

Choix discrets expérimentaux- rappels méthodologiques

Dans le cadre des choix discrets expérimentaux, les individus sont confrontés à un ensemble de choix réalisables (et hypothétiques) concernant par exemple des moyens de transport (comme le bus ou le métro) et qui seraient différenciés par les attributs ou caractéristiques qui caractérisent chaque choix comme les prix et les durées du trajet et/ ou encore par les niveaux de confort. Généralement, les choix sont proposés aux individus de manière à les mettre en situation où ils devront inévitablement faire des compromis. L'intérêt de l'approche se situe d'ailleurs dans cette capacité à mettre en avant les compromis qui pourraient exister entre différents attributs caractérisant l'ensemble des choix. Faisant partie de la méthode des préférences exprimées, la méthode des choix discrets expérimentaux est une méthode d'évaluation fondée sur la nouvelle théorie de la consommation de Lancaster, qui stipule que les préférences pour un bien dépendent non pas de la consommation directe de ce bien mais plutôt indirectement des caractéristiques que ce bien possède ; et sur la théorie de l'utilité aléatoire (Hanley et Mourato, 2001). Le cadre théorique de l'utilité aléatoire décompose la fonction d'utilité indirecte de l'individu en deux composants : un élément observable qu'on assume comme étant une fonction linéaire des caractéristiques des différentes alternatives, et d'un élément aléatoire. La fonction d'utilité indirecte U_{ni} d'un individu n en choisissant une alternative i , est donnée par l'équation (1) :

$$U_{ni} = V_{ni} + \mathcal{E}_{ni} = \beta' X_{ni} + \mathcal{E}_{ni} \quad (1)$$

V_{ni} est la composante observable de l'utilité et \mathcal{E}_{ni} représente les termes aléatoires (tous les facteurs inobservables qui influencent les choix de l'individu), qui sont assumés comme indépendamment et identiquement distribués en valeurs extrêmes (*IID*). X_{ni} sont les vecteurs des différents attributs –spécifiques de l'alternative i pour l'individu n , tandis que β est le vecteur des 'poids' de l'utilité.

On assume que l'individu n choisit l'alternative qui va lui procurer le niveau d'utilité maximum. Ainsi, la probabilité que l'alternative i va être choisie parmi les alternatives j restantes est donnée par l'équation (2) :

$$\Pr [(U_{ni} > U_{nj}) \forall i \neq j] = \Pr [(V_{ni} - V_{nj}) > (\mathcal{E}_{nj} - \mathcal{E}_{ni})] \quad (2)$$

L'hypothèse que les termes aléatoires soient indépendamment et identiquement distribués en valeurs extrêmes (*IID*) nous permet d'exprimer l'équation (2) en termes de distribution logistique (Hanley et Mourato, 2001). Ainsi, on aura :

$$\Pr [(U_{ni} > U_{nj}) \forall i \neq j] = \exp (\mu V_{ni}) / \sum_g \exp (\mu V_{ng}) \quad (3)$$

Où le terme μ désigne le paramètre d'échelle, qui est inversement lié à la variance des termes aléatoires (Swait et Louvière, 1993), et g représente l'ensemble des alternatives restantes disponibles pour l'individu n . Dans cette spécification de choix probabilistique connue sous le nom de modèle logit conditionnel (CLM)¹⁹, on ne peut pas identifier séparément le paramètre d'échelle μ en travaillant sur un seul ensemble de données, c'est pourquoi on assume généralement que celui-ci est égal à l'unité. En conséquence, la sélection des choix doit se conformer à la règle d'indépendance des alternatives non pertinentes (IIA- Independence from Irrelevant Alternatives), qui spécifie que la probabilité relative de deux alternatives sélectionnées ne sera pas affectée par le rajout ou la suppression d'une alternative (Hanley et Mourato, 2001).

Cependant, la validité de la règle IIA est rarement rencontrée dans la pratique, étant donné que celle-ci impose une forte restriction sur les comportements de choix. Ainsi, l'application du CLM doit toujours être suivie d'un test qui va vérifier la validité de l'IIA comme celui de Hausman (Hausman et McFadden, 1984). Si la restriction imposée par l'IIA est violée, alors d'autres modèles comme le modèle logit mixte devraient être utilisés. Le principal avantage du modèle logit mixte comparé au CLM réside dans le fait qu'au lieu de fixer les coefficients des estimateurs des variables observées dans la spécification de la fonction d'utilité, on fait varier ces coefficients aléatoirement dans la population. La variance de β (voir équation 1) dans ce cas va provoquer des corrélations entre les différentes alternatives. En effet, pour chaque individu, β va devenir la somme de la moyenne b de la population et de l'écart-type σ_n , qui représente la variation des préférences individuelles relative à la préférence moyenne. L'équation (1) précédente est réécrite comme suit (Train, 1998) :

$$U_{ni} = b' X_{ni} + \sigma_n' X_{ni} + \varepsilon_{ni} \quad (4)$$

¹⁹ On utilise les modèles logit conditionnel quand la modélisation des choix dépend des caractéristiques des choix eux-mêmes.

Dans cette expression, à cause de la composante inobservable de la fonction d'utilité, $\sigma_n' X_{ni} + \mathcal{E}_{ni}$ est corrélée sur les différentes alternatives (à cause de σ_n), et le modèle logit mixte n'est plus restreint par la règle de l'*IIA*. La probabilité qu'un individu n va choisir l'alternative sera alors donnée par (Train, 1988) :

$$Pr_{ni} = \int L_{ni}(\beta) f(\beta) d\beta \quad (5)$$

Où $L_{ni}(\beta) = \exp(\beta' X_{ni}) / \sum_j \exp(\beta' X_{nj})$ à cause de la propriété des termes aléatoires qui sont indépendamment et identiquement distribués, et $f(\beta)$ représente la fonction de densité des variations des préférences parmi la population.

III.4 Etude de cas

La méthode des choix discrets expérimentaux est particulièrement appropriée quand on cherche à valoriser des biens (ou services) environnementaux avec différentes caractéristiques. L'objectif de l'évaluation est de déterminer comment les individus valorisent ces biens quand leurs caractéristiques varient. Dans notre cas, on s'intéresse sur les consentements des citoyens de la ville d'Antananarivo à faire une donation d'argent pour la préservation des forêts épineuses du plateau Mahafaly dans le sud-ouest de Madagascar, selon différents niveaux de forêts préservées et les impacts distributifs des donations.

Etant donné qu'on cherche à comparer les préférences des citoyens d'un pays en voie de développement à celles d'un pays développé, on a conservé les mêmes paramètres (nombre d'alternatives, nombre d'attributs, et nombre de niveaux pour chaque attribut) utilisés par Markova-Nenova et Wätzold (2017) pour concevoir leur expérimentation sur les choix réalisés à Cottbus, en Allemagne, pour la conception de notre expérimentation. Ainsi, pareillement que les choix discrets expérimentaux de Cottbus, le scénario concerne le lancement d'un projet de lever de fonds qui aurait pour but de conserver les forêts épineuses du plateau Mahafaly, actuellement menacées par les pratiques des cultures sur brûlis. Le projet cherche à mettre en place un fonds de conservation qui serait financé par des donations individuelles, duquel la population locale du plateau Mahafaly pourrait recevoir des compensations monétaires pour l'abandon des pratiques de cultures sur brûlis.

Comme attributs pour notre expérimentation sur les choix, on a considéré le niveau de conservation achevé (exprimé en étendue de forêts épineuses conservée), le montant de la

donation, et les impacts distributifs des paiements parmi les ménages au sein de la communauté locale. Pour ce dernier attribut, on a informé les participants que la manière dont les compensations seraient distribuées parmi les ménages va dépendre des traditions qui prévalent aux niveaux des communautés locales, et qui sont différentes dans la région Mahafaly. On a alors expliqué qu'on a identifié deux traditions parmi les communautés : une où la communauté locale donne la plus grande part d'un bien commun aux ménages les plus pauvres, et une autre où le partage est égal pour tous les ménages au sein de la communauté. Cependant, dans certaines communautés, on a mentionné qu'il n'était pas possible d'identifier quelle tradition prévalait et que de ce fait, on ne pouvait pas prédire comment les compensations monétaires seraient distribuées parmi les ménages locales.

On a sélectionné des niveaux d'attributs identiques à ceux de Markova-Nenova et Wätzold (2017) pour les attributs 'étendue de forêts épineuses conservée' et 'mode de distribution'. On a adapté les niveaux de l'attribut 'montant des donations' à la devise locale et au niveau de richesse et de revenu des donateurs potentiels, qui sont substantiellement bas à Antananarivo comparés à ceux de Cottbus. Le tableau III-1 montre les attributs et les niveaux d'attributs qu'on a pris pour l'expérimentation.

Tableau III- 1 Attributs et niveaux d'attributs

Attributs	Niveaux d'attributs
Etendue de forêts épineuses préservée (km ²)	25, 50, 75.
Mode de distribution : qui reçoit les compensations monétaires ?	Distribution égale Les ménages les plus pauvres reçoivent plus Distribution inconnue
Montant de la donation pour le projet de conservation (en Ariary; 1 euro = 2,872 Ariary)	200- 500- 1,000- 1,500- 2,000- 2,500- 5,000- 7,500- 10,000

Source : Auteur.

Un plan factoriel fractionnaire produit 27 profils pour notre expérimentation. Un design bloqué a permis de construire 9 cartes de choix réparties sur trois blocks de questionnaires. La figure III-1 illustre un exemple de carte de choix qu'on a utilisé pour l'expérimentation.

Figure III- 1 Un exemple de carte de choix

Si on vous demandait de faire une donation pour établir le fonds d'un projet qui vise à préserver les forêts épineuses de la région Mahafaly et que vous aurez seulement trois alternatives (A, B, C), laquelle choisiriez-vous?

	Alternative A	Alternative B	Alternative C: Pas de donation
Étendue de forêts épineuses préservées	 25 km ²	 50 km ²	0 km ²
Distribution des compensations monétaires: qui reçoivent les compensations?	"Distribution égale"	"Distribution inconnue"	X (signifie qu'aucun fonds ne sera mis en place)
Montant de la donation (en une seule fois)	Ar 1 000	Ar 1 500	Pas de donation
Je choisis: [S'il vous plaît, cochez seulement une case]	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Source : Auteur.

III.5 Collecte des données

Durant la phase initiale de collecte des données, on a mené une enquête préliminaire avec 20 participants afin de soulever les difficultés concernant la compréhension du questionnaire et du traitement des cartes de choix. Après avoir complété cette étape et discuté des difficultés rencontrées avec les participants, on a décidé de mener les prochaines enquêtes entièrement en Malagasy, tout en gardant le questionnaire en Français.

Pour recruter les participants, des annonces ont été faites dans différentes localités comme dans des bâtiments administratifs publics, des églises, et des universités. Selon la disponibilité des participants, quelques sessions ont été menées individuellement tandis que la majorité a été conduite en groupe. Pour chaque session de groupe, on a informé les participants que le questionnaire est strictement individuel et on a essayé de prévenir les discussions entre les participants pendant que ceux-ci complètent le questionnaire.

A chaque début de session, on a commencé l'expérimentation par une série de questions qui visaient principalement à familiariser les participants avec le sujet de l'enquête. On a par la suite fourni aux participants quelques informations sur les problématiques de conservation et la pauvreté qui prévalent dans la région Mahafaly, et on les a informés de la possibilité de

mettre en place un fonds de conservation. Toutes ces informations ont été transmises aux moyens d'une courte présentation avec des diapositives et d'une projection de film documentaire sur la biodiversité du plateau Mahafaly. Parallèlement, on a donné des instructions sur la manière de compléter le questionnaire, incluant les cartes de choix. A cause de la nature hypothétique de l'expérimentation, on a communiqué des informations 'cheap-talk' durant nos présentations pour éviter les biais hypothétiques. On a ainsi explicitement mentionné l'importance de considérer la capacité à faire une donation (i.e. revenu) en faisant un choix dans les cartes, et demandé aux participants de faire leurs choix comme si la situation se présentait réellement. Les scripts de 'cheap talk' avaient été efficaces à limiter les biais hypothétiques dans les enquêtes sur les choix discrets expérimentaux (Carlsson et al., 2005).

Après le remplissage des cartes de choix, l'enquête était conclue avec une série de questions qui visaient à identifier les opinions générales des participants sur la conservation de l'environnement à Madagascar, et sur différentes formes de distribution des compensations monétaires. Ces questions touchaient aussi des informations socio-économiques sur les participants et des informations sur leur appartenance à des organismes environnementaux et/ou caritatifs. Chaque participant recevait 6,000 Ariary (approximativement 2 euros) pour sa participation.

L'enquête a été menée par l'un des auteurs et son assistant durant la période de Mai- Juillet 2013 en 23 sessions, avec 300 participants au total. Parmi ces 300, seulement deux questionnaires complétés ont été écartés à cause d'incohérences et de réponses protestataires. Le tableau III-2 présente les données statistiques de la population enquêtée. La fiche questionnaire de l'enquête est consultable dans l'annexe 1.

Tableau III- 2 Statistiques sur la population enquêtée

	Population enquêtée en % (effectif)
Sexe	
Homme	46 (136)
Femme	54 (162)
Age ¹ (années)	
18-28	39 (114)
29-48	40 (117)
49-68	20 (61)
69-88	1 (3)
Revenu net mensuel (en Ariary; 1 euro = 2,872 Ariary)	
< 20 000	60.4 (180)
20,000 à moins de 50,000	15.8 (47)
50,000 à moins de 100,000	6.4 (19)
100,000 à moins de 200,000	8 (24)
> 200,000	9.4 (28)
Taille (effectif)	298

¹ Trois participants n'avaient pas donné leur âge.

Source : Données de l'auteur.

III.6 Résultats

Dans ce qui va suivre, on présentera les résultats des modèles logit conditionnel et logit mixte, les analyses des consentements à payer et la comparaison des préférences des participants venant d'Antananarivo et de ceux venant de Cottbus. On présentera enfin les résultats d'estimation qui illustrent les sources de l'hétérogénéité des préférences pour les participants d'Antananarivo.

III.6.1 Modèle logit conditionnel

On a commencé l'estimation du modèle par celle du logit conditionnel. On a utilisé le logiciel NLOGIT5 pour l'estimation, en considérant seulement les variables relatives aux attributs. Les niveaux de l'attribut 'mode de distribution' ont été codés par effets (*effect-coded*) afin de mieux mettre en exergue les compromis que les participants feraient face durant les processus de choix.

Tableau III- 3 Estimateurs du modèle logit conditionnel

Variables	Estimateurs (écart-type entre parenthèses)
Etendue de forêts épineuses préservée	0.02939*** (0.00141)
Distribution égale	0.39136*** (0.04354)
Les ménages pauvres reçoivent plus	0.46419*** (0.04441)
Distribution inconnue	-0.47620*** (0.05274)
Montant de la donation	-0.97850 x 10 ⁻⁰⁴ *** (0.1222 x 10 ⁻⁰⁴)
Nombre de participants	298
Nombre d'observations	2,682
Log-likelihood	-2,078.27125
AIC (normalisé)	1.554
Pseudo-R ²	0.2947

*** significatif à 1%, ** significatif à 5%, * significatif à 10%

Source : Calculs de l'auteur.

Comme le tableau III-3 le montre, tous les coefficients sont significatifs à 1% de niveau et sont tous correctement signés. Comme attendu, les participants font face à des compromis entre le montant de la donation, les impacts distributifs, et les performances de conservation du projet. Les estimateurs des variables 'étendue de forêts épineuses préservée', 'distribution égale', et 'les ménages pauvres reçoivent plus' sont tous positifs et significatifs à 1% de niveau, ce qui signifie que les participants ont des préférences à la fois pour les objectifs de conservation que pour certaines formes de distribution. En effet, le fait que le coefficient de la variable 'distribution inconnue' soit négatif et significatif à 1% de niveau signifie que les participants préfèrent ne pas choisir les projets qui ne fournissent pas d'information sur les impacts distributifs des donations.

Bien que le modèle logit conditionnel qu'on a estimé ait une performance décente (*pseudo-R²* ~ 0.3), il est recommandé qu'on poursuive l'estimation par un test de validité de la restriction imposée par l'*IIA* comme expliqué auparavant (Hensher et al., 2005), qui consiste typiquement à appliquer du test de spécification de Hausman. Ce test a pour objectif de comparer les estimateurs d'un modèle logit conditionnel restreint, où l'on a enlevé l'une des alternatives des séries de choix, avec le modèle logit conditionnel initial (qui comprend toutes les alternatives). Si la restriction imposée par l'*IIA* est valide (hypothèse nulle), alors les paramètres estimés des deux modèles (logit conditionnel restreinte et non restreinte) ne doivent pas significativement être différents.

Le test de spécification de Hausman qu'on a réalisé avec le modèle logit conditionnel n'a pas écarté l'hypothèse nulle à 1% de niveau de significativité pour l'alternative A. Par contre, la restriction imposée par l'*IIA* est rejetée pour l'alternative B à 1% de niveau de significativité. Pour l'alternative C, le test a donné une valeur négative de la statistique de *Chi-deux*, que l'on

pourrait interpréter comme une indication en faveur de l'hypothèse nulle (Hensher et al., 2005). Pour assurer une conclusion sans ambiguïté pour la validité de l'IIA pour l'alternative C, on a réalisé un test de spécification de Hausman généralisé, qui a la capacité de ne pas produire des valeurs négatives de la statistique de Chi-deux (Wessie, 1999) en utilisant le logiciel STATA. Les tests pour les alternatives B et C ont rejeté la validité de la restriction imposée par l'IIA à 1% de niveau de significativité. Une telle violation reflète les limitations du modèle logit conditionnel à identifier toutes les sources de corrélation entre les alternatives, telle que l'existence de différences de préférences parmi les participants (Train, 2009).

III.6.2 Modèle logit mixte

Aux vus des limitations du modèle logit conditionnel, un modèle différent capable de pallier la restriction imposée par l'IIA devrait être utilisé, comme le modèle logit mixte (McFadden et Train, 2000). Les procédures pour estimer ce type de modèle comprennent généralement : le choix et le nombre des paramètres à faire varier aléatoirement, le choix des distributions aléatoires, et les interprétations des estimateurs du modèle (Hoyos, 2010).

Tout d'abord, on a estimé un modèle logit non restreint pour identifier l'existence de préférences hétérogènes dans la population. Comme pour le cas de Markova-Nenova et Wätzold (2017), on a spécifié une distribution triangulaire pour l'attribut étendue de forêts épineuses préservées et l'attribut donation, tandis que pour l'attribut relatif aux modes de distribution, on a spécifié une distribution normale. L'ajustement global du modèle ainsi estimé est légèrement mieux que celui du modèle logit conditionnel précédent (*pseudo-R*² égal à 0.299 comparé à 0.295 pour le logit conditionnel). Tous les coefficients sont correctement signés et sont significatifs à 1% de niveau. Pour l'estimation du modèle logit mixte, la significativité des estimateurs des écarts-types indique la présence de préférences hétérogènes parmi la population. On a pu observer cela pour les coefficients des variables 'étendue de forêts épineuses préservée' (1% de niveau de significativité), 'les ménages pauvres reçoivent plus' (10% de niveau de significativité), 'distribution inconnue' (5% de niveau de significativité), 'donation' (1% de niveau de significativité). Cependant, l'écart-type de l'estimateur moyen de la variable 'distribution égale' n'est pas significatif, ce qui suppose l'absence de préférences hétérogènes pour ce niveau d'attribut.

Pour obtenir des conclusions plus précises sur l'existence de préférences hétérogènes pour l'estimateur moyen du coefficient 'distribution égale', on a estimé un deuxième modèle logit mixte, dans lequel on a imposé une égalité entre l'estimateur de l'écart-type et l'estimateur

moyen du coefficient 'distribution égale' comme l'avait suggéré Hensher et al. (2005). Comme pour le cas de Markova-Nenova et Wätzold (2017), on a aussi imposé la même contrainte sur la distribution triangulaire pour l'attribut 'montant de la donation' afin d'éviter un changement de signe. On n'a pas observé une amélioration significative de la performance du modèle ainsi estimé, comparé au modèle logit mixte estimé précédemment. Par contre, l'estimateur de l'écart-type du paramètre aléatoire 'distribution égale' est maintenant significatif à 1% de niveau, ce qui suggère l'existence de préférences hétérogènes pour ce niveau d'attribut. Les autres estimateurs des écart-types des paramètres aléatoires sont aussi significatifs, à 1% de niveau pour l'étendue de forêts épineuses préservée, à 10% de niveau pour 'les ménages pauvres reçoivent plus', à 5% de niveau pour la 'distribution inconnue', et à 1% de niveau pour le 'montant de la donation'. Le tableau III-4 montre la comparaison entre les résultats des deux estimations logit mixte qu'on a effectuées.