\mu_m - \mu_m}{\mu_m}} \alpha_{i_{k+1}} \mu_m x_{i_{k+1}}^{\mu_m-1} \\
&= \sum_{m=1}^M \left(\mu_m^{k+1} \prod_{n=i_1}^{i_{k+1}} (\alpha_{nm} x_n^{\mu_m-1}) \prod_{n=0}^k \left(\frac{\mu}{\mu_m} - n \right) \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu - (k+1)\mu_m}{\mu_m}} \right)
\end{aligned}$$

Etudions maintenant le signe de cette expression en fonction de k . Etant donné que x et tous les coefficients de cette expression sont positifs, on peut remarquer que seul le terme $\prod_{n=0}^k \left(\frac{\mu}{\mu_m} - n \right)$ possède un signe à déterminer. Par définition de la fonction $G(x)$, on sait que le premier terme de ce produit, soit $\frac{\mu}{\mu_m}$ est positif, mais inférieur à 1. Il s'ensuit que tous les termes de la forme $\frac{\mu}{\mu_m} - k$ sont négatifs et par conséquent, on peut aisément conclure que :

$$\prod_{n=0}^{k-1} \left(\frac{\mu}{\mu_m} - n \right) \begin{cases} \geq 0 & \text{si } k \text{ est impair} \\ \leq 0 & \text{si } k \text{ est pair} \end{cases}$$

$\forall x \in \mathbb{R}_+^N$ et cela permet d'aboutir à la conclusion recherchée:

$$\forall x \in \mathbb{R}_+^N \quad \frac{\partial^k G(x)}{\partial x_{i_1} \dots \partial x_{i_k}} \begin{cases} \geq 0 & \text{si } k \text{ est impair} \\ \leq 0 & \text{si } k \text{ est pair} \end{cases}$$

Toutes les propriétés qui viennent d'être démontrées prouvent que la fonction $G(x) = \sum_{m=1}^M \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}$ induit un modèle qui appartient à la famille GEV. Il s'agit maintenant d'établir le lien entre une fonction et un modèle et de prouver que la fonction $G(x)$ étudiée jusqu'à présent peut être associée au modèle cross-nested. Ce travail est l'objet de la section suivante.

3.4 La formule de probabilité

A ce stade de l'analyse, deux éléments théoriques ont été présentés séparément : le modèle cross-nested et la famille GEV. Il s'agit maintenant de faire le lien

entre ces deux entités. Pour cela, on utilise la théorie générale concernant la famille GEV qui nous dit notamment que pour un modèle induit par une fonction G (vérifiant les propriétés des fonctions GEV) on a la relation suivante :

$$P(i|C) = \frac{e^{V_i} \frac{\partial G}{\partial e^{V_i}}(e^{V_1}, \dots, e^{V_N})}{\mu G(e^{V_1}, \dots, e^{V_N})}$$

On va démontrer maintenant que la formule ci-dessus, dans laquelle la fonction G sera la fonction $G(x) = \sum_{m=1}^M \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}$ est équivalente à la formule (1):

$$\begin{aligned} P(i|C) &= \frac{e^{V_i} \frac{\partial G}{\partial e^{V_i}}(e^{V_1}, \dots, e^{V_N})}{\mu G(e^{V_1}, \dots, e^{V_N})} \\ &= \frac{e^{V_i} \sum_{m=1}^M \left[\frac{\mu}{\mu_m} \left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)^{\frac{\mu - \mu_m}{\mu_m}} \alpha_{im} \mu_m e^{V_i(\mu_m - 1)} \right]}{\mu \sum_{m=1}^M \left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)^{\frac{\mu}{\mu_m}}} \quad (2) \\ &= \frac{\sum_{m=1}^M \left[\left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)^{\frac{\mu - \mu_m}{\mu_m}} \alpha_{im} e^{V_i \mu_m} \right]}{\sum_{m=1}^M \left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)^{\frac{\mu}{\mu_m}}} \\ &= \sum_{m=1}^M \left[\frac{\left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)^{\frac{\mu}{\mu_m}} \alpha_{im} e^{V_i \mu_m}}{\sum_{l=1}^M \left(\left(\sum_{j \in C_l} \alpha_{jl} e^{V_j \mu_l} \right)^{\frac{\mu}{\mu_l}} \right) \left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)} \right] \end{aligned}$$

Or, on a :

$$V_{c_m} = \mu_m \tilde{V}_{C_m} + \ln \left(\sum_{j \in C_m} \alpha_{jm} e^{\tilde{V}_j \mu_m} \right)$$

donc :

$$e^{V_{c_m}} = e^{\mu_m \tilde{V}_{C_m}} \left(\sum_{j \in C_m} \alpha_{jm} e^{\tilde{V}_j \mu_m} \right)$$

$$\begin{aligned}
&= e^{\mu_m \tilde{V}_{C_m}} \left(\sum_{j \in C_m} \alpha_{jm} e^{(V_j - \tilde{V}_{C_m}) \mu_m} \right) \\
&= \sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m}
\end{aligned}$$

ainsi:

$$\begin{aligned}
P(i|C) &= \sum_{m=1}^M \left[\frac{e^{\frac{\mu}{\mu_m} V_{C_m}} \alpha_{im} e^{V_i \mu_m}}{\left(\sum_{l=1}^M e^{\frac{\mu}{\mu_l} V_{C_l}} \right) \left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)} \right] \\
&= \sum_{m=1}^M \left[\frac{e^{\frac{\mu}{\mu_m} V_{C_m}} \alpha_{im} e^{\tilde{V}_i \mu_m} e^{\tilde{V}_{C_m} \mu_m}}{\left(\sum_{l=1}^M e^{\frac{\mu}{\mu_l} V_{C_l}} \right) e^{\tilde{V}_{C_m} \mu_m} \left(\sum_{j \in C_m} \alpha_{jm} e^{\tilde{V}_j \mu_m} \right)} \right] \\
&= \sum_{m=1}^M \left[\frac{e^{\frac{\mu}{\mu_m} V_{C_m}}}{\sum_{l=1}^M e^{\frac{\mu}{\mu_l} V_{C_l}}} \frac{\alpha_{im} e^{\tilde{V}_i \mu_m}}{\sum_{j \in C_m} \alpha_{jm} e^{\tilde{V}_j \mu_m}} \right] \\
&= \sum_{m=1}^M P(C_m|C) P(i|C_m)
\end{aligned}$$

Tout ce développement prouve donc l'équivalence des deux formules et on peut ainsi conclure que le modèle cross-nested étendu appartient à la famille GEV et que la fonction qui lui est associée est la fonction

$$G(x) = \sum_{m=1}^M \left(\sum_{j \in C_m} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}$$

. A ce stade de l'étude du modèle cross-nested, il reste essentiellement à calculer l'expression analytique des dérivées de $P(i|C)$ par rapport à $V_j, j = 1, \dots, N, \mu_m, m = 1, \dots, M, \mu$ et $\alpha_{jm}, j = 1, \dots, N, m = 1, \dots, M$. Ce sera le contenu de la section suivante.

3.5 Les dérivées partielles

Le calcul des dérivées partielles est absolument nécessaire si l'on entend estimer les coefficients du modèle dans un cas pratique. Supposons que l'on possède les résultats d'un sondage au cours duquel S personnes ont été interrogées et ont effectué un choix parmi un certain nombre d'alternatives, chacune possédant des attributs bien définis. Prenons un exemple pour éclaircir cet aspect:

Supposons que S personnes ont été invitées à choisir entre la voiture et le train en connaissant le coût du trajet et le temps du trajet pour les deux modes de transport. La personne qui effectue le sondage relève la réponse de chaque personne, les attributs des alternatives disponibles (coût, temps,...), mais relèvent aussi d'autres attributs de la personne (sexe, salaire,...) qui selon lui peuvent avoir influencé le choix. A partir de ces données, on peut calculer pour un individu n la probabilité, en fonction de ses attributs, de choisir l'alternative i dans l'ensemble des alternatives C_n . On considère en effet que les alternatives disponibles ne sont peut-être pas les mêmes pour tous les individus. On peut donc calculer $P(i_n|C_n)$, où i_n représente l'alternative choisie par l'individu n . Evidemment, cette grandeur va dépendre de tous les coefficients inconnus. Pour l'estimation de ces coefficients, on utilise généralement une méthode de maximum de vraisemblance:

On écrit :

$$\mathcal{L}(V_j, \mu, \mu_m, \alpha_{jm}) = \prod_{n=1}^N P(i_n|C_n)$$

Toutefois, il est beaucoup plus facile de calculer le maximum du logarithme de cette fonction, qui est atteint pour les mêmes valeurs des coefficients que dans le cas du maximum de vraisemblance. On a alors:

$$\ln \mathcal{L}(V_j, \mu, \mu_m, \alpha_{jm}) = \sum_{n=1}^N \ln P(i_n|C_n)$$

La plupart des algorithmes de programmation non-linéaires nécessitent la connaissance des dérivées. Bien que l'implémentation d'un algorithme de résolution ne fasse pas partie des objectifs de ce projet, c'est néanmoins ce qui motive le calcul de l'expression de ces dérivées.

Tout d'abord, afin de ne pas répéter plus tard les mêmes calculs, on dérive par rapport à une variable générique. Soit donc θ un paramètre quelconque par rapport auquel on va dériver la fonction $\ln P(i_n|C_n)$. On obtient:

$$\frac{\partial}{\partial \theta} \ln P(i_n|C_n) = \frac{1}{P(i_n|C_n)} \frac{\partial}{\partial \theta} P(i_n|C_n)$$

N.B.: par la suite, la fonction $G(e^{V_1}, \dots, e^{V_N})$ sera simplement notée G .

Or:

$$\frac{\partial}{\partial \theta} P(i_n|C_n) = \frac{\partial}{\partial \theta} \left[\frac{e^{V_{i_n}} \frac{\partial G}{\partial e^{V_{i_n}}}}{\mu G} \right]$$

$$\begin{aligned}
&= \frac{\left[\frac{\partial e^{V_{i_n}}}{\partial \theta} \frac{\partial G}{\partial e^{V_{i_n}}} + e^{V_{i_n}} \frac{\partial^2 G}{\partial \theta \partial e^{V_{i_n}}} \right] \mu G - e^{V_{i_n}} \frac{\partial G}{\partial e^{V_{i_n}}} \left[\mu \frac{\partial G}{\partial \theta} + \frac{\partial \mu}{\partial \theta} G \right]}{\mu^2 G^2} \\
\Rightarrow & \frac{1}{P(i_n|C_n)} \frac{\partial}{\partial \theta} P(i_n|C_n) \\
&= \frac{\mu G \left[\frac{\partial e^{V_{i_n}}}{\partial \theta} \frac{\partial G}{\partial e^{V_{i_n}}} + e^{V_{i_n}} \frac{\partial^2 G}{\partial \theta \partial e^{V_{i_n}}} \right] \mu G - e^{V_{i_n}} \frac{\partial G}{\partial e^{V_{i_n}}} \left[\mu \frac{\partial G}{\partial \theta} + \frac{\partial \mu}{\partial \theta} G \right]}{e^{V_{i_n}} \frac{\partial G}{\partial e^{V_{i_n}}} \mu^2 G^2} \\
&= \frac{e^{V_{i_n}} \left[\left(\frac{\partial^2 G}{\partial \theta \partial e^{V_{i_n}}} \right) \mu G - \frac{\partial G}{\partial e^{V_{i_n}}} \frac{\partial G}{\partial \theta} \mu - \frac{\partial G}{\partial e^{V_{i_n}}} \frac{\partial \mu}{\partial \theta} G \right] + \frac{\partial G}{\partial e^{V_{i_n}}} \frac{\partial e^{V_{i_n}}}{\partial \theta} \mu G}{\mu G e^{V_{i_n}} \frac{\partial G}{\partial e^{V_{i_n}}}} \\
&= \frac{\frac{\partial^2 G}{\partial \theta \partial e^{V_{i_n}}} \frac{\partial G}{\partial e^{V_{i_n}}} - \frac{1}{G} \frac{\partial G}{\partial \theta} - \frac{1}{\mu} \frac{\partial \mu}{\partial \theta} + \frac{1}{e^{V_{i_n}}} \frac{\partial e^{V_{i_n}}}{\partial \theta}}{\frac{\partial G}{\partial e^{V_{i_n}}}} \\
&= \frac{\partial}{\partial \theta} \left[\ln \frac{\partial G}{\partial e^{V_{i_n}}} \right] - \frac{\partial}{\partial \theta} (\ln G) + \frac{\partial}{\partial \theta} (\ln e^{V_{i_n}}) - \frac{\partial}{\partial \theta} (\ln \mu) \\
&= \frac{\partial}{\partial \theta} \left[\ln \left(\frac{\partial G}{\partial e^{V_{i_n}}} \right) - \ln G \right] + \frac{\partial V_{i_n}}{\partial \theta} - \frac{\partial}{\partial \theta} (\ln \mu) \\
&= \frac{\partial}{\partial \theta} \left[\ln \left(\frac{1}{G} \frac{\partial G}{\partial e^{V_{i_n}}} \right) \right] + \frac{\partial V_{i_n}}{\partial \theta} - \frac{\partial}{\partial \theta} (\ln \mu) \\
&= \frac{\partial}{\partial \theta} \left[\ln \left(\frac{\partial}{\partial e^{V_{i_n}}} \ln G \right) \right] + \frac{\partial V_{i_n}}{\partial \theta} - \frac{\partial}{\partial \theta} (\ln \mu)
\end{aligned}$$

La démarche sera maintenant de remplacer le paramètre θ successivement par les paramètres V_k , μ , α_{kl} et finalement μ_k afin d'obtenir l'expression de la dérivée de $\ln P(i_n|C_n)$ par rapport à chacun de ces paramètres. Pour les calculs qui vont suivre, on se basera toujours sur la formule de départ suivante:

$$\frac{\partial}{\partial \theta} \left[\ln \frac{\partial G}{\partial e^{V_{i_n}}} \right] - \frac{\partial}{\partial \theta} (\ln G) + \frac{\partial}{\partial \theta} (\ln e^{V_{i_n}}) - \frac{\partial}{\partial \theta} (\ln \mu)$$

Tout d'abord, il est utile de calculer ce que vaut $\ln \frac{\partial G}{\partial e^{V_{i_n}}}$, car ce terme se retrouve dans chacune des dérivées. On a donc selon la formule (2):

$$\ln \frac{\partial G}{\partial e^{V_{i_n}}} = \ln \sum_{m=1}^M \left[\mu \left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)^{\frac{\mu - \mu_m}{\mu_m}} \alpha_{im} e^{V_{i_n} (\mu_m - 1)} \right]$$

Pour alléger l'écriture, on notera par la suite $\frac{\partial G}{\partial e^{V_{i_n}}} = \Delta$

Dérivée par rapport à V_k :

$$\begin{aligned} \frac{\partial}{\partial V_k} \ln P(i_n | C_n) &= \frac{1}{\Delta} \left(\sum_{m=1}^M \left(\frac{\mu}{\mu_m} - 1 \right) \left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)^{\frac{\mu - 2\mu_m}{\mu_m}} \alpha_{km} e^{V_k \mu_m} \mu_m e^{V_{i_n} (\mu_m - 1)} \mu \alpha_{i_n m} \right. \\ &\quad \left. + \mu \alpha_{i_n m} \left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)^{\frac{\mu - \mu_m}{\mu_m}} e^{V_{i_n} (\mu_m - 1)} (\mu_m - 1) \delta_{i_n k} \right) \\ &\quad - \frac{1}{G} \left[\sum_{m=1}^M \left[\frac{\mu}{\mu_m} \left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)^{\frac{\mu - \mu_m}{\mu_m}} \alpha_{km} e^{V_k \mu_m} \mu_m \right] \right] + \delta_{i_n k} \end{aligned}$$

Dérivée par rapport à μ :

$$\begin{aligned} \frac{\partial}{\partial \mu} \ln P(i_n | C_n) &= \frac{1}{\Delta} \left(\sum_{m=1}^M \left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)^{\frac{\mu - \mu_m}{\mu_m}} \ln \left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right) e^{V_{i_n} (\mu_m - 1)} \alpha_{i_n m} \frac{\mu}{\mu_m} \right. \\ &\quad \left. + \left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)^{\frac{\mu - \mu_m}{\mu_m}} e^{V_{i_n} (\mu_m - 1)} \alpha_{i_n m} \right) \\ &\quad - \frac{1}{G} \left[\sum_{m=1}^M \left[\left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right)^{\frac{\mu}{\mu_m}} \ln \left(\sum_{j \in C_m} \alpha_{jm} e^{V_j \mu_m} \right) \frac{1}{\mu_m} \right] \right] - \frac{1}{\mu} \end{aligned}$$

Dérivée par rapport à α_{kl} :

$$\begin{aligned} \frac{\partial}{\partial \alpha_{kl}} \ln P(i_n | C_n) &= \frac{1}{\Delta} \left(\left(\frac{\mu}{\mu_l} - 1 \right) \left(\sum_{j \in C_l} \alpha_{jl} e^{V_j \mu_l} \right)^{\frac{\mu - 2\mu_l}{\mu_l}} e^{V_k \mu_l} e^{V_{i_n} (\mu_l - 1)} \mu \alpha_{i_n l} \right. \\ &\quad \left. + \left(\sum_{j \in C_l} \alpha_{jl} e^{V_j \mu_l} \right)^{\frac{\mu - \mu_l}{\mu_l}} \mu e^{V_{i_n} (\mu_l - 1)} \delta_{i_n k} \right) \end{aligned}$$

$$-\frac{1}{G} \left[\frac{\mu}{\mu_l} \left(\sum_{j \in C_l} \alpha_{jl} e^{V_j \mu_l} \right)^{\frac{\mu - \mu_l}{\mu_l}} e^{V_k \mu_l} \right]$$

Dérivée par rapport à μ_l :

$$\begin{aligned} \frac{\partial}{\partial \mu_l} \ln P(i_n | C_n) &= \frac{1}{\Delta} (e^{\frac{\mu}{\mu_l} - 1})^{\ln(\sum_{j \in C_l} \alpha_{jl} e^{V_j \mu_l})} \\ &\left[\frac{-\mu}{\mu_l^2} \ln \left(\sum_{j \in C_l} \alpha_{jl} e^{V_j \mu_l} \right) + \left(\frac{\mu}{\mu_l} - 1 \right) \frac{\sum_{j \in C_l} \alpha_{jl} e^{V_j \mu_l} V_j}{\sum_{j \in C_l} \alpha_{jl} e^{V_j \mu_l}} \right] \mu \alpha_{i_n l} e^{V_{i_n} (\mu_l - 1)} \\ &+ \left(\sum_{j \in C_l} \alpha_{jl} e^{V_j \mu_l} \right)^{\frac{\mu - \mu_l}{\mu_l}} \mu \alpha_{i_n l} e^{V_{i_n} (\mu_l - 1)} V_{i_n} \\ &- \frac{1}{G} (e^{\frac{\mu}{\mu_l}})^{\ln(\sum_{j \in C_l} \alpha_{jl} e^{V_j \mu_l})} \\ &\left[\frac{-\mu}{\mu_l^2} \ln \left(\sum_{j \in C_l} \alpha_{jl} e^{V_j \mu_l} \right) + \frac{\mu}{\mu_l} \frac{\sum_{j \in C_l} \alpha_{jl} e^{V_j \mu_l} V_j}{\sum_{j \in C_l} \alpha_{jl} e^{V_j \mu_l}} \right] \end{aligned}$$

Tous les calculs présentés dans cette section posent les bases du modèle cross-nested étendu et prouvent l'appartenance de ce modèle à la famille GEV. Ce modèle est une extension du modèle proposé par Vovsha et on peut légitimement espérer que le calcul des dérivées partielles permettra l'implémentation d'un algorithme capable d'estimer un modèle cross-nested étendu et donc certainement de démontrer la puissance de ce modèle dans beaucoup de cas concrets.

4 Etude d'un problème de choix de mode de transport

4.1 Présentation du problème

Cette seconde partie du projet consiste en une étude pratique d'un problème concret de choix de mode de transport. Lors d'un sondage 1139 personnes ont été invitées à déclarer leur préférence concernant le choix d'un mode de

transport. Chaque personne a été interrogée 9 fois, mais chaque fois dans une situation hypothétique différente. Parallèlement, les personnes qui ont effectué le sondage ont demandé aux personnes interrogées leur salaire, leur âge et d'autres caractéristiques susceptibles selon eux d'influencer le choix à effectuer. Ces caractéristiques seront détaillées ultérieurement. A partir de ces données, il s'agit de calibrer un modèle multinomial à l'aide du logiciel HieLoW, c'est-à-dire de déterminer quelles caractéristiques influencent réellement le choix de la personne. Lors du sondage, les personnes devaient choisir entre la voiture, le train(CFF) et Swissmetro. Le jeu de données nous a été aimablement fourni par le professeur Axhausen de l'ETHZ.

4.2 Calibration du modèle

4.2.1 Idée générale

L'idée générale de la calibration du modèle à partir du jeu de données est de débiter avec un modèle extrêmement simple, puis peu à peu d'intégrer les caractéristiques à disposition et d'observer à chaque étape si la modification apportée améliore le modèle. Toutefois, il existe plusieurs manières d'intégrer une caractéristique dans le modèle existant. Par exemple, si l'on décide d'intégrer le paramètre *temps du trajet*, on peut se demander si la partie déterministe de la fonction d'utilité dépend linéairement du temps ou du logarithme du temps. Ainsi, pour une même caractéristique, il peut s'avérer nécessaire de tester plusieurs modèles afin de déterminer la façon optimale d'intégrer la caractéristique au modèle.

Lorsque toutes les caractéristiques sont intégrées, il faut encore "nettoyer" le modèle, c'est-à-dire éliminer les caractéristiques qui influencent de façon négligeable le choix. Afin de déterminer si une caractéristique peut être éliminée du modèle, on se base sur certains tests statistiques que fournit HieLoW lors du calcul des coefficients β_t de la fonction $V_k = \sum_t \beta_t x_{kt}$. On élimine la caractéristique x_{kt} si l'hypothèse nulle $\beta_t = 0$ ne peut pas être rejetée à un niveau de 5%. Dans une autre optique, il s'agit de déceler si certaines caractéristiques sont très fortement corrélées, c'est-à-dire si $\beta_t = \beta_s$. Pour cela, on utilise également un test statistique permettant de décider si l'hypothèse nulle $\beta_t = \beta_s$ peut être rejetée ou pas. Si c'est le cas, on peut poser les coefficients β_t et β_s égaux dans le modèle.

Le premier modèle estimé considère uniquement deux caractéristiques qui sont le temps et le coût. On peut donc pour chaque alternative k écrire:

$$V_k = \text{cout} * \text{cout}[k] + \text{temps} * \text{temps}[k]$$

où coût[k] est le coût du trajet avec le mode transport k et temps[k] est le temps de trajet avec le mode de transport k. Ce modèle très simple sert de point de départ pour l'élaboration de modèles plus sophistiqués. Il s'agit maintenant d'augmenter la complexité de ce modèle initial en lui ajoutant de nouvelles caractéristiques.

4.2.2 Le jeu de données

Avant de débiter la calibration du modèle, il convient par précaution de vérifier si les données sont cohérentes. Dans ce cas précis, cette vérification s'est avérée utile, car 640 lignes de données sur les 1139*9=10251 lignes du jeu comportaient des valeurs aberrantes. On distingue deux types d'aberrations: d'une part des personnes qui ne possèdent pas de permis de conduire choisissent la voiture et d'autre part des personnes ont choisi deux modes de transport. Grâce à un procédé d'exclusion, HieLoW élimine ces données "corrompues" et n'en tient donc pas compte lors de l'estimation des coefficients. Le jeu de données comportait pour chaque observation deux types de caractéristiques: certaines liées aux alternatives et d'autres relatives aux personnes. D'autre part, il existe deux types très différents de caractéristiques: les caractéristiques continues et les caractéristiques discrètes ou à niveaux. Le temps et le coût sont typiquement des caractéristiques continues car elles peuvent théoriquement prendre n'importe quelle valeur positive. L'âge par contre est dans ce modèle une caractéristique à niveaux car lors de l'enquête, on a "découpé" l'âge en 5 classes d'âge et on a simplement demandé aux personnes à laquelle de ces classes ils appartenaient. Dans le tableau ci-dessous, ont été représentées toutes les caractéristiques dont on dispose pour calibrer le modèle. Pour les caractéristiques à niveaux, on a indiqué les différents niveaux, à quoi ils correspondaient et pour chaque niveau, le coefficient qui le représente dans le modèle. De plus, pour certaines caractéristiques, on a indiqué quelle proportion de l'échantillon appartenait à chacun des niveaux

de la caractéristique.

Caractéristique	Niveau	Signification	Coefficient	Proportion
Temps		temps du trajet	temps	
Cout		cout du trajet	cout	
Frequence		frequence des trains	frequence	
Somme			somme	
Confort	1	pas de places assises	confort1	
	2	places assises	confort2	
But	1	pendulaire	but1	14,26%
	2	achats	but2	3,03%
	3	voyage d'affaires	but3	47,42%
	4	loisirs	but4	35,28%
Classe de wagon	1	voyage hab. en 1ère	classe1	
	2	voyage hab. en 2ème	classe2	
Billet	1	aller-retour avec demi	billet1	
	2	?	billet2	
	3	aller-retour sans demi	billet3	
	4	?	billet4	
	5	?	billet5	
	6	abonnement general	billet6	
	7	AG junior ou senior	billet7	
	8	?	billet8	
	10	?	billet10	
Qui paie le trajet	1	soi-meme	quipaie1	
	2	entreprise	quipaie2	
	3	moitie-moitie	quipaie3	
Bagages	1	pas de bagages	bagages1	
	2	bagages	bagages2	
Age	1	jusqu'a 24 ans	age1	3,66%
	2	24-39 ans	age2	30,24%
	3	39-54 ans	age3	38,01%
	4	54-65 ans	age4	20,9%
	5	plus de 65 ans	age5	7,18%
nombre de personnes	1	voyage seul	1personne	
	2	voyage a deux	2personnes	
	3	voyage a trois	3personnes	
	4	voyage a quatre	4personnes	
sexe	1	homme	masculin	78,45%
	2	femme	feminin	21,55%
salaire	1	moins de 50000 francs	salaire1	11,96%
	2	50000-100000 francs	salaire2	36,93%
	3	plus de 100000 francs	salaire3	41,18%
	4	non-revele	salaire4	9,93%
train	0	interroge dans la voiture	train0	71,62%
	1	interroge dans le train	train1	28,38%

Voici quelques précisions qui permettent de bien pouvoir comprendre les caractéristiques, leur signification et leur codification:

1) Les deux caractéristiques qui indiquent le numéro du canton de départ et le numéro du canton de destination n'apparaissent pas dans le tableau, car d'une part elles possèdent de nombreux niveaux et d'autre part on ne possède pas la liste qui à chaque numéro associe un canton.

2) Parfois, la signification des différents niveaux d'une caractéristique nous est parvenue de manière incomplète, ce qui explique les points d'interrogation dans la troisième colonne du tableau.

3) La variable Somme indique la somme du temps qu'il faut jusqu'à la gare de départ et du temps de la gare d'arrivée jusqu'au but du trajet. Pour la voiture, cette caractéristique est nulle.

4) La fréquence représente la durée de l'intervalle de temps entre deux trains consécutifs. Pour la voiture, la fréquence est supposée égale à zéro.

5) Le confort ne concerne que Swissmetro, car les deux autres modes de transport possèdent toujours des places assises.

Voici encore quelques informations supplémentaires sur les variables dites continues:

	Voiture	Swissmetro	CFE
Cout maximum[CHF]	286	422	520
Cout minimumm[CHF]	5	5	7
Cout moyenm[CHF]	83.997	101.5	91.16
Temps maximum[Min]	470	796	1049
Temps minimumm[Min]	23	8	20
Temps moyenm[Min]	149.13	91.478	165.424
Frequence maximumm[Min]	0	120	120
Frequence minimumm[Min]	0	10	10
Frequence moyennem[Min]	0	20.768	52.638

4.2.3 Le codage des variables

Leur modélisation ne présente aucune difficulté, mais illustrons-la en décrivant le modèle de base:

$$V_{voiture} = cout * cout[voiture] + temps * temps[voiture]$$

$$\begin{aligned}
V_{train} &= \text{cout} * \text{cout}[train] + \text{temps} * \text{temps}[train] \\
V_{swissmetro} &= \text{cout} * \text{cout}[Swissmetro] + \text{temps} * \text{temps}[Swissmetro]
\end{aligned}$$

Illustrons maintenant comment il est possible de modéliser une variable à niveaux comme l'âge. Il convient tout d'abord de préciser que seule la différence entre deux coefficients est représentative. De ce fait, on peut choisir un niveau de base, c'est-à-dire une alternative, et poser le coefficient relatif à cette alternative égal à zéro. Ensuite, la comparaison des coefficients permet de mesurer l'impact d'une caractéristique dans le choix d'un individu. Pour l'ensemble de la modélisation, la convention a été de poser la voiture comme niveau de base pour toutes les variables à niveaux.

$$\begin{aligned}
V_{voiture} &= \text{cout} * \text{cout}[voiture] + \text{temps} * \text{temps}[voiture] \\
V_{train} &= \text{cout} * \text{cout}[train] + \text{temps} * \text{temps}[train] + \text{age1}_{CFE} * (\text{age} = 1) \\
&\quad + \text{age2}_{CFE} * (\text{age} = 2) + \text{age3}_{CFE} * (\text{age} = 3) \\
&\quad + \text{age4}_{CFE} * (\text{age} = 4) + \text{age5}_{CFE} * (\text{age} = 5) \\
V_{swissmetro} &= \text{cout} * \text{cout}[Swissmetro] + \text{temps} * \text{temps}[Swissmetro] \\
&\quad + \text{age1}_{Swissmetro} * (\text{age} = 1) + \text{age2}_{Swissmetro} * (\text{age} = 2) \\
&\quad + \text{age3}_{Swissmetro} * (\text{age} = 3) + \text{age4}_{Swissmetro} * (\text{age} = 4) \\
&\quad + \text{age5}_{Swissmetro} * (\text{age} = 5)
\end{aligned}$$

Le logiciel HieLoW va alors estimer, à partir des données, la valeur des 12 coefficients qui maximise la vraisemblance. Si le coefficient est significativement différent de zéro (cf. test statistique), alors on peut interpréter son signe. Si le coefficient est positif, alors la caractéristique qu'il représente augmentera l'utilité en grandissant. Inversement, si le coefficient est négatif, alors la caractéristique qu'il représente diminuera l'utilité en grandissant. Cet aspect sera approfondi lors de l'analyse des résultats, mais on peut déjà faire quelques observations par rapport à l'exemple ci-dessus. On peut logiquement supposer que *cout* sera négatif. En effet, plus le coût du trajet est élevé, plus l'utilité d'un mode de transport diminue. Il en va de même avec le coefficient *temps*. Supposons que l'estimation nous fournisse $\text{age5}_{Swissmetro} \geq 0$. Cela signifie que les personnes de plus de 65 ans préfèrent Swissmetro à la voiture (niveau de base), indépendamment des autres caractéristiques. Ainsi, au travers de chaque caractéristique, il sera possible de tirer des informations quant

au comportement des personnes, mais également de prévoir le comportement futur d'une personne. En effet, il suffira de connaître les caractéristiques de cette personne pour calculer les fonctions d'utilité de chaque mode de transport et pour ensuite déduire la probabilité qu'une personne choisisse tel mode de transport. Il convient toutefois de se méfier de ce genre de raisonnement, car l'on considère ici uniquement la partie déterministe des fonctions d'utilité. L'utilité comporte aussi une composante aléatoire et de ce fait, toutes les conclusions qui seront établies le seront avec une certaine probabilité. Ce ne seront en aucun cas des vérités absolues.

Malheureusement, le codage des niveaux des variables ne nous est parvenu que partiellement. Toutefois, cette déficience dans le codage ne perturbe presque pas la calibration du modèle, mais ne permet pas une analyse détaillée de certaines caractéristiques. En effet, il est inutile de savoir qu'un coefficient est positif si l'on ignore la signification du niveau de la caractéristique qu'il représente. De ce fait, l'analyse de certains résultats sera quelque peu délicate et nous nous concentrerons donc sur les caractéristiques correctement codifiées et dont l'interprétation ne laisse planer aucun doute.

Maintenant que le problème est un peu plus clairement présenté, il est temps de discuter de la calibration du modèle à partir du modèle de base décrit à la page 19.

4.2.4 Les modèles successifs

Pour aboutir au modèle final, il a fallu calibrer de nombreux modèles successifs et effectuer une multitude d'essais. En annexe de ce document sont d'ailleurs présentés les principaux modèles intermédiaires qui ont été calibrés et qui ont permis de faire progresser significativement le modèle. On y trouve notamment la valeur de la log-vraisemblance, la valeur des coefficients qui ont été estimés, ainsi que l'erreur standard asymptotique qui leur est associée et qui permet de décider si un coefficient est significativement différent de zéro. On trouve également la formulation analytique des fonctions d'utilité des trois modes de transport, ce qui permet d'observer les améliorations apportées au modèle étape après étape. À noter aussi que le modèle 4 est le modèle obtenu avant le nettoyage final, c'est-à-dire avant l'élimination des coefficients non significatifs. Voici tout de même un bref descriptif des différents modèles qui ont permis d'aboutir au modèle final avant nettoyage. On précise pour chaque modèle les caractéristiques ajoutées par rapport au modèle précédent:

1) modèle 0 : le coût et le temps sont les seules caractéristiques prises en

considération.

2) modèle 1 : la somme, la fréquence, les bagages et le confort.

3) modèle 2 : le salaire, la classe de wagon, qui paie le trajet et la caractéristique train qui indique si la personne a été interrogée dans la voiture ou dans le train.

4) modèle 3 : l'âge, le sexe, le nombre de personnes dans le groupe de la personne interrogée.

5) modèle 4 : c'est le modèle final avant nettoyage. On a donc rajouté le but, le type de billet, les numéros des cantons de départ et de destination.

Soulignons ici la stratégie utilisée pour calibrer le modèle: aucune simplification n'a été entreprise entre le modèle de base et le modèle 4. Les tests statistiques n'ont pas été pris en considération et ceci afin de ne pas tirer de conclusions trop hâtives. En effet, l'introduction de nouvelles caractéristiques ultérieurement peut rendre significatif un coefficient qui ne l'était pas avant. De ce fait, on peut simplement signaler qu'à chaque étape de la calibration, la valeur de la log-vraisemblance augmente et donc que le nouveau modèle traduit mieux la réalité. Il s'agit donc maintenant de nettoyer le modèle obtenu en utilisant les tests statistiques fournis par le logiciel HieLoW. Ce nettoyage se fait de manière assez systématique. Par exemple, on tente de retirer un coefficient du modèle et on compare le résultat obtenu par rapport au modèle précédent. Si le résultat n'est pas significativement différent, alors on garde le modèle modifié. Evidemment, ce procédé n'est utilisé que pour les coefficients pour lesquels l'hypothèse *le coefficient est différent de zéro* ne peut pas être rejetée. Ainsi, étape après étape, on parvient au modèle final qui est décrit dans la section 4.3.

4.2.5 Le nettoyage du modèle

Le nettoyage du modèle s'est concentré sur l'élimination des coefficients non significatifs. Dans le modèle avant nettoyage, plus de la moitié des coefficients estimés se sont avérés non significatifs. Toutefois, en les supprimant tous du modèle, on obtient un modèle significativement différent du modèle initial. Ce paradoxe s'explique assez simplement: les coefficients éliminés sont individuellement non significatifs, mais pris ensemble ils influencent significativement le modèle. De ce fait, la démarche s'est axée sur l'élimination de certains coefficients, essentiellement ceux qui apparaissaient comme moins intéressants. Par exemple, les coefficients non significatifs des caractéristiques déficientes (cf. page 22) ont tous été enlevés du modèle, car de toute façon,

il aurait été impossible de les interpréter. Quant au choix des autres coefficients à éliminer, il est purement subjectif et certainement que quelqu'un d'autre aurait effectué d'autres choix. La calibration d'un modèle n'est de loin pas une science exacte et fait souvent appel à la clairvoyance et à l'intuition de la personne qui calibre. De ce fait, il existe beaucoup de modèles cohérents différents capables de modéliser une même situation. Le modèle qui est présenté dans la section suivante est certainement l'un de ces modèles et l'analyse de la section 4.3 permettra d'appréhender la réalité à la simple lecture des résultats de l'estimation.

Avant d'aborder la section consacrée aux résultats, il convient d'ouvrir une parenthèse concernant les problèmes rencontrés au cours de la calibration. En effet, certaines caractéristiques se sont avérées extrêmement corrélées, sans que cela puisse être expliqué raisonnablement. Par exemple, l'âge et le but étaient fortement corrélées et de toute évidence il apparaît difficile de justifier cette corrélation. Pour éliminer ces corrélations qui perturbait HieLoW lors de l'estimation des coefficients, il a fallu éliminer certains niveaux des caractéristiques concernées et concéder ainsi une perte d'information. Outre ces problèmes, il est apparu dans le jeu de données des valeurs aberrantes (un niveau quipaie0) difficilement interprétables, car sans codification. La calibration s'est donc attachée à trouver le juste milieu entre un modèle qui traduit le plus précisément possible la réalité et un modèle le moins influencé possible par les incohérences du jeu de données.

4.3 Résultats

Dans ce paragraphe sont simplement présentés sous forme synthétisée et plus agréable à consulter les principaux résultats de l'estimation du modèle, mais également les résultats obtenus lors de l'estimation d'un modèle nested n'intégrant que le coût et le temps.

4.3.1 Les fonctions d'utilité

Voici les trois fonctions d'utilité qu'a permis d'établir la calibration d'un modèle multinomial, puis le nettoyage du modèle:

$$V_{voiture} = cout * cout[voiture] + temps * temps[voiture]$$

$$\begin{aligned}
V_{train} = & +somme * somme[voiture] \\
& cout * cout[CFF] + temps * temps[CFF] \\
& +somme * somme[CFF] + but2_{CFF} * but2 \\
& +but3_{CFF} * but3 + but4_{CFF} * but4 + billet2_{CFF} * billet2 \\
& +billet3_{CFF} * billet3 + billet5_{CFF} * billet5 \\
& +billet8_{CFF} * billet8 + billet10_{CFF} * billet10 \\
& +1pers_{CFF} * 1pers + 2pers_{CFF} * 2pers \\
& +3pers_{CFF} * 3pers + 4pers_{CFF} * 4pers \\
& +age1_{CFF} * age1 + age2_{CFF} * age2 + age3_{CFF} * age3 \\
& +age4_{CFF} * age4 + salaire1_{CFF} * salaire1 \\
& +salaire2_{CFF} * salaire2 + salaire3_{CFF} * salaire3 \\
& +bagages1_{CFF} * bagages1 + dep17_{CFF} * dep17 \\
& +dest22_{CFF} * dest22 + frequence * frequence[CFF] \\
& +masculin_{CFF} * sexe1 + feminin_{CFF} * sexe2 \\
V_{Swissmetro} = & cout * cout[Swissmetro] + temps * temps[Swissmetro] \\
& +somme * somme[Swissmetro] + but2_{Swissmetro} * but2 \\
& +but3_{Swissmetro} * but3 + but4_{Swissmetro} * but4 \\
& +billet2_{Swissmetro} * billet2 + billet4_{Swissmetro} * billet4 \\
& +billet4_{Swissmetro} * billet6 + billet7_{Swissmetro} * billet7 \\
& +2pers_{Swissmetro} * 2pers + 4pers_{Swissmetro} * 4pers \\
& +age1_{Swissmetro} * age1 + age2_{Swissmetro} * age2 \\
& +age3_{Swissmetro} * age3 + age4_{Swissmetro} * age4 \\
& +classe2_{Swissmetro} * classewagon + bagages1_{Swissmetro} * bagages1 \\
& +dep25_{Swissmetro} * dep25 + dest17_{Swissmetro} * dest17 \\
& +masculin_{Swissmetro} * sexe1 + feminin_{Swissmetro} * sexe2 \\
& +train0_{Swissmetro} * train0 + frequence * frequence[Swissmetro]
\end{aligned}$$

Ces expressions méritent quelques éclaircissements afin de mieux pouvoir comprendre les fonctions d'utilité:

1) La variable billet5(par exemple) ne prend que les deux valeurs 0 et 1. En effet, il faut comprendre $billet5 = (billet==5)$, c'est-à-dire que si la personne interrogée a pris un billet du type 5 alors la variable billet5 vaut 1, tandis que dans le cas contraire elle vaut 0. Ainsi pour un individu donné,

seule une des variables *billetX* vaut 1, toutes les autres étant égales à zéro.

2) Les deux caractéristiques *numéro du canton de départ* et *numéro du canton de destination* sont considérées dans ce modèle comme des caractéristiques à 5 niveaux, car presque toutes les observations donne un résultat dans lequel on retrouve ces 5 cantons. Malheureusement, comme cela a déjà été précisé auparavant, la liste qui fait correspondre à chaque numéro un canton n'est pas disponible.

4.3.2 La valeur des coefficients

Voici la valeur des coefficients du modèle final avec pour chacun d'eux la valeur de test t qui sert à déterminer si le coefficient est significatif. Il s'agit évidemment d'une estimation réalisée à partir du logiciel HieLoW. Il convient de souligner que la fonction d'utilité n'a pas d'unités, mais que les coefficients possèdent une unité. Par exemple, l'unité du coefficient *cout* est $\frac{1}{min}$. On ne peut toutefois les interpréter uniquement dans le cadre de comparaisons et ce travail est l'objet de la section 4.3.

<i>cout</i>	=	-0.01672	-29.23
<i>temps</i>	=	-0.01611	-29.51
<i>somme</i>	=	-0.00344	-5.047
<i>frequence</i>	=	-0.00814	-10.36
<i>but2_{CFE}</i>	=	+0.1225	+0.413
<i>but3_{CFE}</i>	=	-0.01737	-0.102
<i>but4_{CFE}</i>	=	-0.2424	-1.404
<i>but2_{Swissmetro}</i>	=	-0.5906	-3.669
<i>but3_{Swissmetro}</i>	=	-0.1261	-1.513
<i>but4_{Swissmetro}</i>	=	-0.6517	-7.428
<i>bagages1_{CFE}</i>	=	-0.8603	-7.448
<i>bagages1_{Swissmetro}</i>	=	-0.4697	-8.444
<i>1personne_{CFE}</i>	=	-0.1446	-0.322
<i>2personnes_{CFE}</i>	=	+0.6067	+1.362
<i>3personnes_{CFE}</i>	=	-0.4926	-0.93
<i>4personnes_{CFE}</i>	=	+0.0572	+0.108

<i>2personnes</i> _{Swissmetro}	= +0.1604	+2.556
<i>4personnes</i> _{Swissmetro}	= -0.4041	-2.784
<i>age1</i> _{CFE}	= -0.274	-0.842
<i>age2</i> _{CFE}	= +0.887	+4.011
<i>age3</i> _{CFE}	= +0.8927	+4.038
<i>age4</i> _{CFE}	= +1.264	+5.836
<i>age1</i> _{Swissmetro}	= +0.2941	+1.619
<i>age2</i> _{Swissmetro}	= +0.8321	+7.535
<i>age3</i> _{Swissmetro}	= +0.7252	+6.732
<i>age4</i> _{Swissmetro}	= +0.5092	+4.6
<i>salaire1</i> _{CFE}	= -0.0268	-0.151
<i>salaire2</i> _{CFE}	= -0.2769	-1.778
<i>salaire3</i> _{CFE}	= -0.4393	-2.673
<i>billet2</i> _{CFE}	= +1.46	+3.568
<i>billet3</i> _{CFE}	= -0.4393	-3.493
<i>billet5</i> _{CFE}	= +1.175	+4.485
<i>billet8</i> _{CFE}	= +0.8689	+3.402
<i>billet10</i> _{CFE}	= +1.379	+3.931
<i>billet2</i> _{Swissmetro}	= +1.164	+3.795
<i>billet4</i> _{Swissmetro}	= +1.515	+3.18
<i>billet6</i> _{Swissmetro}	= -0.4912	-5.209
<i>billet7</i> _{Swissmetro}	= -0.5635	-2.665
<i>dep17</i> _{CFE}	= -0.6044	-2.698
<i>dest22</i> _{CFE}	= +0.3863	+2.925
<i>dep25</i> _{Swissmetro}	= -0.1878	-2.915
<i>dest17</i> _{Swissmetro}	= -0.3462	-4.229
<i>classe2</i> _{Swissmetro}	= -0.3475	-6.248
<i>train0</i> _{Swissmetro}	= -0.612	-7.155
<i>masculin</i> _{CFE}	= -1.779	-3.643
<i>feminin</i> _{CFE}	= -1.884	-3.76
<i>masculin</i> _{Swissmetro}	= +0.3939	+2.7

$$feminin_{Swissmetro} = +0.3766 + 2.493$$

On peut encore signaler que la valeur de la log-vraisemblance pour le modèle final est de -6249.41 avec 48 coefficients estimés, tandis que la valeur de la log-vraisemblance pour le modèle avant nettoyage est de -6190 avec 91 coefficients estimés, ce qui justifie le fait de déclarer les deux modèles non significativement différents et de pouvoir ainsi travailler sur un modèle avec 43 coefficients de moins. Comme cela a déjà été précisé auparavant, on trouve en annexe les résultats de quelques modèles intermédiaires entre le modèle n'intégrant que le coût et le temps et le modèle avant nettoyage.

4.3.3 L'estimation d'un modèle nested

Afin de pouvoir comparer le modèle multinomial et le modèle nested, on a estimé deux modèles nested simples, c'est-à-dire n'intégrant que les caractéristiques coût et temps. Dans les deux modèles, les alternatives ont été réparties dans deux nids. Le premier contient la voiture et le train, car ce sont des moyens de transport existants et le deuxième contient uniquement Swissmetro qui n'existe pas encore. Lors de l'estimation d'un modèle nested, on estime non seulement des coefficients relatifs aux caractéristiques, mais pour chaque nid m on estime un coefficient θ_m . Si pour ce coefficient l'hypothèse nulle *le coefficient est égal à 1* est rejetée, alors cela signifie que l'introduction de ce nid est justifiée et qu'elle a amélioré le modèle. Le premier modèle testé part de l'hypothèse que les coefficients θ_1 et θ_2 sont égaux. Le résultat obtenu est très intéressant: le coefficient est significativement différent de 1. Cela signifie que le fait qu'un mode de transport existe ou non influence l'individu dans son choix. Le second modèle testé ne fait aucune hypothèse sur les coefficients θ_1 et θ_2 . Là encore, le résultat est intéressant: non seulement les deux coefficients sont significativement différents de 1, mais ils sont aussi significativement différents l'un de l'autre. En conclusion, on peut dire que dans l'exemple qui nous concerne, pourtant assez simple puisqu'il ne concerne que trois alternatives, l'estimation d'un modèle cross-nested est significativement différente (et surtout meilleure) que l'estimation d'un modèle multinomial. Ce résultat nous conforte dans l'idée que l'estimation d'un modèle cross-nested étendu peut fournir des résultats qui reflètent avec une très grande précision la réalité et qui de ce fait sont beaucoup plus fiables.

4.4 Analyse des résultats

L'analyse des résultats va se subdiviser en 4 parties: une brève étude des coefficients non significatifs, une interprétation de certains coefficients du modèle final, un éclairage sur ce que l'on appelle la valeur du temps et finalement quelques applications qui montrent la façon dont on peut utiliser les résultats. Remarquons que l'interprétation d'un coefficient ne peut se faire que par rapport à un autre coefficient. Ce sera souvent la voiture, car les coefficients relatifs à la voiture (niveau de base) sont tous nuls et donc le signe d'un coefficient permet déjà de tirer certaines conclusions. A chaque fois, les comparaisons se font indépendamment des autres caractéristiques. Les bases étant posées, nous pouvons maintenant commencer l'analyse.

4.4.1 Les coefficients non significatifs

Entre le modèle avant nettoyage et le modèle final, 43 coefficients ont été éliminés parce qu'ils n'étaient pas suffisamment significatifs. Il est illusoire de vouloir trouver une explication satisfaisante pour chacun d'eux, mais on peut toutefois relever certains aspects fort intéressants. Ainsi, le fait que le trajet soit payé par l'entreprise ou par le voyageur lui-même n'influence pas le choix de ce dernier. Cela peut paraître étonnant, car on aurait pu s'attendre à ce que les personnes qui ne paient pas le voyage optent plus facilement pour Swissmetro, qui est un mode de transport rapide, mais cher. Cependant, ce pourrait être aussi un indicateur d'une certaine honnêteté des personnes interrogées qui se refusent à profiter des avantages qui leur sont offerts. On peut également remarquer que le niveau de salaire de la personne n'influence que faiblement son choix. En effet, la plupart des coefficients relatifs au salaire ne sont pas significatifs, mais certains ont tout de même été incorporés dans le modèle final afin de rendre plus claires les applications qui suivront. Toutefois, ce résultat nous conforte dans l'idée que l'aspect financier du trajet n'est pas fondamental pour les gens. On verra par la suite que l'aspect temporel joue lui un rôle déterminant. Enfin, on peut signaler que les personnes interrogées dans le train n'accordent pas spécialement leur préférence à ce mode de transport et ne lui voue pas une admiration sans bornes.

4.4.2 Les coefficients significatifs

Le modèle final comporte 48 coefficients. Cette analyse va s'attacher à parcourir les caractéristiques les plus importantes et à essayer d'interpréter les coefficients qui y sont relatifs. Il est évidemment impossible d'être exhaustif dans ce parcours, mais nous essayerons tout de même de souligner les principales enseignements que fournit la valeur des coefficients estimés.

le sexe: $masculin_{Swissmetro}$ et $feminin_{Swissmetro}$ sont tous les deux positifs ce qui tend à prouver l'attrait qu'exerce Swissmetro en général. Sans doute que l'impact de la technologie nouvelle fascine beaucoup de monde. Toutefois, on peut se poser la question: les gens garderont-ils le même enthousiasme lorsque Swissmetro sera réalisé ou bien l'attrait laissera-t-il place à une certaine crainte de voyager dans un tunnel sur un coussin magnétique? La question reste ouverte et seul l'avenir est susceptible de nous fournir une solution... Inversement, les deux coefficients relatifs au train sont fortement négatif et déjà maintenant une tendance se dégage pour reléguer le train bien souvent en dernière position. Les applications de la section 4.4.4 confirmeront cette tendance.

les bagages: l'estimation fournit les résultats auxquels on pouvait s'attendre. En effet, si une personne voyage avec des bagages, alors elle affiche une nette préférence pour la voiture. On remarque cela à la négativité des coefficients $bagages1_{CFE}$ et $bagages1_{Swissmetro}$. Cela est assez normal, car les bagages sont généralement assez encombrants dans un train surtout au moment de changer de train, tandis qu'une fois déposés dans le coffre de la voiture, ils ne posent plus aucun tracas.

l'âge: la valeur des coefficients relatifs à l'âge est très étonnante puisque toutes les classes d'âge sauf une préfèrent le train et Swissmetro à la voiture. Ce résultat est partialement en contradiction avec le résultat obtenu avec les coefficients relatifs au sexe. En fait, seules les personnes âgées de moins de 24 ans préfèrent la voiture au train. Sans doute le train leur apparaît-il comme ringard à un âge où on vient d'obtenir son permis de conduire.

le but: beaucoup de coefficients relatifs au but du voyage ne sont pas significativement différents de zéro, mais apparaissent tout de même dans le modèle. Cependant, on remarque que le train est assez apprécié pour aller faire ses achats, mais que par contre la voiture est privilégiée pour les loisirs.

Par contre, et cela est étonnant, Swissmetro n'est pas un moyen de transport adéquat pour les voyages d'affaires.

la fréquence: le résultat obtenu est des plus logiques. En effet, si la fréquence augmente (il faut comprendre si la durée entre deux trains augmente) alors l'utilité du train et de Swissmetro diminue car le coefficient fréquence est négatif.

la somme: les mêmes constatations que ci-dessus peuvent être effectuées.

4.4.3 La valeur du temps

La valeur du temps est certainement l'un des résultats les plus importants que l'on peut déduire de l'estimation des coefficients. Donnons tout d'abord une définition de la valeur du temps: la valeur du temps est le prix qu'une personne est prête à payer pour gagner une minute de trajet. Cette valeur est calculée en divisant le coefficient relatif au temps (unité: $\frac{1}{min}$) par le coefficient relatif au coût (unité: $\frac{1}{fr}$). Dans l'exemple calibré, on trouve pour la valeur du temps 96.3 centimes. Ce chiffre semble très élevé. Sans doute que l'attrait pour Swissmetro explique ce résultat. En effet, Swissmetro est l'exemple type du mode de transport qui propose de payer plus cher pour gagner du temps. On peut signaler au passage que dans l'ensemble du sondage on devine un puissant attrait pour Swissmetro qui quelque part fausse légèrement les résultats. Remarquons enfin que le signe des coefficients relatifs au coût et au temps est conforme aux attentes: plus le temps augmente (respectivement le coût) plus l'utilité diminue. C'est la logique des choses.

Toutefois, le résultat obtenu permet de tirer un enseignement important: la durée du trajet est primordiale pour les gens. Ils semblent accorder beaucoup plus d'importance à la durée qu'au coût du trajet. S'il fallait trouver une explication rationnelle, ce serait peut-être celle-ci: instinctivement, les gens associent à Swissmetro un gain de temps, mais parallèlement "oublient" le coût supplémentaire qui en découle. On pourrait même soupçonner qu'au moment d'être interrogés, les gens s'imaginaient que le but du questionnaire était de sonder leur avis sur la question du gain de temps à nouveau à cause de l'association Swissmetro/gain de temps.

4.4.4 Quelques applications

Pour terminer cette analyse, il est fort intéressant de calculer dans quelques cas précis la probabilité pour un individu de choisir tel ou tel mode de transport. Pour cela, on utilise la formule de probabilité relative au modèle multinomial :

$$P(i|C) = \frac{e^{V_i}}{\sum_{j=1}^N e^{V_j}}$$

où N est le nombre d'alternatives (dans le cas qui nous occupe N=3)

Pour chacune des applications, on fixe tous les paramètres (temps, coût, sexe, salaire,...) sauf un qui est le coût du trajet avec Swissmetro. Les informations relatives aux alternatives sont répertoriées dans un tableau, tandis que celle relatives à l'individu sont décrites à chaque fois. Il y a en tout six applications qui se rassemblent en trois groupes de deux. Pour chaque groupe, les données qui concernent les alternatives sont constantes et seules celles relatives à l'individu varient. Le premier groupe traite d'un trajet court, le deuxième d'un trajet de longueur moyenne et le dernier d'un trajet beaucoup plus long. Ainsi, il est intéressant de découvrir quel mode de transport est le plus apprécié pour un type de trajet particulier. De même, on peut savoir approximativement combien doit coûter Swissmetro pour qu'au moins un certain pourcentage de la population emprunte ce mode de transport.

Exemple 1

	Voiture	Swissmetro	CFF
Cout[CHF]	15	x	18
Temps[Min]	20	5	18
Frequence[Min]	0	30	30
Somme[Min]	0	5	5

Pour la situation type 1, considérons le cas d'un homme, entre 54 et 65 ans, qui gagne plus de 100'000 francs, qui va faire des achats avec sa famille (4 personnes), qui voyage en train, qui possède un abonnement général, qui n'a pas de bagages et qui voyage habituellement en première classe.

Pour la situation type 2, considérons le cas d'une femme, entre 39 et 54 ans, qui gagne moins de 50'000 francs, qui va faire des loisirs, qui voyage

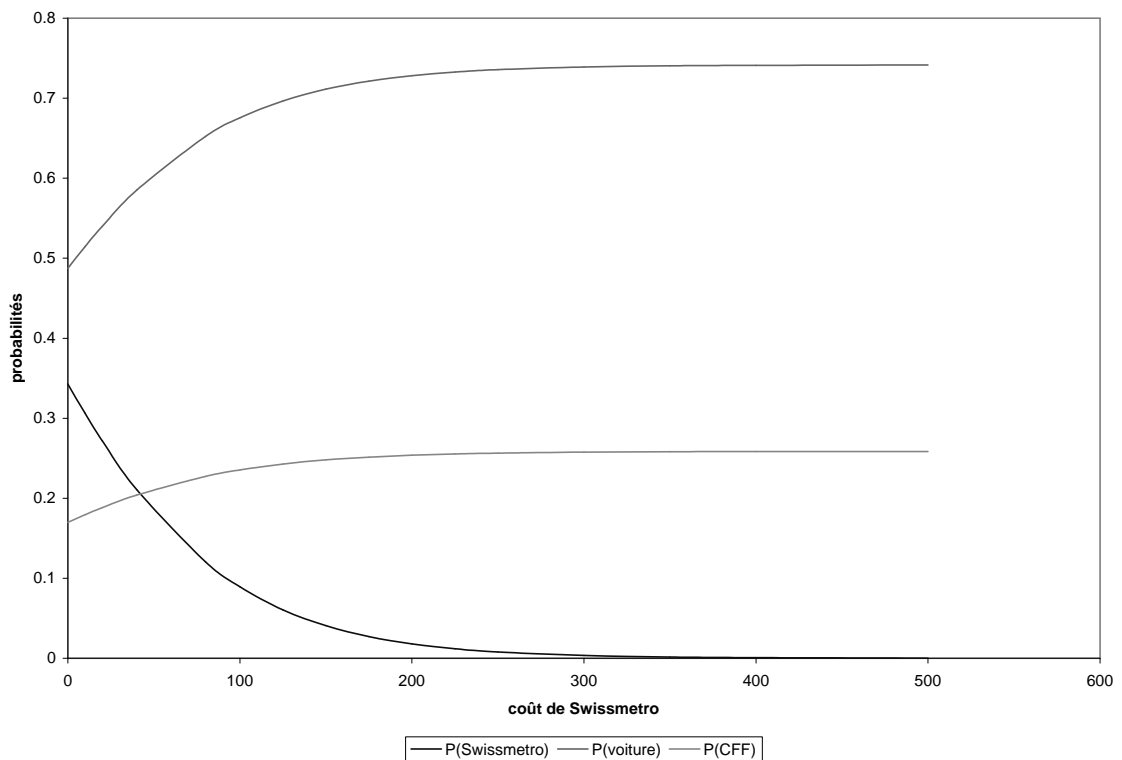


FIG. 1 – *Situation type 1.1*

seule, qui voyage en train, qui a un billet aller-retour avec demi-tarif, qui a des bagages et qui voyage habituellement en deuxième classe.

Exemple 2

	Voiture	Swissmetro	CFF
Cout	45	x	40
Temps	150	35	120
Frequence	0	30	30
Somme	0	25	15

Pour la situation type 1, considérons un homme entre 24 et 39 ans, qui gagne plus de 100'000 francs, qui est en voyages d'affaires, qui possède un abonnement général, qui voyage seul, qui voyage en train, qui voyage d'habitude en première classe et qui a des bagages.

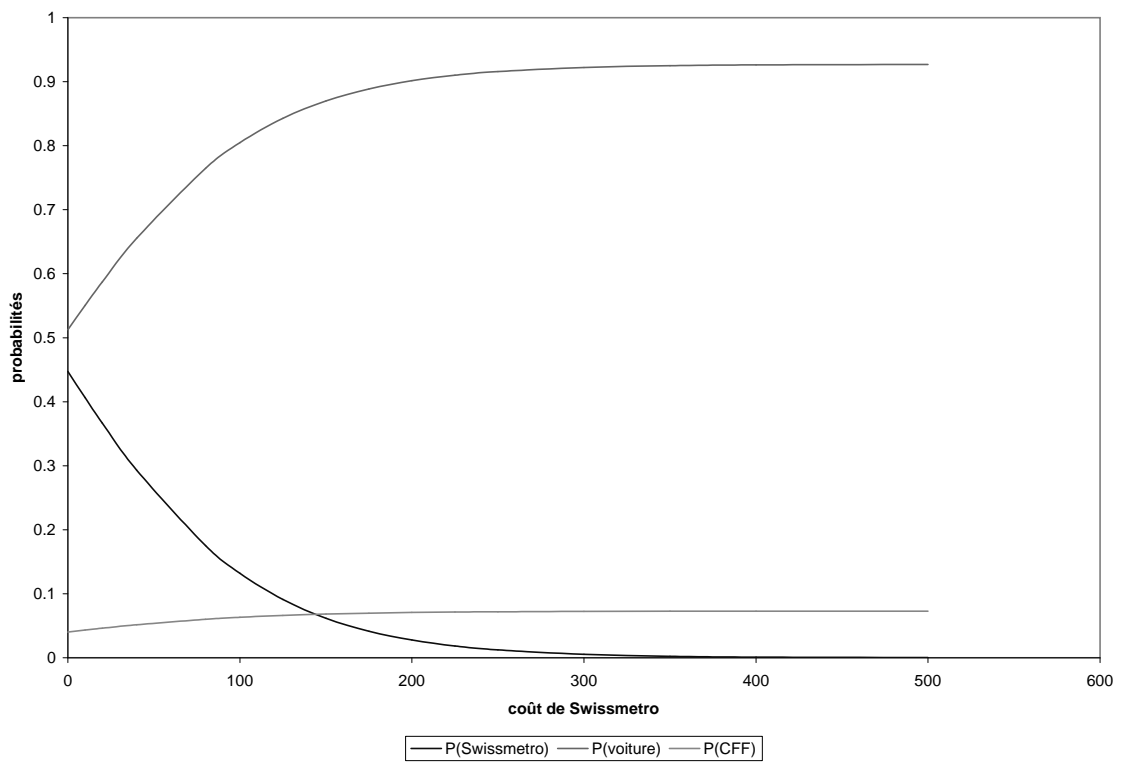


FIG. 2 – *Situation type 1.2*

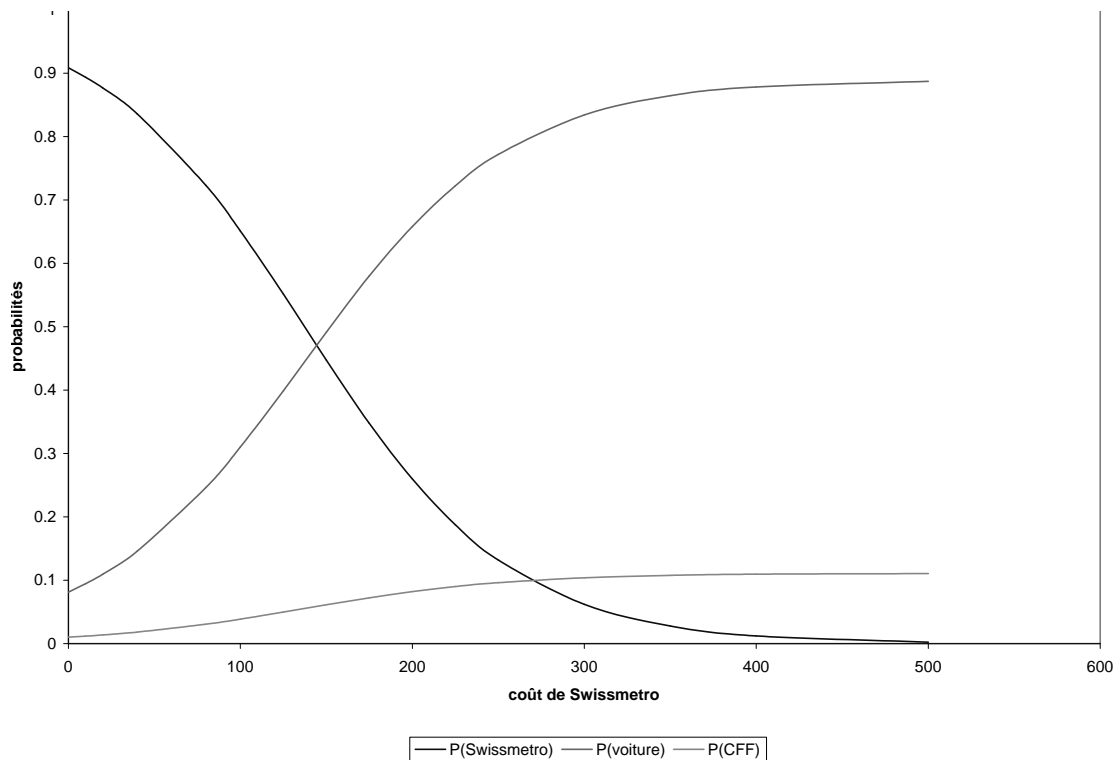


FIG. 3 – *Situation type 2.1*

Pour la situation type 2, considérons une femme entre 54 et 65 ans, qui gagne entre 50'000 et 100'000 francs, qui va faire des loisirs avec des amies (4 personnes), qui voyage en voiture, qui a un billet aller-retour sans demi-tarif, qui voyage habituellement en deuxième classe et qui n'a pas de bagages.

Exemple 3

	Voiture	Swissmetro	CFF
Cout	60	x	75
Temps	200	80	240
Frequence	0	30	30
Somme	0	10	10

Pour la situation type 1, considérons le cas d'un homme, entre 39 et 54 ans, qui gagne entre 50'000 et 100'000 francs, qui est pendulaire, qui

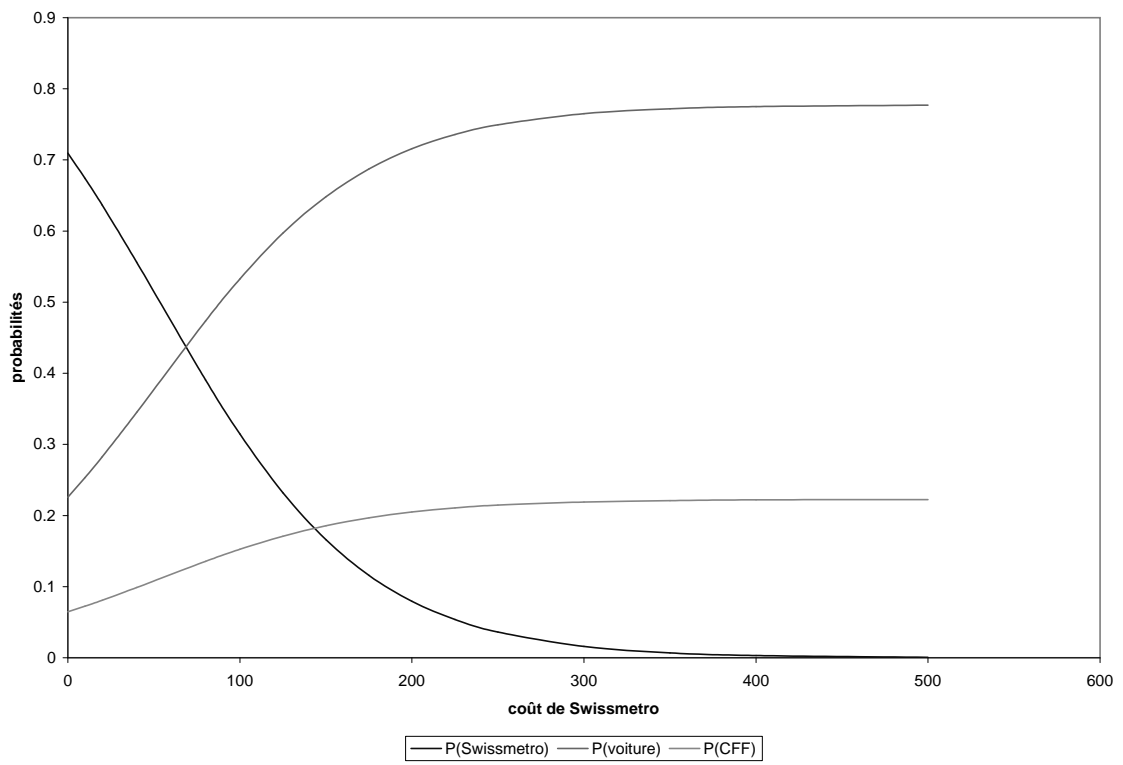


FIG. 4 – *Situation type 2.2*

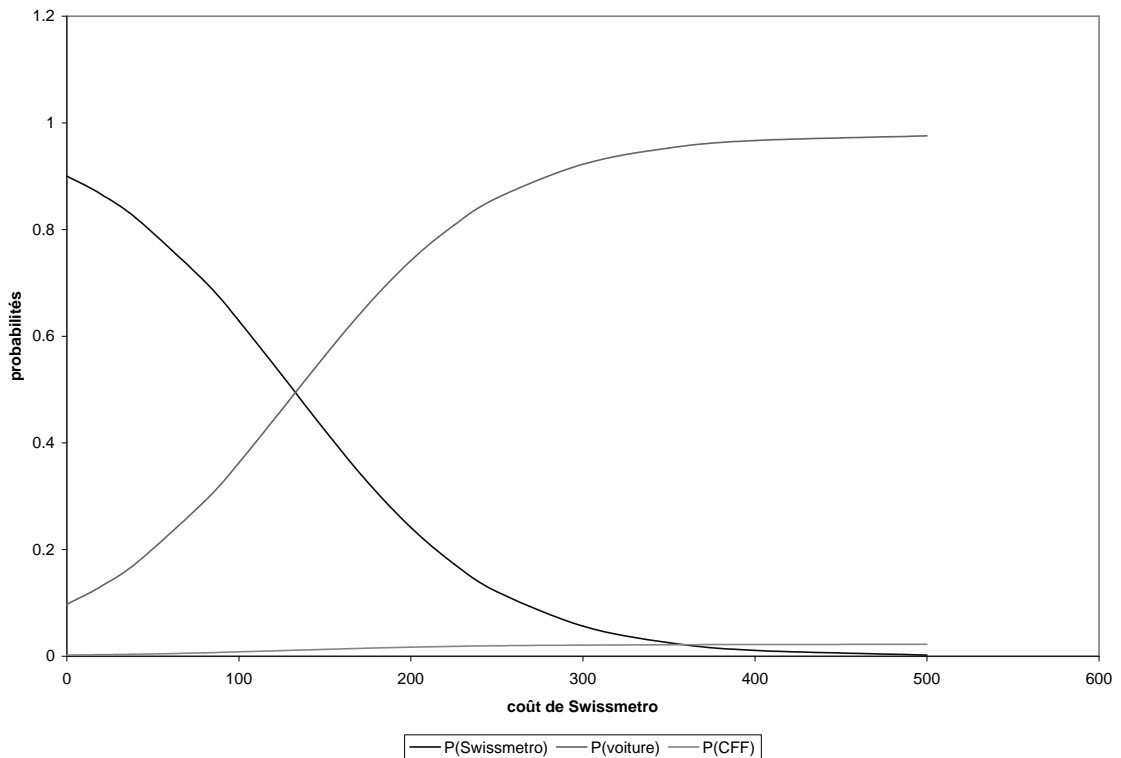


FIG. 5 – *Situation type 3.1*

voyage seul en voiture, qui a un billet aller-retour sans demi-tarif, qui voyage habituellement en deuxième classe et qui a des bagages.

Pour la situation type 2, considérons une femme de moins de 24 ans, qui gagne moins de 50'000 francs, qui va faire des loisirs avec une copine (2 personnes), qui voyage en train, qui possède un abonnement général junior, qui voyage d'habitude en deuxième classe et qui n'a pas de bagages.

Remarquons que ces situations types ont été choisies subjectivement en essayant tout de même de tenir compte d'une certaine logique par exemple entre l'âge et le salaire. De même, les personnes qui voyagent en voiture sont considérées comme ayant un billet aller-retour deuxième classe. Pour obtenir un résultat plus général, il faudrait tester toutes les possibilités puis pondérer par la proportion que chaque type de personne a dans la population suisse.

Les graphiques obtenus permettent de tirer de précieux enseignements, dont on aurait pu se douter parfois. Ainsi, on remarque que Swissmetro obtient les faveurs des gens sur les longs trajets. Cela est compréhensible, car

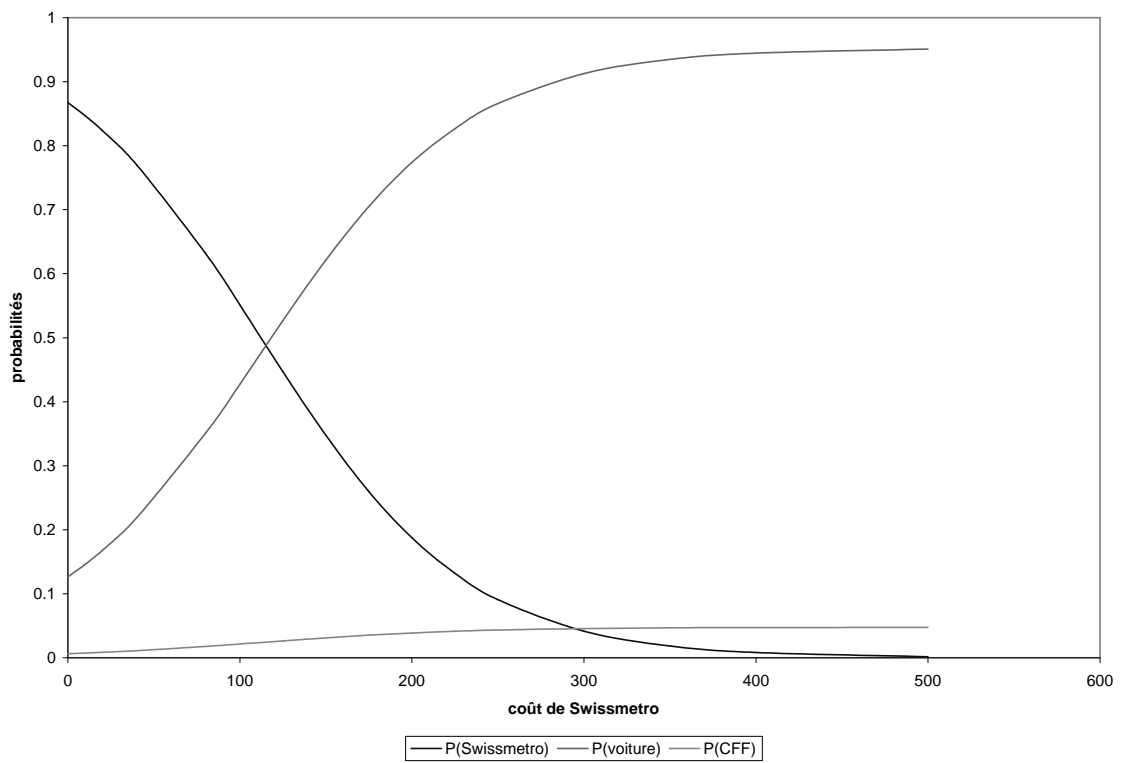


FIG. 6 – *Situation type 3.2*

c'est sur ce type de trajet que le gain de temps est substantiel. Les longs trajets permettent aussi de se rendre compte de la valeur élevée du temps. En effet, si l'on considère la situation type 2.1, on voit qu'une personne est prête à payer plus de trois fois plus cher pour aller environ quatre fois plus vite. Notons tout de même que la personne considérée est aisée. D'ailleurs la situation type 2.2 concerne une personne moins aisée et dans les mêmes conditions, on se rend compte qu'elle est prête à payer le double seulement pour aller environ quatre fois plus vite. On remarque également que le train est très peu prisé. Peut-être qu'au cours du sondage, les personnes interrogées ont perçu Swissmetro comme une alternative au train, mais pas à la voiture et ils ont ainsi focalisé leur choix entre la voiture et Swissmetro. Beaucoup de personnes ont considéré Swissmetro comme un train amélioré et de ce fait il apparaissait évident de préférer Swissmetro au train. Cet argument n'explique certainement pas tout, mais contribue à le faire. Sur les courts trajets, on remarque que même dans le cas où Swissmetro est gratuit, une majorité de personnes choisit la voiture pour leur déplacement. Ceci s'explique par le fait que le temps jusqu'à la gare et le temps de la gare au but représente un gros pourcentage du temps total et de ce fait les personnes ont vraiment l'impression de perdre leur temps. De toute façon, cette situation est très hypothétique car Swissmetro prévoit de relier uniquement les grandes villes de Suisse. Personne n'empruntera Swissmetro pour aller faire ses achats.

5 Conclusion

Les modèles de choix discrets ont encore de beaux jours devant eux. C'est en tout cas l'enseignement principal que l'on peut tirer après s'être impregné de ce sujet pendant 4 mois. D'une part, il est évident au vu des développements théoriques que le modèle cross-nested proposé par Vovsha peut-être amélioré. Mais la généralisation de ce modèle n'est-elle pas au fond qu'une étape dans le développement d'un modèle cross-nested définitif? Peut-être permettra-t-elle de développer d'autres modèles encore plus complexes mais surtout encore plus représentatifs de la réalité de la vie? L'avenir nous le dira. Toujours est-il que le modèle étudié dans la partie théorique de ce projet fait certainement avancer d'un pas la théorie des modèles de choix discrets. D'autre part, la calibration d'un modèle multinomial a déjà permis de recueillir des résultats plus qu'intéressants. Qu'en sera-t-il lorsque des logiciels auront été programmés pour calibrer des modèles cross-nested. Les études

d'impact pourront être affinées et permettront de meilleures prévisions. Quoiqu'il en soit, la "mort" des modèles de choix discrets ne semblent pas programmée pour bientôt et c'est tant mieux, car c'est réellement un outil de travail pratique et agréable.

6 Bibliographie

Ben-Akiva et Bierlaire (1999) Discrete choice methods and their applications to short-term travel decisions, in R. Hall (ed), *Handbook of Transportation Science*, Kluwer.

Ben-Akiva et Lerman (1985) *Discrete Choice Analysis*, MIT Press.

Bierlaire (1998) Discrete choice models, in M.Labbé, G.Laporte, K.Tanczos and Ph. Toint (eds), *Operations Research and Decision Aid Methodologies in Traffic and Transportation Management*, Vol. 166 of *NATO ASI Series, Series F: Computer and Systems Sciences*, Springer Verlag, pp.203-227.

Vovsha P. (1998) Application of Cross-Nested logit model to mode choice in Tel Aviv, Israel, metropolitan area, *Transportation Research Record* **1607** pp.6-15.

Ben-Akiva (1973) Structure of passenger travel demand models, MIT Press.