

Stratégies d'échantillonnage et modèles de comptage dans la méthode des coûts de transport

Sébastien TERRA

*Ministère de l'Écologie et du Développement Durable
Direction des Études Économiques et de l'Évaluation
Environnementale*



La méthode des coûts de transport

✧ Origine : Hotelling (1947)

✧ Méthode de valorisation

✧ fondée sur des préférences révélées

✧ indirecte

✧ De la fréquentation d'un site naturel au cours d'une année à la fonction de demande

✧ Enquêtes

✧ Téléphonique auprès d'un échantillon représentatif de la population

✧ En face-à-face sur le site étudié

Les modèles de comptage dans la méthode des coûts de transport

- ✿ Comment modéliser la demande de visites ?
- ✿ L'enquête téléphonique et les modèles de comptage associés
- ✿ L'enquête en face-à-face et les modèles associés
- ✿ Les données et les variables utilisées
- ✿ Les résultats obtenus
- ✿ Le calcul des variations de bien-être des visiteurs

3/20

Comment modéliser la demande de visites ?



✧ Objectifs de l'étude

- ✧ Modéliser le nombre de visites effectuées au cours des 12 mois précédant l'enquête
 - ✧ Nature de cette variable rend intéressante l'utilisation de modèles de comptage
- ✧ Étudier les modèles économétriques de comptage
- ✧ Comparer les différents modèles en termes de surplus du consommateur, variations équivalente et compensatrice

✧ Données utilisées

- ✧ 2 enquêtes réalisées en 2002 pour estimer la valeur d'usage récréatif d'une zone humide intérieure
- ✧ Lac du Der (Marne) : barrage-réservoir de la Seine – Plus grand lac artificiel d'Europe

L'enquête téléphonique

5/20

✧ Avantages

- ✧ Permet d'interroger des visiteurs et des non-visiteurs
- ✧ Coût unitaire plus faible que pour l'enquête en face-à-face

✧ Inconvénients

- ✧ Quand la proportion de visiteurs est faible, la taille de l'échantillon nécessaire devient grande
- ✧ Comment déterminer l'étendue du « marché » ?

✧ Déroulement

- ✧ Sondage par la méthode des quotas (sexe, âge, PCS)
- ✧ 2 021 personnes interrogées en décembre 2002
- ✧ Durée du questionnaire : 10 minutes

Modèles pour l'enquête téléphonique (1)

6/20

- ✧ Visiteurs et non-visiteurs sont interrogés
- ✧ Hypothèse 1 : un seul processus
 - ✧ Le nombre de visites effectuées par les visiteurs et les non-visiteurs peut être décrit par le même processus stochastique
- ✧ Hypothèse 2 : deux décisions
 - ✧ « participation » ou non : modélisation de la probabilité d'effectuer au-moins 1 visite
 - ✧ Niveau de fréquentation, conditionnellement à la décision de participation
- ✧ Trois modèles :
 - ✧ Cragg
 - ✧ ZIP (Zero-Inflated Poisson)
 - ✧ ZINB (Zero-Inflated Negative Binomial)

Modèles pour l'enquête téléphonique (2)

✧ Hypothèse 1 : même processus pour visiteurs et non-visiteurs

✧ 2 modèles de base

✧ Modèle de Poisson

$$\forall k \geq 0, P(n_i = k / \mathbf{x}_i) = e^{-\lambda_i} \frac{\lambda_i^k}{k!}$$



Problème : sur-dispersion

✧ Modèle binomial négatif

$$\forall k \geq 0, P(n_i = k / \mathbf{x}_i) = \frac{\Gamma\left(k + \frac{1}{\alpha}\right)}{\Gamma(k+1)\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{\frac{1}{\alpha}}{\frac{1}{\alpha} + \lambda_i}\right)^{\frac{1}{\alpha}} \left(\frac{\lambda_i}{\frac{1}{\alpha} + \lambda_i}\right)^k$$

Modèles pour l'enquête téléphonique (3)

8/20

✧ Hypothèse 2 - 2 décisions : Participation et Fréquentation

✧ 3 modèles

✧ Cragg

✧ Décision de participation modélisée par un modèle Logit

✧ Niveau de fréquentation modélisée par un modèle de poisson tronqué en 0

$$\forall k \geq 1, P(n_i = k / \mathbf{x}_i, n_i > 0) = e^{-\lambda_i} \frac{\lambda_i^k}{k!(1 - \exp(-\lambda_i))}$$

✧ Estimation séparée des deux modèles

✧ ZIP (Zero-Inflated Poisson)

$$P(n_i = 0 / \mathbf{x}_i) = \pi_i + (1 - \pi_i)e^{-\lambda_i}$$

$$\forall k > 0, P(n_i = k / \mathbf{x}_i) = (1 - \pi_i)e^{-\lambda_i} \frac{\lambda_i^k}{k!}$$

$$\text{avec } \pi_i = \frac{1}{1 + \exp(-\mathbf{z}_i \gamma)}$$

✧ ZINB (Zero-Inflated Negative Binomial)

L'enquête en face-à-face, sur site

✿ Avantage

Permet d'interroger des visiteurs : intéressant quand la proportion de visiteurs est faible

✿ Inconvénients

- ✿ Ne permet pas d'interroger les non-visiteurs
- ✿ Comment obtenir un échantillon représentatif ?
- ✿ Stratification endogène

✿ Déroulement

- ✿ Sélection aléatoire des personnes interrogées (1 par groupe)
- ✿ 241 personnes interrogées en novembre 2002
- ✿ Entretiens réalisés en fin de semaine
- ✿ Durée du questionnaire : 10 minutes

Modèles pour l'enquête sur site

✧ Deux problèmes avec l'enquête sur site

- ✧ Données tronquées : seules les personnes qui visitent le site sont interrogées
- ✧ Stratification endogène : les personnes qui visitent fréquemment le site ont une probabilité d'être interrogées plus grande que les personnes qui visitent le site de façon occasionnelle.

✧ Shaw (1988) : modèle de Poisson

$$\forall k \geq 1, P(n_i = k / \mathbf{x}_i) = e^{-\lambda_i} \frac{\lambda_i^{k-1}}{(k-1)!}$$

✧ Englin-Shonkwiler (1995) : modèle binomial négatif

$$\forall k \geq 1, P(n_i = k / \mathbf{x}_i) = \frac{k \Gamma\left(k + \frac{1}{\alpha}\right) \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha}}}{\Gamma(k+1) \Gamma\left(\frac{1}{\alpha}\right) \left(\frac{1}{\alpha} + \lambda_i\right)} \left(\frac{1}{\alpha} + \lambda_i\right)^k \lambda_i^{k-1}$$

Les données

✧ Observations supprimées = valeurs manquantes pour

- ✧ Revenu
- ✧ Niveau de fréquentation
- ✧ Séjour de plusieurs jours dans la région (face-à-face)
- ✧ Nombre de visites supérieur à 50 (téléphonique)

✧ Echantillon utilisé :

- ✧ Téléphonique : 1 477 personnes
- ✧ Face-à-face : 111 personnes

Les variables utilisées

✧ Variable expliquée : nombre de visites effectuées au cours des 12 mois précédant l'enquête

✧ Variables explicatives

- ✧ Distance et temps de trajet entre le domicile et le site
- ✧ Moyen de transport utilisé
- ✧ Nombre de personnes effectuant le trajet
- ✧ Niveau de revenu
- ✧ Profession et catégorie socio-professionnelle (PCS)
- ✧ Nombre d'enfants
- ✧ Type d'activité pratiquée
- ✧ ...

Le calcul des coûts de transport

13/20

✧ Coût de transport simple (CT)

✧ $CT = 2 * D * CM / P$

✧ D : distance domicile-lac du Der

✧ CM : coût de revient kilométrique

✧ P : nombre de personnes effectuant le voyage

✧ Coût d'opportunité du temps (COT)

✧ COT : le temps consacré à se rendre sur le site aurait pu être consacré à d'autres activités (dont le travail)

✧ COT est fondée sur une fraction du salaire horaire (ici 1/3)

✧ $COT = (T/60) * 2 * (1/3) * (R/135)$

✧ Coût de transport total (CTT)

✧ $CTT = CT + COT$

Les sites substitués

- ✧ Deux sites substitués sont considérés :
 - ✧ Le lac de Madine
 - ✧ Le lac d'Orient
- ✧ CT et COT pour chacun des deux sites ont été calculés, à partir de la ville de résidence des personnes interrogées.
- ✧ Deux variables incluses dans le modèle :
 - ✧ Mesurent les coûts de transport totaux pour se rendre sur le site substitué le plus proche ou le deuxième plus proche à partir du domicile de la personne interrogée.
 - ✧ Signe positif attendu
- ✧ Coûts de transport vers les sites substitués non calculés pour l'enquête sur site.

Résultats des estimations pour l'enquête téléphonique (1)

✧ Modèle binomial négatif avec sites substitués

✧ Coûts transport : ↘

✧ Coûts transport substitut 1 : ↗

✧ Coûts transport substitut 2 : ↘

✧ Revenu : ↗ (non significatif)

✧ Homme : ↗

✧ Plus de 50 ans : ↘

✧ Nombre d'enfants : ↗

✧ Profession intermédiaire et employé : ↗ par rapport à ouvrier

✧ Autres PCS : pas de différence par rapport à ouvrier

✧ Observation des oiseaux : ↗

✧ Connaissance des intérêts du lac : ↗

Résultats des estimations pour l'enquête téléphonique (2)

16/20

- ✧ Modèles avec distinction entre participation et fréquentation
 - ✧ Probabilité de se rendre sur le site
 - ✧ Revenu : ↗
 - ✧ Age : ↘
 - ✧ Nombre d'enfants : ↗
 - ✧ Niveau de fréquentation
 - ✧ Mêmes signes que dans le modèle précédent

- ✧ Bilan statistique :
 - ✧ Modèles avec sites substitués sont préférés aux modèles sans prise en compte des sites substitués (Test de Vuong)
 - ✧ Modèles avec distinction entre participation et fréquentation sont préférés aux modèles simples (Test du rapport de vraisemblance)

Résultat de l'enquête sur site

✧ Modèle sans site substitut avec prise en compte de la stratification endogène

✧ Coûts transport : ↘

✧ Revenu : ↗ mais non significatif (NS)

✧ Homme : NS

✧ Plus de 50 ans : NS

✧ Profession libérale : signe négatif par rapport à ouvrier

✧ Autres PCS : NS

✧ Observation des oiseaux : NS

✧ Connaissance des intérêts du lac : NS

Estimation du bien-être des visiteurs

✧ Complémentarité faible (Maler, 1974)

✧ Modèles simples

✧ Surplus du consommateur (SC) $SC_i^v = -\frac{1}{\beta_{CT}}$ et $SC_i^a = -\frac{\lambda_i}{\beta_{CT}}$

✧ Variation équivalente (VE) $VC_i = \frac{1}{\beta_R} \ln\left(1 + \lambda_i \frac{\beta_R}{\beta_{CT}}\right)$

✧ Variation compensatrice (VC) $VE_i = -\frac{1}{\beta_R} \ln\left(1 - \lambda_i \frac{\beta_R}{\beta_{CT}}\right)$

✧ Modèle ZIP et ZINB

✧ Surplus du consommateur $SC_i^a = -(1 - \pi_i) \frac{\lambda_i}{\beta_{CT}}$

	NB	ZINB	NB + Subs	ZINB + Subs	Poisson FaF	Poisson FaF / tel
Valeur annuelle (€)						
SC	263,7 (53,7)	262,0 (47,7)	149,7 (22,1)	155,3 (19,1)	317,2 (125,2)	497,3 (167,5)
VC	257,4 (51,8)		148,8 (21,8)		300,5 (121,4)	463,5 (174,8)
VE	270,6 (56,0)		150,6 (22,3)		338,1 (136,7)	543,1 (169,5)
Valeur par visite (€)						
SC	37,9 (7,6)	37,9 (6,9)	21,3 (3,1)	22,5 (2,8)	20,4 (8,9)	20,40 (8,9)
VC	37,1 (7,3)		21,2 (3,1)		19,7 (8,6)	19,4 (8,6)
VE	38,7 (7,9)		21,4 (3,1)		21,2 (9,5)	21,7 (9,5)

Conclusion

✧ Résultats obtenus

- ✧ Ignorer les sites substitués sur-estime de 75 % la valeur du site
- ✧ Distinction participation / fréquentation est nécessaire

✧ Perspectives méthodologiques

- ✧ Meilleure prise en compte des valeurs manquantes
- ✧ Généralisation de l'usage des modèles de comptage
- ✧ Modélisation multi-site à l'aide de modèles à utilité aléatoire

✧ Etape suivante de la démarche

- ✧ Estimer la valeur d'usage totale du lac du Der