



**HAL**  
open science

# La congestion : une variable clé dans la méthode des coûts de déplacement

Mélody Leplat

► **To cite this version:**

Mélody Leplat. La congestion : une variable clé dans la méthode des coûts de déplacement. [Travaux universitaires] auto-saisine. 2012, 33 p. hal-01133266

**HAL Id: hal-01133266**

**<https://hal.science/hal-01133266>**

Submitted on 18 Mar 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**INRA**

AGRO  
CAMPUS  
OUEST

# La congestion : une variable clé dans la méthode des coûts de déplacement

Mélody LEPLAT

**Working Paper SMART – LERECO N°12-05**

September 2012

*Les Working Papers SMART-LERECO ont pour vocation de diffuser les recherches conduites au sein des unités SMART et LERECO dans une forme préliminaire permettant la discussion et avant publication définitive. Selon les cas, il s'agit de travaux qui ont été acceptés ou ont déjà fait l'objet d'une présentation lors d'une conférence scientifique nationale ou internationale, qui ont été soumis pour publication dans une revue académique à comité de lecture, ou encore qui constituent un chapitre d'ouvrage académique. Bien que non revus par les pairs, chaque working paper a fait l'objet d'une relecture interne par un des scientifiques de SMART ou du LERECO et par l'un des deux éditeurs de la série. Les Working Papers SMART-LERECO n'engagent cependant que leurs auteurs.*

*The SMART-LERECO Working Papers are meant to promote discussion by disseminating the research of the SMART and LERECO members in a preliminary form and before their final publication. They may be papers which have been accepted or already presented in a national or international scientific conference, articles which have been submitted to a peer-reviewed academic journal, or chapters of an academic book. While not peer-reviewed, each of them has been read over by one of the scientists of SMART or LERECO and by one of the two editors of the series. However, the views expressed in the SMART-LERECO Working Papers are solely those of their authors.*

---

**La congestion : une variable clé  
dans la méthode des coûts de déplacement**

Mélody LEPLAT

*ESC Bretagne Brest, F-29000 Brest, France  
Agrocampus Ouest, UMR1302 SMART, F-35000 Rennes, France  
INRA, UMR1302 SMART, F-35000 Rennes, France*

**Auteur pour la correspondance / Corresponding author**

**Mélody LEPLAT**

Ecole Supérieure de Commerce de Bretagne-Brest

2 avenue de Provence, CS 23812

29238 Brest Cedex 3, France

Email: [melody.leplat@esc-bretagne-brest.com](mailto:melody.leplat@esc-bretagne-brest.com)

Téléphone / Phone: +33 (0)2 29 00 62 29

Fax: +33 (0)2 23 48 53 80

*Les Working Papers SMART-LERECO n'engagent que leurs auteurs.*

*The views expressed in the SMART-LERECO Working Papers are solely those of their authors.*

## **La congestion : une variable clé dans la méthode des coûts de déplacement**

### **Résumé**

La méthode des coûts de déplacement est largement utilisée pour évaluer la valeur de sites récréatifs et simuler l'impact de politiques publiques sur le bien-être. La congestion, déterminant très probable de la demande récréative (Jakus et Shaw, 1997), doit faire l'objet d'une attention particulière de la part de l'analyste. En n'intégrant pas cette variable, on s'expose au biais d'omission. Mais l'insérer, c'est risquer d'introduire de l'endogénéité économétrique dans le modèle, car la congestion est susceptible d'être corrélée au terme d'erreur. De plus, la congestion est à la fois un déterminant du choix de site et le résultat de l'agrégation des choix individuels. Elle est donc le résultat d'un équilibre de Nash en probabilités. Plusieurs techniques d'estimation ont été récemment proposées pour résoudre tout ou partie de ces difficultés. Cet article présente ces méthodes et les illustre dans un modèle de choix discrets hiérarchisé appliqué au choix de site naturel récréatif.

**Mots-clefs :** congestion, méthode des coûts de déplacement, endogénéité, modèles de choix discret

**Classifications JEL :** Q26, Q51

### **Congestion: a key variable in travel costs demand models**

The travel cost method is widely used to estimate the value of recreational sites and to simulate the impact of public policies on welfare. Congestion is likely to be a determinant of recreational demand (Jakus and Shaw, 1997) and has to be given special attention. Ignoring this variable leads to the omission bias. But introducing congestion is also likely to involve endogeneity problems because of a potential correlation with the error term. In addition, congestion is determined by a sorting equilibrium, *i.e.*, it is both the result of the choice process and an explanatory variable of this choice. Some methods have been recently propose to address these difficulties. These methods are presented and illustrated in a nested model of recreational choice.

**Keywords:** congestion, travel costs methods, endogeneity, random utility models

**JEL classifications:** Q26, Q51

## **La congestion : une variable clé dans la méthode des coûts de déplacement**

### **1. Introduction**

Alors que la littérature a beaucoup exploré l'impact d'attributs de qualité sur le bien-être (Smith et Desvougues, 1985 ; McConnell, 1992 ; Parsons et Massey, 2003) celui de la congestion n'a été que peu étudié. Pourtant, la congestion est considérée depuis longtemps comme un élément central de la demande récréative (Anderson et Bonsor, 1974 ; Fisher et Krutilla, 1972) et la négliger implique un biais d'omission (Cesario, 1980) sur la valeur estimée des sites<sup>1</sup> et sur l'estimation des variations de bien-être. Par exemple, la fermeture d'un site entraîne le report de sa fréquentation sur ses sites substituts, et donc une hausse de la congestion globale tandis qu'une amélioration de la qualité d'un site a pour conséquence une hausse de la congestion à ce site. Si l'effet de la congestion est négatif, ces mesures entraînent une diminution du bien-être, qui n'est pas prise en compte dans un modèle où la congestion est omise. Il en résulte une sous-estimation de la valeur d'un site, et une surestimation de l'impact sur le bien-être d'une hausse de qualité (Timmins et Murdock, 2007 ; O'Hara, 2007a). En outre, sans indicateur de congestion, on ne dispose pas d'éléments empiriques sur lesquels s'appuyer pour la réguler. Ainsi, Hanley *et al.* (2002), dans une étude sur la régulation de l'accès à des parcours d'escalade en montagne, préviennent que leurs résultats sont à prendre avec précaution car la congestion n'a pas pu être introduite de manière satisfaisante. Il apparaît surprenant que, dans une étude qui examine différentes mesures de régulation de l'accès, les effets d'encombrement ne soient pas pris en compte.

Mais introduire la congestion présente des défis pour l'analyste d'un point de vue économique et économétrique. Un premier problème tient à sa mesure. Quel indicateur de congestion faut-il retenir ? La congestion réelle, anticipée, espérée ou perçue ? L'indicateur adéquat dépend du problème analysé (Jakus et Shaw, 1997) :

- si le choix de site est postérieur à l'observation de la congestion, ce sont les congestions réelle ou perçue qui expliquent ce choix ;
- si le choix est antérieur à l'observation de la congestion, il convient d'utiliser la congestion « *anticipée* » ou, mieux, la congestion « *espérée* » pour expliquer le choix du site.

---

<sup>1</sup> Un site récréatif est un espace naturel homogène où se pratiquent des activités récréatives.

Jakus et Shaw distinguent la congestion « *anticipée* » où les individus ont une connaissance subjective sur la distribution de probabilités des autres individus, de la congestion « *espérée* » où les individus en ont une connaissance objective. Dans ce second cas, on est en présence d'anticipations rationnelles. Sous cette hypothèse, le vecteur des niveaux de congestion anticipée aux différents sites est le même pour tous les individus, et égal au vecteur des espérances de congestion. La congestion est alors le résultat d'un équilibre de Nash en probabilités mis en évidence par Timmins et Murdock (2007), Boxall et Adamowicz (2000) et Phaneuf *et al.* (2009). Timmins et Murdock (2007) et Boxall et Adamowicz (2000) utilisent le terme de congestion « *anticipée* » plutôt que celui de congestion « *espérée* ». Dans la suite de cet article, on parlera de congestion *anticipée* dans le sens d'*espérée* de Jakus et Shaw (1997). Phaneuf *et al.* (2009) différencient les modèles d'équilibres dits « *simple* », qui traitent des interactions sociales, des modèles d'équilibres « *complexes* », qui étudient les interactions entre les agents économiques et le milieu naturel. Ainsi, sous l'hypothèse d'anticipations rationnelles, la congestion est le résultat d'un équilibre *simple* et elle revêt un caractère endogène au sens économique : elle est à la fois un déterminant du choix de site, et le résultat de ce choix. Cette particularité de la congestion est à l'origine d'un second problème dit d'identification parfois difficile à résoudre dans le cas des méthodes basées sur les comportements observés comme la méthode des coûts de déplacement (MCD). Ce problème est lié au fait que les niveaux de congestion des sites co-varient avec le niveau de qualité des sites, ce qui fait que la distinction entre les effets de la congestion et les effets propres de la qualité des sites est parfois difficile. Cette co-variation des niveaux de qualité et de congestion est un problème difficilement surmontable lorsque qu'on ne dispose que d'une observation, c'est-à-dire qu'on n'observe qu'une seule situation d'équilibre. Il est alors nécessaire d'imposer des hypothèses sur la forme des relations entre les niveaux de congestion et de qualité pour espérer distinguer les effets d'intérêt (Bayer et Timmins, 2007, proposent différentes solutions et commentent leurs limites). L'observation de plusieurs équilibres différents permet de s'affranchir, au moins en partie, de l'utilisation de ce type d'hypothèses. La méthode d'évaluation contingente (MEC) offre une solution alternative simple à ce problème d'identification. Le caractère hypothétique des scénarios permet de faire varier les niveaux de congestion indépendamment des niveaux de qualité, sous réserve que ces variations demeurent crédibles pour les enquêtés. Ainsi, l'essentiel des études qui introduisent la congestion utilisent la MEC (Boxall *et al.*, 2003 ; McConnell, 1977 ; Cicchetti et Smith, 1973). Cette méthode s'appuie sur des comportements déclarés, hypothétiques, contrairement à la MCD qui s'appuie sur des comportements observés. De ce fait, la MEC « casse » le lien qui existe entre le comportement individuel et la congestion,

c'est-à-dire la somme des décisions individuelles de visite, qui devient alors exogène.

Le problème de la corrélation entre les niveaux de congestion et de qualité peut être accompagné d'un autre problème économétrique. Si les sites considérés sont hétérogènes et si les variables disponibles décrivant les sites ne permettent pas un contrôle suffisant de cette hétérogénéité, alors l'économètre doit introduire des variables latentes dans ses modèles d'utilité aléatoire, une par site. Ces variables latentes contiennent l'effet sur l'utilité procuré par un site, des attributs non-observés de ce site, et sont typiquement « absorbées » par les termes d'erreur des modèles d'utilité aléatoire. Si ces attributs sont des déterminants importants des choix des visiteurs, alors les niveaux de congestion seront corrélés avec les termes d'erreur du modèle de choix discrets. Ce problème d'endogénéité « économétrique » est classique en micro-économétrie. S'il est résolu à l'aide de techniques à variables instrumentales simples dans le cas des modèles continus, la difficulté augmente dans le cas de modèles présentant de fortes discontinuités, par exemple les modèles de choix discrets (Bayer et Timmins, 2007 ; Timmins et Murdock, 2007). Dans la suite de l'article, le terme d' « *endogénéité économique* » sera utilisée pour qualifier le fait que la congestion est le résultat d'un équilibre de Nash en probabilités, tandis que le terme d' « *endogénéité économétrique* » fera référence à la potentielle corrélation entre la variable de congestion et le terme aléatoire du modèle utilisé.

Cet article passe en revue les différentes manières d'intégrer la congestion dans un modèle explicatif de la demande récréative appliquant les méthodes de préférences révélées. La première section est consacrée aux modèles qui introduisent la congestion comme une variable explicative « lambda », c'est-à-dire exogène, puis la section suivante explique les techniques utilisées dans la littérature pour tenir compte de l'endogénéité économique de la congestion. Dans la mesure où ces techniques ne permettent pas de contrôler la potentielle endogénéité économétrique de la congestion, la troisième section analyse les articles qui proposent de traiter ce problème par l'application de variables instrumentales dans les modèles de choix. La quatrième section présente une méthode initialement développée dans les modèles de choix de biens différenciés (Berry, 1994 ; Berry *et al.*, 1995), qui permet d'estimer le modèle à l'équilibre de congestion tout en contrôlant l'endogénéité économétrique probable de cette variable. Cet article se termine par l'illustration des techniques présentées sur un modèle de choix discrets hiérarchisé appliqué au choix de sites récréatifs naturels.



## 2. Première approche : la congestion comme une variable exogène

A partir des années 70 et plus encore après l'article de Cesario (1980) mettant en évidence le biais d'omission, les études expliquant le comportement récréatif ont commencé à introduire la congestion parmi les variables explicatives, d'abord dans des modèles de demande continue, c'est-à-dire qui intègrent une fonction de demande de visite continue adressée à un seul site, puis dans les modèles de choix discrets qui se sont développés depuis les années 90.

Les principaux résultats des modèles de demande continue traitant de la congestion sont exposés dans le Tableau 1. Dans ces modèles, les auteurs utilisent soit des indicateurs objectifs de congestion réelle (nombre de visiteurs/surface) (Deyak et Smith, 1978 ; Salanié, 2006) soit des indices de congestion perçue, obtenus en interrogeant les visiteurs sur leur perception de la congestion (Allen et al., 1981 ; Bell et Leeworthy, 1990), et l'impact de la congestion sur l'utilité diffère selon les études. Ainsi, il est positif chez Bell et Leeworthy (1990), négatif chez Allen et al. (1981) ou encore quadratique pour Salanié (2006), c'est-à-dire que des niveaux faibles de congestion influent positivement sur la demande de visite, mais cet impact devient négatif pour des niveaux de congestion plus forts. Brown et Mendelsohn (1984) ont une approche un peu différente car ils utilisent la méthode hédonique des coûts de déplacement. Ils expliquent le coût total de visite à un site par les attributs de ce site, dont la congestion perçue par les pêcheurs. Pour deux tiers des 126 sites, l'impact de la congestion est négatif mais il est généralement non significatif. Les auteurs concluent donc que la congestion n'est pas un facteur explicatif consistant du choix de sites de pêche. Deyak et Smith (1978) observent également des impacts différents de la congestion en estimant des fonctions de demande pour des sites de campings où la décision de participer est modélisée en amont. Ils obtiennent un impact négatif de la congestion sur le choix des sites mais un impact positif sur la participation.

**Tableau 1 : Modèles de demande continue avec congestion (MCD)**

| Référence                  | Activité  | Type d'indicateur | Indicateur   | Effet       | Significativité      | Nombre d'alternatives |
|----------------------------|-----------|-------------------|--------------|-------------|----------------------|-----------------------|
| Brown et Mendelsohn (1984) | Pêche     | Indicatrices      | Cong. perçue | <0 ou >0    | Non significatif     | 126                   |
| Deyak et Smith (1978)      | Randonnée | Densité           | Cong. réelle | >0 et <0    | Selon spécifications | 2                     |
| Bell et Leeworthy (1990)   | Plages    | Indicatrices      | Cong. perçue | >0          | 1%                   | 1                     |
| Allen <i>et al.</i> (1981) | Camping   | Indicatrices      | Cong. perçue | >0 ou <0    | Non significatif     | 1                     |
| Salanié (2006)             | Pêche     | Densité           | Cong. réelle | Quadratique | Selon spécifications | 28                    |

A partir des années 90, les applications de la méthode des coûts de déplacement ont progressivement abandonné les modèles de demande continue car ils ne permettent pas de prendre en compte de manière satisfaisante la substitution entre les différents sites visités. D'autre part, Mullahy (1986) a montré que ne pas tenir compte de la nature discrète du nombre de visites conduisait à des évaluations de surplus erronées. Ces deux problèmes peuvent certes être résolus en estimant les fonctions de demande par des modèles de comptage, le prix des substituts étant alors introduit dans les fonctions de demande (voir par exemple Ozuna et Gomez, 1994 ; Englin et Shonkwiler, 1995 ; Bin *et al.*, 2005). Cependant, seul un nombre limité de substituts peut être introduit. Il est donc préférable de recourir aux modèles de choix discrets à utilité aléatoire (*Random Utility Model* ou RUM) quand le nombre de substituts est important (Parsons, 2004), ce qui explique l'engouement pour ces modèles à partir des années 90.

Dans les modèles de choix discrets, chaque site  $j$  est supposé conférer une utilité  $V_{ij}$  à l'individu  $i$  qui peut être décomposée entre une partie déterministe  $U_{ij}$  qui dépend des variables observables telles que les caractéristiques  $Z_i$  de l'individu  $i$ , les attributs  $X_j$  du site  $j$ , dont la congestion, le coût de déplacement  $TC_{ij}$  de l'individu  $i$  pour se rendre sur le site  $j$ , et une partie aléatoire,  $\varepsilon_{ij}$  :

$$V_{ij} = U_{ij}(Z_i, X_j, TC_{ij}) + \varepsilon_{ij} \quad (1)$$

Les individus choisissent le site qui maximise leur utilité  $V_{ij}$  et leur choix peut ensuite être déterminé selon une probabilité  $p_{ij}$  dont la forme dépend des hypothèses sur la distribution du terme aléatoire. Par exemple, le modèle logit multinomial, modèle de choix discrets le plus répandu, suppose que les aléas sont indépendants et identiquement distribués selon une loi de Gumbel. Les modèles de choix permettent de calculer facilement le surplus et d'estimer de manière satisfaisante l'impact sur le bien-être de variations de qualité, ce qui a contribué à leur développement.

Cependant, la prise en compte de la congestion dans les RUM a fait l'objet de peu de travaux de modélisation comme le montre le Tableau 2 synthétisant les principales caractéristiques et résultats des RUM estimés avec une variable de congestion. Quelques auteurs ont commencé à introduire la congestion comme variable explicative dans les années 90. Dans ces analyses, la

congestion n'est pas un élément central mais une variable explicative parmi d'autres, introduite pour éviter le biais d'omission (Cesario, 1980). L'impact de la congestion est le plus souvent négatif et significatif (Berman *et al.*, 1997 ; Hansen *et al.*, 1999 ; Yen et Adamowicz, 1994) mais on observe également parfois des effets quadratiques (Salanié, 2006 ; Schuhman et Schwabe, 2004), voire positifs (Schuhman et Schwabe, 2004 ; Lin *et al.*, 1996). Beaucoup de ces articles s'intéressent à la pêche récréative (Berman *et al.*, 1997 ; Lin *et al.*, 1996 ; Salanié, 2006 ; Schuhman et Schwabe, 2004).

Les études mettant en oeuvre des modèles de demande continue comme celles utilisant les RUM ne permettent pas d'établir un effet clair de la congestion sur l'utilité. Schneider et Hammitt (1995) expliquent ces résultats contrastés par trois effets :

- l'effet de « *product shift* » : devant des conditions de visites imprévues, par exemple une fréquentation plus importante que prévue, le visiteur change sa conception de l'expérience récréative. Le visiteur abordera la visite différemment en termes d'activités, de durée de visite etc., de sorte que l'expérience récréative vécue ne correspond pas à celle attendue. Le visiteur peut alors considérer le site comme un nouveau site : le site congestionné et le site non congestionné deviennent des biens différents ;
- l'effet de rationalisation : la congestion pousse le visiteur à s'interroger sur la qualité de la visite récréative. Finalement, il décide que l'encombrement n'a pas eu d'impact sur son expérience récréative ;
- l'effet d'adaptation : si le visiteur considère que la congestion au site cible est trop importante, il peut décider de renoncer à sa visite. Les activités récréatives qui génèrent de fortes dépenses seraient plus sujettes à l'effet de rationalisation (Shelby *et al.*, 1988).

**Tableau 2 : Analyses empiriques traitant de la congestion dans des modèles de choix discrets**

| Référence   | Endogénéité économétrique traitée | Activité  | Endogénéité économique prise en compte | Type d'indicateur               | Indicateur de congestion | Effet                | Significativité      | Nombre d'alternatives |
|---|-----------------------------------|-----------|--|---------------------------------|--------------------------|----------------------|----------------------|-----------------------|
| Berman <i>et al.</i> (1997)                             | Non                               | Pêche     | Non                                    | Indicatrice                     | Non précisé              | <0                   | 10%                  | 30                    |
| Lin <i>et al.</i> (1996)                                | Non                               | Pêche     | Non                                    | Densité                         | Cong. réelle             | >0                   | 5%                   | 4                     |
| Kerkvliet et Nowell (2000)                              | Non                               | Pêche     | Non                                    | Densité                         | Cong. réelle             | <0                   | n.s. <sup>b</sup>    | 5                     |
| Schuhman et Schwabe (2004)                              | Non                               | Pêche     | Non                                    | Densité                         | Cong. réelle             | >0 ou Q <sup>d</sup> | Selon spécifications | 9                     |
| Hansen <i>et al.</i> (1999)                             | Non                               | Chasse    | Non                                    | Densité                         | Cong. réelle             | <0                   | 1%                   | 16                    |
| Yen et Adamowicz (1994)                                 | Non                               | Chasse    | Non                                    | Densité                         | Cong. réelle             | <0                   | 1%                   | 27                    |
| Salanié (2006)  | Non                               | Pêche     | Non                                    | Densité                         | Cong. réelle             | Q <sup>d</sup>       | 1%                   | 28                    |
| O'Hara (2007a, 2007b)                                   | Non pertinent                     | Escalade  | Oui                                    | Densité relative                | Cong. anticipée          | <0                   | 5%                   | 7                     |
| Timmins et Murdock (2007)                               | Traitée                           | Pêche     | Oui                                    | Densité relative                | Cong. anticipée          | <0                   | 1%                   | 569                   |
| Boxall <i>et al.</i> (2005)                             | VI <sup>a</sup>                   | Randonnée | Non                                    | Indicatrices                    | Cong. anticipée          | <0                   | 10%                  | 5                     |
| Boxall <i>et al.</i> (2005), Boxall et Adamowicz (2000) |                                   |           |  | Indicatrices (VI <sup>a</sup> ) | Cong. anticipée          | <0 <sup>c</sup>      | 1%                   | 5                     |
| Hindsley <i>et al.</i> (2007)                           | VI <sup>a</sup>                   | Plage     | Non                                    | Quantitatif                     | Cong. réelle             | 1%                   | 7                    | 2                     |

<sup>a</sup> Variables instrumentales

<sup>b</sup> Non significatif

<sup>c</sup> Effet plus important dans ce cas qui prend en compte l'endogénéité

<sup>d</sup> Quadratique

Une autre explication peut cependant être avancée. Dans les RUM, les variables explicatives observées constituent la partie déterministe du modèle tandis que les variables explicatives non observées sont reléguées dans la partie aléatoire ou terme d'erreur. En effet, le terme d'erreur contient tout ce qui n'est pas observé par le chercheur. Or, si des attributs désirables ou au contraire répulsifs n'ont pas pu être intégrés explicitement dans la partie déterministe du modèle, alors ils se retrouvent dans le terme d'erreur. Si ces attributs expliquent en partie le choix de sites, ils influencent la fréquentation et donc la congestion. Il en résulte une corrélation entre la congestion et le terme d'erreur. En l'absence de contrôle de cette endogénéité au sens économétrique, les valeurs des paramètres estimés sont biaisées. De plus, la congestion n'est pas simplement une variable explicative du choix de site, elle en est également le résultat. Les individus choisissent un site en fonction de ses caractéristiques et de la congestion qu'ils anticipent. La congestion réelle est le résultat de ce processus de choix. Elle est alors issue d'un équilibre de Nash en probabilités. Cette information supplémentaire, si elle est intégrée dans l'estimation peut améliorer la qualité du modèle estimé. La section suivante explique comment prendre en compte cette information dans l'estimation du modèle.

### **3. Estimation du modèle à l'équilibre de congestion**

Dans les modèles à utilité aléatoire, le choix d'un site récréatif n'est pas connu avec certitude mais selon une probabilité, résultat de la maximisation individuelle de l'utilité  $V_{ij}$ . Dans ce processus, les individus choisissent un site récréatif étant données leurs anticipations sur le comportement de visite des autres individus. Si on suppose que les individus anticipent rationnellement les comportements de visites des autres individus, le vecteur des niveaux de congestion anticipée aux différents sites est le même pour tous les individus et égal au vecteur des espérances de congestion (Timmins et Murdock, 2007). La congestion anticipée est alors égale à la somme des probabilités de visiter un site par tous les individus :

$$C_j = \sum_i p_{ij} \quad (2)$$

En supposant que le terme d'erreur suit une loi de Gumbel, on se place dans le cadre traditionnel du modèle logit multinomial et les probabilités de visite s'écrivent:

$$P_{ij} = \frac{e^{U_{ij}(TC_{ij}, X_j, \sum_i P_{ij})}}{\sum_k e^{U_{ik}(TC_{ik}, X_k, \sum_i P_{ik})}} \quad (3)$$

La congestion anticipée est alors définie de manière implicite dans le modèle de choix : elle représente la congestion anticipée telle qu'elle est calculée à partir du modèle lui-même. Le choix de visiter un site en présence de congestion devient le résultat d'un équilibre de Nash en probabilités (Boxall et Adamowicz, 2000 ; Timmins et Murdock, 2007 ; Bayer et Timmins, 2007), qui correspond au point fixe de l'équation (3). Bayer et Timmins (2005) ont montré l'existence de cet équilibre. Cependant, son unicité n'est prouvée que dans le cas d'un effet négatif de congestion, c'est-à-dire quand l'augmentation du nombre de visiteurs diminue l'utilité individuelle. Dans le cas d'un impact positif de la congestion, dit effet d'agglomération, une multiplicité d'équilibres peut apparaître surtout quand l'effet est fortement positif.

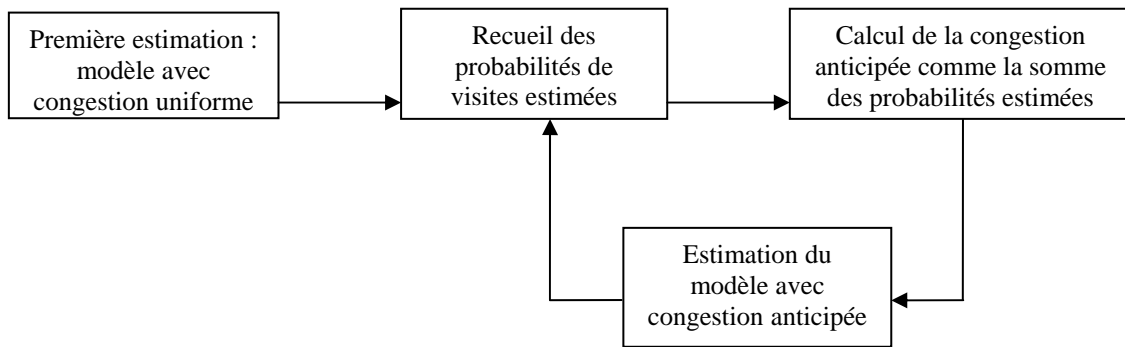
Tenir compte de ce que la congestion est le résultat d'un équilibre de Nash en probabilités doit être considéré. Certains modèles de choix assurent, par construction, la cohérence du modèle estimé avec l'équilibre de Nash observé (l'ensemble des visites observées étant supposé s'analyser dans ce cadre), d'autres non. Tout d'abord, imposer que les estimateurs des paramètres du modèle doivent respecter les conditions de l'équilibre de Nash, à savoir que les niveaux de congestion utilisés en tant que déterminants des choix des visiteurs doivent coïncider avec les fréquences observées des visites (réelles ou espérées), accroît l'efficacité de ces estimateurs dès lors que les hypothèses de comportement aboutissant à l'équilibre de Nash sont valides. Ensuite, et peut-être surtout, l'utilisation du modèle estimé pour des exercices de simulation d'effets de politiques de régulation (voire des calculs d'optima sociaux) nécessitent une situation initiale à l'équilibre. On analyse typiquement les effets d'instruments de régulation en comparant les propriétés des équilibres initiaux, avant la mise en œuvre de l'instrument, et finaux, résultant de la mise en œuvre de l'instrument, c'est-à-dire dans le cadre d'analyse de statique comparative.

Sans l'estimation de constantes spécifiques aux sites, la congestion estimée par le modèle n'a aucune raison d'être égale à la congestion anticipée, introduite comme variable explicative.

Bien entendu, imposer la cohérence de la congestion estimée avec la congestion anticipée complique singulièrement l'estimation des paramètres du modèle de choix ainsi défini. Néanmoins, la procédure itérative proposée par O'Hara (2007a), elle-même inspirée de celle proposée par Bayer et Timmins (2007), est plutôt intuitive. Elle définit les estimateurs des paramètres des modèles de choix considérés comme des estimateurs du maximum de vraisemblance contraint par le respect de l'équilibre de Nash, ce qui, en fait, revient à choisir les estimateurs du maximum de vraisemblance parmi ceux qui vérifient l'équilibre de point fixe que constitue l'équilibre de Nash. L'équilibre de point fixe est unique, dès lors que l'effet de la congestion est négatif sur le choix des sites, c'est-à-dire quand les visiteurs évitent les sites trop encombrés (Bayer et Timmins, 2005). Cette condition assure la convergence numérique de la procédure itérative de O'Hara (2007a).

L'approche proposée par O'Hara (2007a) s'avère particulièrement utile dans ce contexte, et ce, même si elle n'a pas été développée explicitement pour assurer la cohérence des estimations avec les conditions de l'équilibre. En définissant la mesure de congestion utilisée dans le modèle comme la congestion anticipée telle qu'elle est définie à partir du modèle lui-même, cette approche garantit la cohérence des résultats obtenus avec l'équilibre de Nash en moyenne. Cette approche a, en outre, un autre avantage. Elle n'utilise que les données de l'échantillon disponible, c'est-à-dire qu'elle ne nécessite pas de mesures spécifiques des niveaux de congestion. Dans le cas où la congestion a un effet négatif sur la probabilité de choix, l'équilibre résultant est unique, ce qui suggère l'utilisation d'une procédure itérative pour le calcul, dont la logique est décrite par la Figure 1.

O'Hara (2007a) définit cette procédure comme suit. La première étape consiste en l'estimation du modèle de choix en supposant une répartition uniforme des visiteurs sur les sites considérés. On estime donc le modèle de choix de sites en définissant la congestion comme le nombre d'individus de l'échantillon divisé par le nombre de sites de l'espace de choix. Cette estimation permet ensuite de calculer une première approximation des congestions anticipées en sommant les probabilités de visites estimées. Le processus itératif peut alors commencer. Il consiste à estimer, à chaque étape  $t$ , un estimateur des vecteurs de paramètres du modèle à partir du modèle de choix décrit précédemment en approximant les niveaux de congestion anticipée grâce aux dernières estimations de paramètres obtenues. Le processus itératif converge, numériquement, dès lors que le vecteur de paramètres estimé à l'étape  $t$  et celui estimé à l'étape  $t-1$  sont jugés suffisamment proches.

**Figure 1 : Procédure itérative d'O'Hara (2007a)**

Bayer et Timmins (2005) ont montré que cette procédure conduit à un équilibre unique et stable si le paramètre de la congestion est négatif. En revanche, si le paramètre est positif, il peut exister plusieurs équilibres, stables ou instables selon les valeurs initiales. Malheureusement, il n'y a aucun moyen de connaître *a priori* le signe du paramètre de congestion et donc de savoir si la procédure convergera vers un équilibre unique (Bayer et Timmins, 2005).

O'Hara (2007a) estime un modèle de choix de sept parcours d'escalade. Il explique que ces alternatives sont très homogènes, de sorte que le problème d'endogénéité de la congestion au sens économétrique ne se pose pas. En effet, lorsque des mesures fines et précises des caractéristiques des alternatives sont disponibles et/ou lorsque les alternatives considérées sont homogènes, le problème d'endogénéité économétrique lié à la congestion tend à diminuer. De fait, ce problème d'endogénéité disparaît totalement lorsqu'il est possible de supposer que, compte-tenu de l'information disponible pour caractériser les alternatives et compte-tenu de l'homogénéité même de ces alternatives, la partie non observée des valeurs des sites dépend des préférences des individus considérés. L'hétérogénéité des alternatives est alors complètement expliquée dans la partie déterministe, observée, de l'utilité. Dans ce cas, les termes d'erreur des utilités sont « *purement exogènes* » au sens où ils sont essentiellement le reflet de l'hétérogénéité des préférences des individus, voire d'événements indépendants vis-à-vis des caractéristiques des sites, comme les effets d'émotions passagères, la proximité d'un site par rapport au lieu de résidence d'un ami à visiter, etc. Ceci revient alors à dire qu'il n'est plus possible d'extraire des termes d'erreur des utilités conférées par les alternatives, une partie qui serait commune à tous les individus, c'est-à-dire des effets spécifiques des alternatives considérées. La potentielle corrélation entre la partie déterministe et la partie aléatoire disparaît.

La procédure itérative développée par O'Hara (2007a) permet d'estimer le modèle à l'équilibre de congestion, *i.e.*, de prendre en compte l'endogénéité économique de la congestion, mais ne



contrôle pas la potentielle endogénéité économétrique. Par conséquent, il est nécessaire de s'assurer que ce problème d'endogénéité au sens économétrique ne se pose pas avant d'appliquer cette procédure d'estimation. Puisque les alternatives d'O'Hara (2007a) sont homogènes, celui-ci peut se concentrer sur le calcul de l'équilibre et sur l'estimation des paramètres de son modèle à cet équilibre. A l'issue de cette procédure, son paramètre estimé de la congestion est négatif comme pour Timmins et Murdock (2007). Dans les modèles qui introduisent la congestion sans tenir compte de son endogénéité économique, les impacts sur le bien-être de politiques qui modifient le niveau de fréquentation ou sa répartition entre les sites sont mesestimés. Ainsi, O'Hara (2007a) montre que si on ne recalcule pas les niveaux de congestion lors de la simulation de la réouverture d'un site d'escalade, l'impact sur le bien-être de cette extension de l'espace de choix est sous-estimée. A l'inverse, l'impact sur le bien-être d'une politique d'amélioration de la qualité est surestimée quand elle ne tient pas compte des coûts de congestion induits par la hausse de fréquentation qui en résulte (résultat obtenu également par Leplat et Le Goffe, 2009, de manière analytique).

Dans un autre article, O'Hara (2007b) utilise aussi un modèle lié à une fonction de comptage pour expliquer la fréquence des visites et les choix de site. La congestion est désormais définie par site et par occasion de choix. En itérant, il obtient des résultats similaires à ceux de son précédent modèle : un effet négatif de la congestion et donc une sous-estimation de la valeur d'un site récréatif si la congestion est omise. Il est intéressant de remarquer que si l'approche de O'Hara (2007a, 2007b) est correcte pour mesurer des effets de congestion anticipée, elle n'est toutefois pas adaptée à son analyse puisqu'il met en avant que les individus choisissent leur parcours d'escalade en fonction de la congestion observée, les grimpeurs pouvant observer *de visu* la congestion des différentes voies avant de choisir celle qu'ils vont emprunter. Leur choix de site est alors postérieur à l'observation de la congestion. Dans ce cas, ce sont les mesures réelles ou perçues qui constituent les variables explicatives pertinentes (Jakus et Shaw, 1997), alors que O'Hara (2007a, 2007b) utilise la congestion anticipée. Or, dans le modèle considéré, il n'y a coïncidence entre les choix réels et les choix observés « *qu'en moyenne* ». Autrement dit, si les termes de congestion calculée par la procédure itérative sont des mesures correctes des niveaux de congestion anticipée, ce ne sont pas nécessairement des mesures correctes de congestion réelle. Un modèle avec constantes spécifiques aux alternatives assurerait la coïncidence exacte entre la congestion réelle et la congestion calculée à partir du modèle.

Toutefois, la procédure d'Ohara, si elle permet d'estimer le modèle à l'équilibre de congestion, ne s'attaque pas au contrôle de l'endogénéité économétrique qui se pose dans les cas où les

alternatives ne sont pas homogènes et où il n'est pas possible de disposer d'une description suffisamment fine des alternatives pour écarter tout soupçon de corrélation entre la partie déterministe et le terme d'erreur. Plusieurs articles proposent d'utiliser les variables instrumentales pour résoudre cette difficulté (Boxall *et al.*, 2005 ; Boxall et Adamowicz, 2000 ; Hindsley *et al.*, 2007).

#### **4. Traitement de l'endogénéité par variables instrumentales dans les modèles RUM**

Dans les modèles estimés par les moindres carrés ordinaires, le problème de corrélation entre une variable explicative et le terme d'erreur est résolu en utilisant des variables instrumentales. Ces variables doivent satisfaire deux critères :

- elles doivent être corrélées à la variable endogène ;
- elles doivent être non corrélées au terme d'erreur.

Boxall et Adamowicz (2000) et Hindsley *et al.* (2007) ont calqué leur méthode de contrôle de l'endogénéité économétrique sur cette procédure développée dans le cadre des modèles linéaires. Ainsi, ils utilisent des fonctions d'attributs exogènes pour construire un instrument de la congestion. Dans une application à la randonnée, Boxall et Adamowicz (2000) estiment deux modèles, le premier avec une congestion anticipée obtenue par questionnaire tandis que le second modèle vise à contrôler l'endogénéité de la congestion par une procédure de variables instrumentales. L'instrument est construit en régressant la congestion anticipée sur des attributs de sites tels que la taille du parking, le nombre d'accès au site et des caractéristiques individuelles. Dans les deux cas, la congestion a un impact négatif sur l'utilité et donc sur la probabilité de choisir un site, mais cet effet est moins prononcé sans instrument et plus faiblement significatif.

Hindsley *et al.* (2007) adoptent une approche similaire dans une application relative au choix de plages mais des variables météorologiques servent également d'instruments pour la congestion. En effet, l'impact de l'encombrement des sites est déterminé certes par la répartition des visites entre les sites, *i.e.*, la congestion relative, mais également par le nombre total de visites, le niveau global de congestion intégré dans l'analyse grâce au modèle de comptage. Ils obtiennent ensuite un effet quadratique de la congestion sur l'utilité des usagers : pour des niveaux faibles de congestion, l'utilité des usagers des plages augmente jusqu'à atteindre un maximum pour diminuer ensuite.

La limite principale de cette stratégie est qu'elle est issue de la technique des variables instrumentales dans les modèles linéaires où les propriétés de cette approche sont bien connues. Mais ce n'est pas le cas dans les modèles de choix. Utiliser des variables instrumentales dans les modèles de choix discrets ne garantit pas l'orthogonalité entre l'instrument et le terme d'erreur. Dans ce contexte, il est difficile de savoir si l'endogénéité est corrigée. Hindsley *et al.* (2007) reconnaissent cette faiblesse mais argumentent en faveur de leur approche en disant qu'elle est préférable à l'omission pure et simple de la congestion. Il est vrai que la littérature sur la question de l'endogénéité économique et économétrique de la congestion n'est pas stabilisée. Les articles, publiés ou non, qui traitent de cette question sont récents. A part Boxall et Adamowicz (2000), ils sont tous postérieurs à 2005 (cf. Tableau 2).

Bayer et Timmins (2005) proposent une stratégie générale à l'ensemble des modèles de choix pour instrumenter la congestion : ils recommandent d'utiliser la part prédite des individus ayant choisi le site  $j$  estimée uniquement en fonction des attributs exogènes des sites et des caractéristiques individuelles. Cette probabilité dépend à la fois des attributs du site  $j$  et des attributs des autres sites. Elle est donc idéale pour instrumenter la congestion. En effet, le choix d'un site  $j$  est déterminé non seulement par ses attributs, mais également par les attributs des sites substitués, puisque l'individu les compare dans ce processus de choix. Les attributs exogènes des autres sites influencent l'équilibre de fréquentation mais pas l'utilité de la visite au site. Par conséquent, ils ne sont pas corrélés avec le terme d'erreur.

L'efficacité de cet instrument repose principalement sur le caractère discriminant des caractéristiques mesurées des sites, par rapport aux choix des visiteurs. En effet, si les sites sont tous très similaires ou si les attributs les différenciant n'ont pu être mesurés, les variables exogènes expliquent peu le choix des sites. L'instrument est alors peu explicatif. En outre, si les caractéristiques mesurées des individus interagissent peu avec les attributs mesurés des sites dans les probabilités de choix des sites visités, l'instrument proposé ne dépend que des attributs du site considéré. L'identification de ce que l'on peut nommer les effets directs de congestion (les effets de la congestion qui jouent en interaction avec les caractéristiques des attributs sont alors qualifiés d'indirects) repose alors sur des différences de formes fonctionnelles. En effet, puisque l'instrument est une combinaison non linéaire des attributs observés des sites, l'effet de congestion mesuré peut essentiellement résulter d'effets non linéaires des attributs observés des sites sur les niveaux d'utilité, c'est-à-dire qu'il peut faire apparaître un effet de congestion significatif là où il n'existe pas. Cette « identification par les formes fonctionnelles » est de fait assez fragile car elle repose sur l'hypothèse selon laquelle le modèle proposé pour les choix

d'alternatives est exact, c'est-à-dire que le terme d'erreur est bien indépendant et identiquement distribué selon la loi de Gumbel et que les formes fonctionnelles des utilités latentes utilisées sont également correctes. Bayer et Timmins (2007) décrivent, plus ou moins en détails, des cas où les instruments qu'ils proposent assurent l'identification des effets de congestion sans recours exclusif à des arguments de « différences de formes fonctionnelles ». L'introduction des coûts de transport dans l'instrument peut considérablement améliorer celui-ci –à condition que ceux-ci ne soient pas endogènes (Timmins et Murdock, 2007). En effet, les coûts de transport interagissent fortement avec les attributs mesurés des sites dans les probabilités de choix des sites visités. L'hétérogénéité des lieux de résidence des visiteurs potentiels crée d'importantes variations exogènes des coûts de transport qui procurent leur efficacité aux instruments proposés.

On a vu que le traitement de l'endogénéité économétrique de la congestion par variables instrumentales dans les modèles de choix discrets n'est pas rigoureux car il ne garantit pas l'orthogonalité entre l'instrument et le terme d'erreur. De plus, la cohérence avec l'équilibre de Nash en probabilités n'est pas assurée. La procédure d'O'Hara (2007a) résout cette difficulté mais la corrélation entre le terme d'erreur et la congestion reste un problème si les alternatives sont hétérogènes. Bayer et Timmins (2007) proposent une autre procédure inspirée de la méthode de Berry *et al.* (1995), développée initialement pour l'analyse des marchés des biens différenciés, qui permet d'estimer le modèle à l'équilibre de congestion tout en contrôlant l'endogénéité économétrique. Cette méthode a été utilisée par Timmins et Murdock (2007) pour estimer un modèle de choix de site récréatif.

## **5. Méthode de Berry, Levinsohn et Pakes (BLP) : une seule méthode pour résoudre deux difficultés**

L'approche de Bayer et Timmins (2007) repose essentiellement sur l'idée que les caractéristiques non observées des alternatives jouent un rôle majeur dans les choix des individus et, par là-même, génèrent un problème d'endogénéité économétrique de la mesure de la congestion dans le modèle de choix. On retrouve cette idée dans les modèles de choix de biens différenciés (Berry, 1994 ; Berry *et al.*, 1995) : les choix des individus dépendent des caractéristiques non observées (ou non mesurées) de ces biens. Ces caractéristiques sont donc des déterminants essentiels des prix des biens en raison de leur formation via les équilibres de marché. Ceci implique finalement que les prix des biens sont fortement liés aux caractéristiques non observées des biens.

Pour expliquer la méthode BLP, il convient d'abord de travailler sur l'expression de l'utilité  $V_{ij}$  dans laquelle on introduit explicitement la congestion  $C_j$ .

$$V_{ij} = U_{ij}(X_j, Z_i, TC_{ij}, C_j) + \varepsilon_{ij} \quad (4)$$

$$\text{avec } U_{ij} = \beta X_j + \phi Z_i X_j + \delta TC_{ij} + \alpha C_j \quad (5)$$

Les paramètres  $\beta, \phi, \delta, \alpha$  constituent les paramètres à estimer. On divise le terme d'erreur  $\varepsilon_{ij}$  en deux parties : la première,  $\zeta_j$ , qui ne dépend que des sites et qui correspond aux attributs non observés des sites  $j$ , et la seconde,  $\varepsilon_{ij}$ , qui dépend du site et de l'individu. L'utilité devient :

$$V_{ij} = U_{ij}(X_j, Z_i, TC_{ij}, C_j) + \zeta_j + \varepsilon_{ij} \quad (6)$$

En séparant l'utilité  $U_{ij}$  entre sa partie qui ne dépend que des sites,  $\overline{U}_j$ , et sa partie qui dépend de l'individu et des sites,  $\tilde{U}_{ij}$ , le modèle s'écrit alors :

$$V_{ij} = \overline{U}_j(X_j, C_j) + \zeta_j + \tilde{U}_{ij}(TC_{ij}, Z_i X_j, Z_i C_j) + \varepsilon_{ij} \quad (7)$$

La partie déterministe  $\overline{U}_j$  et les aléas  $\zeta_j$ , qui dépendent uniquement du site, sont remplacés par des constantes spécifiques au site,  $\gamma_j$  (*Alternative Specific Constants* ou ASC). Le modèle devient alors :

$$V_{ij} = \gamma_j + \tilde{U}_{ij}(TC_{ij}, Z_i X_j, Z_i C_j) + \varepsilon_{ij} \quad (8)$$

$$\text{où } \gamma_j = \overline{U}_j(X_j, C_j) + \zeta_j \quad (9)$$

Le modèle peut alors être estimé par la méthode du maximum de vraisemblance. L'une des constantes  $\gamma_j$  doit être supposée nulle afin que le modèle soit identifiable. La probabilité de visiter le site  $j$  par l'individu  $i$  s'écrit :

$$P_{ij} = \frac{\exp(\gamma_j + \tilde{U}_{ij}(TC_{ij}, Z_i X_j, Z_i C_j))}{\sum_k \exp(\gamma_k + \tilde{U}_{ij}(TC_{ik}, Z_i X_k, Z_i C_k))} \quad (10)$$

Tous les effets liés au site sont maintenant intégrés dans la constante  $\gamma_j$  y compris les attributs non observés  $\zeta_j$ . Il n'y a alors plus de problème d'endogénéité économétrique dans le modèle de choix puisque les aléas  $\varepsilon_{ij}$  sont désormais indépendants des variables explicatives. L'introduction de ces constantes spécifiques aux sites,  $\gamma_j$ , garantit l'égalité entre les fréquences moyennes de choix observées dans l'échantillon et les probabilités estimées de choix de ces mêmes alternatives. De fait, dans ce cas, le modèle est estimé à l'équilibre de Nash en probabilités. Il convient désormais de régresser la constante sur les attributs des sites pour connaître leur impact sur l'utilité. Dans cette régression, la congestion est corrélée avec le terme d'erreur, d'où un problème d'endogénéité économétrique, mais l'artifice de la méthode BLP a permis de transférer cette endogénéité d'un modèle de choix à un modèle linéaire où le problème peut être résolu par la technique des variables instrumentales, bien connue dans ce cadre.

L'instrument retenu est celui proposé par Bayer et Timmins (2007), qui utilise les attributs observés des sites excepté la congestion. Seulement, à ce stade, l'impact des différents attributs des sites est compris dans la constante  $\gamma_j$ . On ne peut donc pas séparer l'effet de la congestion de ceux des autres attributs. Il faut alors régresser la constante sur les attributs des sites mais en contrôlant l'endogénéité de la congestion. Pratiquement, Bayer et Timmins (2007) suggèrent d'utiliser comme instrument de la congestion la moyenne des probabilités de visite calculée en négligeant a) l'effet de congestion et b) les attributs non observés des sites. Cet instrument est donc essentiellement construit à partir des attributs observés des sites, des caractéristiques des visiteurs et des paramètres du modèle sauf  $\alpha$ , le paramètre de la congestion. Le calcul de cet instrument est problématique : la valeur de  $\beta$  est nécessaire au calcul de l'instrument, dont l'objectif est précisément d'estimer la valeur de  $\beta$ . De fait, n'importe quelle (ou presque) valeur

de  $\beta$  peut servir pour la construction d'un instrument valide. Néanmoins l'instrument est d'autant plus efficace que la valeur de  $\beta$  utilisée est proche de sa « vraie » valeur.

Bayer et Timmins (2007) proposent la procédure itérative suivante pour le calcul des estimateurs de  $\alpha$  et  $\beta$  à partir des meilleurs instruments possibles pour la congestion (cf. Figure 2) :

1. On régresse d'abord les constantes spécifiques aux sites sur les attributs des sites en ne tenant pas compte de l'endogénéité économétrique de la congestion  $C_j$ , observée dans l'échantillon. La constante  $\tau$  est introduite de façon à capturer l'effet de la fixation à zéro d'une des constantes spécifiques au site  $\gamma_j$  :

$$\gamma_j = \tau + \beta X_j + \alpha C_j \quad (11)$$

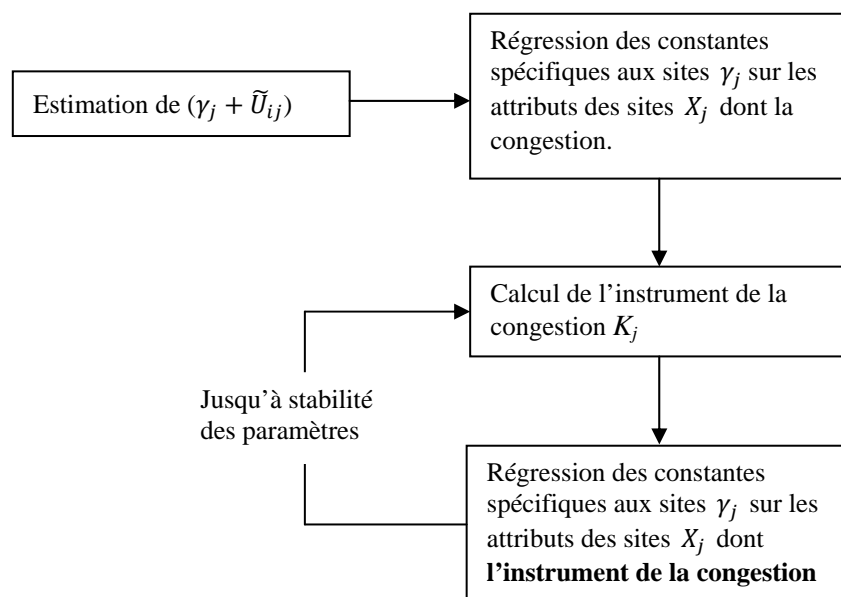
Les paramètres de l'équation (11) étant connus, on peut alors calculer l'instrument, noté  $K_j$ . Dans la probabilité de visite de l'équation (10), la constante est remplacée par son expression, définie par l'équation de régression, mais en enlevant la congestion pour calculer l'instrument. Celui-ci s'écrit alors :

$$K_j = \frac{1}{N} \sum_i \frac{\exp(\tau + \beta(X_j) + \phi(Z_i)X_j + \delta TC_{ij})}{\sum_k \exp(\tau + \beta(X_k) + \phi(Z_i)X_k + \delta TC_{ik})} \quad (12)$$

2. On réestime l'équation (11) par la procédure des moindres carrés<sup>2</sup> mais en remplaçant la congestion  $C_j$  par son instrument  $K_j$ .
3. L'instrument de la congestion est recalculé et les étapes précédentes sont reconduites jusqu'à stabilisation du processus.

Les paramètres  $\tau$ ,  $\phi$ ,  $\delta$ ,  $\alpha$  et  $\beta$  décrivent alors complètement l'utilité.

<sup>2</sup> Si toutes les alternatives sont choisies au moins une fois, les constantes sont toutes positives et il est possible de les estimer par la méthode des doubles moindres carrés. Dans le cas contraire, une procédure plus complexe est nécessaire : voir Timmins et Murdock (2007) pour plus de détails.

**Figure 2 : Procédure de Bayer et Timmins (2007)**

L'approche de Bayer et Timmins (2007) pose également quelques problèmes d'ordre technique. Elle a été mise en oeuvre par Timmins et Murdock (2007) dans un modèle de choix de sites récréatifs. Pour que la régression des constantes spécifiques aux sites soit robuste, il faut disposer d'un nombre important d'alternatives (559 dans le cas de Timmins et Murdock, 2007). Ces constantes doivent en général être calculées en utilisant l'opérateur contractant de Berry (1994) qui, lui-même, utilise la propriété des  $\gamma_j$  dans les modèles de choix qui égalisent la proportion observée avec celle estimée par le modèle. Timmins et Murdock (2007) adaptent la méthode de Bayer et Timmins (2007) au cas de la congestion dans un modèle de choix de site récréatif appliqué à la pêche. Ils estiment leur modèle à l'équilibre de Nash et contrôlent l'endogénéité économétrique de la congestion. Ils mettent en évidence un effet négatif significatif de la congestion. Dans ce cas, en ignorant l'endogénéité économétrique de cette variable, la congestion a un impact positif sur l'utilité et les autres paramètres estimés sont biaisés. Ils calculent ensuite la valeur d'un site récréatif et montrent qu'elle est sous-estimée de plus de 50% si la congestion n'est pas introduite. La valeur d'un site se calcule par la différence entre le bien-être collectif quand ce site est présent dans l'espace de choix et quand il en est absent. Or, quand la congestion est omise, l'impact (négatif chez Timmins et Murdock, 2007) de la hausse de fréquentation sur le bien-être des visiteurs des autres sites récréatifs n'est pas pris en compte, ce qui entraîne la sous-estimation de la valeur du site.



## 6. Illustration

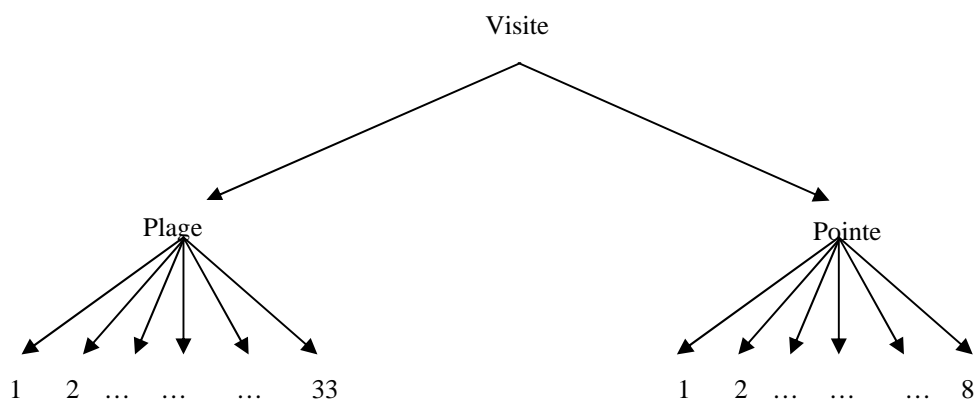
Afin d'illustrer le propos, les techniques d'estimation précédentes sont utilisées sur un même jeu de données. Après avoir présenté la structure du modèle de choix discrets répétés utilisé et les données disponibles, une approche ascendante est adoptée : un modèle sans congestion est d'abord estimé puis la congestion est introduite dans les modèles suivants, de manière exogène avec un indicateur de congestion observée, puis de manière endogène avec le calcul de l'équilibre de congestion (on a différencié dans les résultats les paramètres estimés à l'issue de la première itération et ceux estimés une fois la convergence atteinte). Enfin, la dernière estimation est l'application de la méthode de Berry *et al.* (1995) à nos données, dite méthode BLP.

### 6.1 Structure du modèle et présentation des données

Cette illustration est conduite sur des données d'enquête concernant les visites effectuées par 318 résidents dans 41 sites littoraux finistériens et recueillies durant trois semaines de l'été 2007. Cette base de données comprend à la fois des données individuelles (932 observations, car les résidents interrogés pouvaient déclarer plusieurs visites), des données sur les sites visités ainsi que quelques données météorologiques sur la période d'enquête.

Etant donné que deux types de sites très différents, des pointes et des plages, figurent dans l'espace de choix des résidents, on a choisi d'estimer un modèle de choix discrets hiérarchisés, dont la Figure 3 montre la structure. Les plages sont au nombre de 33 tandis que les pointes constituent les 8 sites restants.

**Figure 3 : Structure du modèle estimé**



La description des sites comprend une variable de congestion issue du traitement de photographies recueillies grâce à un survol aérien conduit le samedi le plus ensoleillé de la période d'enquête. Cette mesure de la congestion est un indicateur de la congestion maximale observée sur la période d'enquête. Le Tableau 3 présente les variables utilisées dans les estimations.

**Tableau 3 : Description des caractéristiques individuelles et des attributs des sites**

|                                       | Description  | Moyenne | Ecart-type |
|---------------------------------------|--|---------|------------|
| <i>Caractéristiques individuelles</i> |  |         |            |
| Groupe                                | Indicatrice =1 si l'individu est venu sur le site en famille ou en groupe le                     | 0,52    | 0,50       |
| Vacances                              | jour de la visite  | 0,48    | 0,50       |
| CSP sup                               | Indicatrice =1 si l'individu est en congé le jour de la visite                                   | 0,37    | 0,67       |
| Bac+2                                 | Indicatrice =1 si l'individu appartient à la catégorie cadre ou profession intermédiaire         | 0,43    | 0,50       |
| Type Rando                            | Indicatrice =1 si la personne interrogée a une formation égale ou supérieure à un bac+2          | 0,21    | 0,41       |
| Coût de transport                     | Indicatrice =1 si la personne interrogée pratique la randonnée                                   | 10,41   | 9,09       |
|                                       | Coût de transport calculé par l'auteur pour se rendre au site j                                  |         |            |
| <i>Attributs des jours de visites</i> |  |         |            |
| Pluie                                 | Indicatrice=1 s'il a plu le jour de la visite  | 0,33    | 0,47       |
| <i>Attributs des sites</i>            |  |         |            |
| Parking                               | Indicatrice =1 si le site dispose d'un parking   | 0,80    | 0,39       |
| Plage                                 | Indicatrice =1 si le site est une plage  | 0,81    | 0,39       |
| Pointe                                | Indicatrice =1 si le site est une pointe   | 0,19    | 0,37       |
| Loin                                  | Indicatrice =1 si le parking est situé plus de 300 mètres du site                                | 0,09    | 0,29       |
| Grd Parking                           | Indicatrice =1 si le site dispose d'un parking de plus de 300 places                             | 0,27    | 0,44       |
| Camping                               | Indicatrice =1 si le site est proche d'un camping (<500m)  | 0,58    | 0,49       |
| Environnement naturel                 | Indicatrice =1 si l'environnement du site est naturel <sup>3</sup>                               | 0,56    | 0,50       |
| Accès difficile                       | Indicatrice =1 si l'accès au site est difficile <sup>4</sup>                                     | 0,15    | 0,37       |
| Jeux                                  | Indicatrice =1 si le site est équipé de jeux ou d'un parcours de santé                           | 0,10    | 0,29       |
| Restaurant                            | Indicatrice = 1 si un restaurant ou un café est présent dans la bande des 200m                   | 0,34    | 0,47       |
| Surveillée                            | Indicatrice =1 si le site est surveillée   | 0,46    | 0,50       |
| Congestion observée (M2)              | Part des individus sur le site (comptabilisés grâce au survol aérien)*1000 par mètre de linéaire | 1,15    | 1,27       |
| Congestion à l'équilibre (M3)         | Probabilité de visite au site*1000 par mètre de linéaire   | 1,33    | 1,47       |
| Instrument de la congestion (M4)      | Probabilité de visite au site*1000 par mètre de linéaire   | 1,13    | 1,23       |

Les résidents  $i = 1, \dots, n$  qui choisissent l'alternative  $j = 1, \dots, J$  ( $J = 41$ ) le jour  $t = 1, \dots, T$  ( $T = 21$ ), obtiennent l'utilité  $V_{ijt}$  :

$$V_{ijt} = \beta(X_j) + \zeta Z_i + \phi X_j + \delta TC_{ij} + \alpha C_j + \theta Q_t + \varepsilon_{ijt} \quad (13)$$

avec:

- $X_j$  le vecteur des attributs de l'alternative  $j$  ;

<sup>3</sup> Cette variable a été construite en recoupant les informations réglementaires sur les zones naturelles protégées, les couches d'informations géographiques de la DIREN et les photographies issues du survol aérien.

<sup>4</sup> L'accès est considéré comme difficile si peu de parkings sont disponibles et que le site n'est desservi ni par des voies principales ni par des voies secondaires d'après les cartes IGN.

- $Z_i$  le vecteur des caractéristiques de l'individu  $i$  ;
- $Q_t$  le vecteur des attributs du jour  $t$  ;
- $TC_{ij}$  le coût de déplacement de l'individu  $i$  au site  $j$  (le coût de déplacement pour la non-participation est considéré comme nul) ;
- $C_j$  la congestion au site  $j$  ;
- $\beta, \zeta, \phi, \delta, \theta$  et  $\alpha$ , les paramètres à estimer ;
- $\varepsilon_{ijt}$  le terme aléatoire.

Sous l'hypothèse que les aléas sont indépendants et identiquement distribués selon une loi de Gumbel, la probabilité de choisir l'alternative  $j$  pour l'individu  $i$ , le jour  $t$  s'écrit :

$$p_{ijt} = \frac{e^{U_{ijt}}}{\sum_k e^{U_{ikt}}} \quad (14)$$

## 6.2 Résultats des estimations

Les résultats des estimations des différents modèles sont détaillés dans les Tableaux 4 et 5. Les variables retenues pour les estimations, congestion exceptée, sont les mêmes quel que soit le modèle estimé. Dans le premier modèle, la congestion est omise. Le second modèle est estimé avec la variable de congestion observée. Sous l'hypothèse d'anticipations rationnelles, les congestions observées et anticipées (« *expected* » au sens de Jakus et Shaw (1997)<sup>5</sup>) sont très proches à défaut d'être identiques. Malheureusement, cette mesure de la congestion observée est de qualité médiocre, de sorte qu'on ne peut l'assimiler à la congestion anticipée. Le troisième modèle utilise la moyenne des probabilités de choix de sites estimées pour construire un indicateur de la congestion anticipée. Afin de garantir la cohérence avec l'équilibre de Nash, une procédure itérative inspirée d'O'Hara (2007a) est utilisée. Si des variables attractives ou au contraire répulsives non disponibles expliquent le choix des sites, alors le terme d'erreur est corrélé avec la congestion et les paramètres estimés du modèle sont biaisés. Or, aucune des estimations précédentes ne permet de contrôler cette potentielle endogénéité. C'est l'objet de la quatrième estimation qui applique la procédure développée par Bayer et Timmins (2007).

---

<sup>5</sup>Dans leur article, Jakus et Shaw (1997) parlent de congestion « *expected* » si les individus ont des préférences objectives sur la distribution de probabilité de visite des autres individus.

Cependant, cette technique demande un grand nombre d'alternatives, ce qui n'est pas le cas dans cette application. L'introduction des constantes spécifiques aux alternatives garantit donc que les paramètres du choix du type de sites ne sont pas biaisés mais le faible nombre d'alternatives (41) invite à la mesure sur la robustesse de la régression de ces ASC sur les caractéristiques des sites.

**Tableau 4 : Résultats des estimation sur le choix de type de site**

| Choix de type de site<br>(plage ou pointe) | M1<br>congestion<br>omise | M2<br>congestion<br>observée | M3<br>équilibre de<br>congestion | M4<br>BLP                      |
|--|---------------------------|------------------------------|----------------------------------|--------------------------------|
| Nombre d'observations                      | 932                       | 932                          | 932                              | 932                            |
| Rho de Mc Fadden                           | 0.27                      | 0.28                         | 0.27                             | 0.33                           |
| Type d'endogénéité traitée                 | Aucune                    | Aucune                       | Economique                       | Economique et<br>économétrique |
| CSP sup                                    | 0.5001<br>(0.3955)        | 0.4960<br>(0.3958)           | 0.4981<br>(0.3956)               | 0.5380<br>(0.4772)             |
| Bac+2                                      | 1.1550***<br>(0.3890)     | 1.1233***<br>(0.3897)        | 1.1434***<br>(0.3894)            | 1.0132**<br>(0.4472)           |
| Pluie                                      | -0.1829<br>(0.3187)       | -0.2064<br>(0.3193)          | -0.1921<br>(0.3190)              | -0.264<br>(0.3824)             |
| Type Rando                                 | -1.6071***<br>(0.3124)    | -1.6614***<br>(0.3140)       | -1.6286***<br>(0.3133)           | -1.9195***<br>(0.3883)         |
| Vacances                                   | 0.8638**<br>(0.3553)      | 0.8111**<br>(0.3562)         | 0.843**<br>(0.3561)              | 0.6496*<br>(0.3942)            |
| Groupe                                     | 0.4344<br>(0.3015)        | 0.4045<br>(0.3028)           | 0.4223<br>(0.3021)               | 0.3031<br>(0.3608)             |
| Valeur inclusive plages                    | 0.7884***<br>(0.0942)     | 0.7580***<br>(0.0900)        | 0.7875***<br>(0.0939)            | 0.6583***<br>(0.1598)          |
| Valeur inclusive pointes                   | 0.6021***<br>(0.0864)     | 0.5724***<br>(0.0832)        | 0.5995***<br>(0.0864)            | 0.4737***<br>(0.0988)          |

Ecarts-types entre parenthèses ; \*\*\*, \*\*, \* : variables significatives aux niveaux respectifs de 1%, 5% et 10%.

Les quatre modèles présentent des résultats proches en termes de qualité d'ajustement. Le modèle 4 issu de la méthode BLP présente un meilleur ajustement, ce qui est normal car les constantes spécifiques aux alternatives « prennent » tous les effets liés aux sites. L'effet « site » est par conséquent intégralement pris en compte mais n'est pas expliqué dans l'estimation du modèle de choix.

De la même manière, les paramètres estimés relatifs aux variables de choix du type de site visité (caractéristiques individuelles pour la plupart) sont assez robustes entre les différentes estimations. Les paramètres du modèle 4 (méthode BLP) sont quelques peu différents des trois autres modèles, cependant ces différences ne sont pas statistiquement significatives quand les paramètres sont eux-mêmes significatifs.

**Tableau 5 : Résultats des estimations sur le choix de site**

| Choix de site              | M1<br>congestion<br>omise          | M2<br>congestion<br>observée       | M3<br>équilibre de<br>congestion   | M4<br>BLP                                      |
|----------------------------|------------------------------------|------------------------------------|------------------------------------|--|
| R <sup>2</sup>             | Non pertinent                      | Non pertinent                      | Non pertinent                      | 0.61   |
| Type d'endogénéité traitée | Aucune                             | Aucune                             | Economique                         | Economique et<br>économétrique                 |
| Coût de transport          | -0.3452 <sup>***</sup><br>(0.0131) | -0.3463 <sup>***</sup><br>(0.0132) | -0.3450 <sup>***</sup><br>(0.0132) | -0.3448 <sup>***</sup><br>(0.0093)             |
| Constante                  | Non<br>pertinent                   | Non<br>pertinent                   | Non<br>pertinent                   | <b>-0.9476<sup>**</sup></b><br><b>(0.4590)</b> |
| Congestion                 |                                    | 0.1634 <sup>***</sup><br>(0.0298)  | 0.0455 <sup>*</sup><br>(0.0258)    | <b>-0.0754</b><br><b>(0.1582)</b>              |
| Loin*pointe                | 0.2665<br>(0.2920)                 | 0.3136<br>(0.2941)                 | 0.3076<br>(0.2932)                 | <b>0.2265</b><br><b>(0.6065)</b>               |
| Parking                    | 1.1407 <sup>***</sup><br>(0.2092)  | 1.0972 <sup>***</sup><br>(0.2115)  | 1.1146 <sup>***</sup><br>(0.2098)  | <b>0.7728<sup>*</sup></b><br><b>(0.4598)</b>   |
| Camping                    | 0.1892 <sup>*</sup><br>(0.1116)    | 0.1823<br>(0.1154)                 | 0.1649<br>(0.1134)                 | <b>0.3499</b><br><b>(0.3628)</b>               |
| Environnement naturel      | 0.0674<br>(0.0811)                 | 0.2487 <sup>***</sup><br>(0.0904)  | 0.1246<br>(0.0881)                 | <b>0.0393</b><br><b>(0.3527)</b>               |
| Grd parking                | 0.2091 <sup>**</sup><br>(0.0866)   | 0.2257 <sup>***</sup><br>(0.0876)  | 0.2081 <sup>**</sup><br>(0.0867)   | <b>0.4931</b><br><b>(0.3683)</b>               |
| Restaurant                 | 0.6827 <sup>***</sup><br>(0.0767)  | 0.5892 <sup>***</sup><br>(0.0790)  | 0.5812 <sup>***</sup><br>(0.0969)  | <b>0.7407<sup>*</sup></b><br><b>(0.3889)</b>   |
| Surveillée*plage           | 0.5886 <sup>***</sup><br>(0.0840)  | 0.4602 <sup>***</sup><br>(0.0882)  | 0.5916 <sup>***</sup><br>(0.0843)  | <b>1.3284<sup>***</sup></b><br><b>(0.3665)</b> |
| Jeux                       | -0.6151 <sup>***</sup><br>(0.1282) | -0.4792 <sup>***</sup><br>(0.1308) | -0.5497 <sup>***</sup><br>(0.1337) | <b>-0.7575</b><br><b>(0.5301)</b>              |
| Accès difficile            | -0.5411 <sup>***</sup><br>(0.1393) | -0.4059 <sup>***</sup><br>(0.1465) | -0.4854 <sup>***</sup><br>(0.1433) | <b>-0.1853</b><br><b>(0.4292)</b>              |

Ecarts-types entre parenthèses ; <sup>\*\*\*</sup>, <sup>\*\*</sup>, <sup>\*</sup> : variables significatives aux niveaux respectifs de 1%, 5% et 10%.

Les différences de paramètres estimés sont nettement plus marquées en ce qui concernent les variables de choix de site. Notre paramètre d'intérêt, celui de la congestion, est positif et significatif dans les modèles 2 et 3. Cependant, l'ampleur de l'effet de la congestion est quatre fois plus importante dans le modèle avec la variable de congestion observée (0.1634 pour le M2 contre 0,0455 pour le M3). En revanche, dans le modèle 4, le coefficient de la congestion est négatif mais non significatif. On retrouve ici le même résultat que Timmins et Murdoch (2007) : la congestion instrumentée a un impact négatif mais non significatif sur l'utilité. Ce changement de signe peut s'expliquer par l'omission d'une ou plusieurs variables « attractives » dans la partie déterministe du modèle. La corrélation, positive alors, entre la congestion et ces variables tendraient à surestimer le paramètre de congestion. Il est cependant difficile d'être plus affirmatif étant donné la qualité de l'ajustement de la régression linéaire ( $R^2$  ajusté = 0,53). En outre, les deux modèles avec congestion anticipée s'appuient sur les probabilités de visite estimées: dans le modèle 3 la congestion est calculée en agrégeant les probabilités de visite estimées et l'instrument de la congestion du modèle 4 somme les probabilités de visite estimées

en retirant l'impact de la congestion. Or, la congestion est le fait des résidents, mais également des touristes dont les choix de site peuvent différer des résidents. Seul le modèle 2 utilise une congestion qui reflète les deux types de préférences. Malheureusement, on peut s'interroger sur la fiabilité de cette mesure de la congestion sur l'ensemble de la période d'enquête.

Le traitement de la congestion entraîne également de fortes variations pour les autres paramètres explicatifs du choix de site. La variable « parking » est significative à 1% dans les trois premiers modèles mais ne l'est qu'à 10% dans le modèle avec constantes spécifiques aux sites (M4), et avec un coefficient beaucoup plus faible dans ce dernier modèle. La variable « Environnement naturel » est très significative (1%) uniquement dans le modèle 2 avec congestion instrumentée. Les variables « Jeux », « Grand parking » et « accès difficile » sont significatives à 1% ou 5% dans les trois premiers modèles avec des paramètres proches, mais ne sont pas significatives dans le modèle avec congestion instrumentée (M4). La surveillance des plages a un effet positif sur l'utilité dans les quatre modèles mais son impact est nettement plus marqué dans le modèle 4.

Finalement, seuls les effets du coût de transport, de la présence d'un parking ou d'un restaurant sont significativement similaires entre les quatre modèles.

Cette illustration permet d'observer que, selon la manière d'introduire la congestion, des différences importantes apparaissent sur la valeur du paramètre de congestion lui-même, mais également sur les valeurs des paramètres estimés des autres attributs des sites, les paramètres des autres niveaux de choix étant en revanche peu affectés par le traitement de la congestion. Si toutes les procédures d'estimation mises en œuvre ici aboutissent à un paramètre de congestion positif, les différences de valeur observées tant sur le paramètre congestion que sur ceux des variables explicatives du choix de site montrent que la congestion est une variable explicative cruciale dans les modèles explicatifs de choix de site.

Dès lors que le nombre d'alternatives est important, l'approche par les constantes spécifiques aux alternatives développée par Bayer et Timmins (2007) est à privilégier car elle contrôle l'endogénéité économétrique de la congestion et garantit que l'estimation est conduite à l'équilibre de Nash. En revanche, quand le nombre d'alternatives est faible, l'estimation itérative est possible à condition que ces alternatives soient homogènes ou qu'une description détaillée des alternatives permette de lever le soupçon d'endogénéité économétrique. L'avantage de ces deux stratégies d'estimation est qu'elles ne demandent pas de disposer d'une mesure de la congestion et qu'elles permettent de simuler l'impact de politiques publiques sur le bien-être en prenant en compte la variation de congestion qui en résulte. Cependant, si une mesure de

congestion est disponible, que celle-ci est de bonne qualité et que l'échantillon d'individus est représentatif, alors le modèle explicatif de choix de site peut être estimé directement sur cette donnée, sans passer par les procédures décrites précédemment. Il convient de s'assurer que la mesure de la congestion observée ou anticipée est proche de la congestion calculée avec le modèle estimé. Des simulations de politiques publiques peuvent alors être conduites en endogénéisant la congestion par une procédure itérative.

## 7. Conclusion

Depuis la conception du cadre théorique de Fisher et Krutilla (1972) permettant de calculer la congestion d'équilibre et la congestion optimale, les modèles d'estimation de la demande récréative ont beaucoup progressé. Dans cette évolution, la question de la congestion a un peu disparu, surtout dans les modèles mettant en œuvre la méthode des coûts de transport. Si quelques articles l'ont introduite dans des modèles de choix discrets (Salanié, 2006 ; Yen et Adamowicz, 1994 ; Berman *et al.*, 1997 ; Hansen *et al.*, 1999 ; Schuhman et Schwabe, 2004 ; Lin *et al.*, 1996), ils la traitent comme une variable explicative comme les autres. La congestion est cependant susceptible d'introduire deux types d'endogénéité dans un modèle de choix :

- une endogénéité économique car la congestion est la résultante du processus de choix ;
- une endogénéité économétrique car la congestion est potentiellement corrélée au terme d'erreur du modèle.

Jusque 2007, peu d'alternatives au traitement exogène de la congestion étaient disponibles. Depuis les articles de Bayer et Timmins (2007), Timmins et Murdoch (2007) et O'Hara (2007a), différentes stratégies d'estimation des modèles avec congestion ont émergé, leur efficacité dépendant du cas d'étude. L'article de Bayer et Timmins (2007) analyse la question de l'endogénéité économique et économétrique de la congestion pour l'estimation des modèles de choix discrets et propose une approche générale, adaptée de celle de Berry *et al.* (1995), pour gérer la corrélation de la congestion avec les effets des attributs inobservés des alternatives de choix. Cette méthode a été appliquée pour la première fois par Timmins et Murdock (2007) dans le cadre de la modélisation des choix de sites récréatifs. Cette méthode d'estimation permet de traiter les deux types d'endogénéité dans les applications où les alternatives de choix sont nombreuses mais elle perd de sa pertinence quand leur nombre est faible. C'est dans ce cas que la procédure itérative d'O'Hara (2007a) est particulièrement intéressante. En présence d'un petit nombre d'alternatives, il est plus aisé de construire une typologie détaillée de ces

alternatives et donc de réduire l'endogénéité économétrique. Si ces alternatives sont homogènes, alors le soupçon de corrélation entre le terme d'erreur et la partie déterministe du modèle de choix disparaît. La procédure d'O'Hara (2007a) permet alors de se concentrer sur l'estimation du modèle à l'équilibre de congestion.

Dans la méthode BLP et celle d'O'Hara (2007a), la prise en compte de l'endogénéité économique de la congestion autorise les simulations de politiques publiques. Malheureusement, parfois, aucune des deux méthodes ne peut être mise en œuvre et seul une variable de congestion exogène est disponible. Il est alors encore possible de simuler des politiques publiques mais la fiabilité de la mesure de la congestion est alors décisive. En outre, la qualité des simulations repose sur la cohérence entre la congestion mesurée et les choix observés dans l'échantillon. Celui-ci doit donc être représentatif.

Les différences entre ces différents traitements de la congestion sont illustrées dans une application au choix de sites récréatifs naturels. Si les paramètres des caractéristiques individuels sont peu affectés par la manière d'introduire la congestion, ce n'est pas le cas pour les paramètres des attributs des sites. Cette absence de robustesse montre que, dans les applications avec peu de sites, une réflexion sur le traitement de la congestion en amont pour le recueil d'attributs de site précis et pertinents est cruciale pour bien expliquer le choix des sites. Dans les applications avec beaucoup d'alternatives (supérieures à 500 chez Timmins et Murdoch, 2007), la procédure de Bayer et Timmins (2007) permet d'estimer des paramètres d'attributs d'alternatives non biaisés.

Chez O'Hara (2007a, 2007b), comme chez Timmins et Murdock (2007) et Bayer et Timmins (2007), seule la congestion d'équilibre est calculée. Or, la congestion est une externalité : le comportement d'un individu est altéré par le comportement des autres individus via la congestion. Il existe donc une congestion d'équilibre mais également une congestion optimale, que le cadre de Fisher et Krutilla (1972) permettait de calculer. Ce cadre a bien été transposé de façon théorique dans un modèle de choix discrets (Leplat et Le Goffe, 2009), mais ce modèle est restrictif car limité à deux sites sans véritable application. Or, calculer l'optimum de fréquentation permet 1) de confirmer ou d'infirmer la nécessité de réguler la fréquentation du point de vue du bien-être collectif, 2) d'avoir un objectif de fréquentation qui ne soit pas simplement basé sur la capacité de charge du milieu et 3) de réfléchir aux politiques de régulation qui permettraient de se rapprocher de cet optimum. Cette voie devrait donc être creusée dans l'avenir.



## Références

- Allen, P.G., Stevens, T.H., Barret, S.A. (1981). The effects of variable omission in the travel cost technique. *Land Economics*, 57(2): 173–180.
- Anderson, F.J., Bonsor, N.C. (1974). Allocation, congestion et the valuation of recreational resources. *Land Economics*, 50(1): 51–57.
- Bayer, P., Timmins, C. (2005). On the equilibrium properties of locational sorting models. *Journal of Urban Economics*, 57(3): 462–477.
- Bayer, P., Timmins, C. (2007). Estimating equilibrium models of sorting across locations. *The Economic Journal*, 117(518): 353–374.
- Bell, F.W., Leeworthy, V.R. (1990). Recreational demand by tourists for saltwater beach days. *Journal of Environmental Economics et Management*, 18: 189–205.
- Berman, M., Haley, S., Kim, H. (1997). Estimating net benefits of reallocation: discrete choice models of sport et commercial fishing. *Marine Resource economics*, 12(4): 307–327.
- Berry, S. (1994). Estimating discrete choice models of product differentiation. *RAND Journal of Economics*, 25: 242–262.
- Berry, S., Levinsohn, J., Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63: 842–889.
- Bin, O., Landry, C.E., Ellis, C.L., Vogelsong, H. (2005). Some consumer surplus estimates for North Carolina beaches. *Marine Resource Economics*, 20: 145–161.
- Boxall, P., Rollins, K.S., Englin, J.E. (2003). Heterogeneous preferences for congestion during a wilderness experience. *Resource et Energy Economics*, 25(2): 177–195.
- Boxall, P.C., Adamowicz, W.L. (2000). Incorporating endogenous perceptions of environmental attributes in RUMs: The case of congestion. IIFET.
- Boxall, P.C., Hauer, G., Adamowicz, W.L. (2005). Modeling congestion as a form of interdependence in random utility models. Document de travail Université d’Alberta, Edmonton, Canada.
- Cesario, F.J. (1980). Congestion et the valuation of recreation benefits. *Land Economics*, 56(3): 329–338.

- Cicchetti, C.J., Smith, V.K. (1973). Congestion, quality deterioration, et optimal use: Wilderness recreation in the Spanish peaks primitive area. *Social Science Research*, 2(1): 15–30.
- Deyak, T.A., Smith, V.K. (1978). Congestion et participation in outdoor recreation: a household function production approach. *Journal of Environmental Economics et Management*, 5: 63–80.
- Englin, J., Shonkwiler, JS. (1995). Estimating social welfare using count data models: An application to long-run recreation demand under conditions of endogenous stratification et truncation. *The Review of Economics et Statistics*, 77(1): 104–12, 1995.
- Fisher, A.C., Krutilla, J.V.(1972). Determination of optimal capacity of resource-based recreation facilities. *Natural Resources Journal*, 12: 417–444.
- Hanley, N., Alvarez-Farizo, B., Shaw, D.W. (2002). Rationing an open-access resource: mountaineering in Scotland. *Land Use Policy*, 19:167-176.
- Hansen, L., Feather, P., Skank, D. (1999). Valuation of agriculture’s multi-site environmental impacts: an application to pheasant hunting. *Agricultural et Resource Economics Review*, 28(2): 199–207.
- Hindsley, P., Landry, C.E., Bin, O., Vogelsong, H. (2007). Site congestion in recreation choice models: A generated regressors approach to beach site selection. Working paper, East Carolina University, USA.
- Jakus, P., Shaw, W.D.(1997). Congestion at recreation areas: Empirical evidence on perceptions, mitigating behaviour et management preferences. *Journal of Environmental Management*, 50(4): 389–401.
- Kerkvliet, J., Nowell, C. (2000). Tools for recreation management in parks: the case of the greater Yellowstone’s blue ribbon fishery. *Ecological Economics*, 34: 89–100.
- Leplat, M., Le Goffe, P.(2009). Faut-il réguler l’encombrement des sites récréatifs: un modèle de choix discret avec participation. *Review of Agricultural et Environmental Studies*, 90(1): 51–77.
- Lin, P.C., Adams, R.M., Berrens, R.P. (1996). Welfare effects of fishery policies: Native american treaty rights et recreational salmon fishing. *Journal of Agricultural et Resource Economics*, 21(2): 263–276.

- McConnell, K.E. (1992). On-site time in the demand for recreation. *American Journal of Agricultural Economics*, 74: 919–925.
- McConnell, K.E. (1977). Congestion et willingness to pay: A study of beach use. *Land Economics*, 53(2): 185–195.
- Mullahy, J. (1986). Specification et testing of some modified count data models. *Journal of Econometrics*, 33: 341–365.
- O’Hara, M.E. (2007a). Congestion effects et mitigating behavior on recreation sites: An application to rock climbing access. Working paper, Colgate University, New-York, USA.
- O’Hara, M.E. (2007b). Congestion effects in a linked trip frequency et site choice model of recreation demand. Working paper, Colgate University, New-York, USA.
- Ozuna, T., Gomez, I.A. (1994). Estimating a system of recreation demand function using a seemingly unrelated Poisson regression approach. *Review of Economics et Statistics*, 2: 356–360.
- Parsons, G.R., Massey, D.M. (2003). A random utility model of beach recreation. In *The New Economics of Outdoor Recreation*, pages 241–267. Edward Elgar: Cheltenham, GB.
- Parsons, G.R. (2004). Travel cost models. In *A Primer on Nonmarket Valuation*. Kluwer Academic Publishers.
- Phaneuf, D.J., Carbone, J. C., Herriges, J.A. (2009). Non-price equilibria for non-marketed goods. *Journal of Environmental Economics et Management*, 57(1): 45–64.
- Salanié, J. (2006). *Analyse économique d’une activité récréative : la pêche au saumon en France*. PhD thesis, Ecole Nationale Supérieure Agronomique de Rennes.
- Schneider, I., Hammitt, W. (1995). Visitor response to outdoor recreation conflict: a conceptual approach. *Leisure Science*, 12: 223–234.
- Schuhman, P.W., Schwabe, K.A. (2004). An analysis of congestion measures et heterogeneous angler preferences in a random utility model of recreational fishing. *Environmental et Resource Economics*, 27: 487–499.
- Shelby, B., Vaske, J., Harris, R. (1988). User standards for ecological impacts at wilderness campsites. *Journal of Leisure Research*, 20: 245–256.
- Smith, V.K., Desvousges, W.H. (1985). The generalized travel cost model et water quality benefits: a reconsideration. *Southern Economic Journal*, 52: 371–382.

Timmins, C., Murdock, J. (2007). A revealed preference approach to the measurement of congestion in travel cost models. *Journal of Environmental Economics et Management*, 53: 230–249.

Yen, S.T., Adamowicz, W.L. (1994). Participation, trip frequency et site choice: a multinomial-Poisson hurdle model of recreation demand. *Canadian Journal of Agricultural Economics*, 42(1): 65–76.

**Les Working Papers SMART – LERECO sont produits par l'UMR SMART et l'UR LERECO**

- **UMR SMART**

L'Unité Mixte de Recherche (UMR 1302) *Structures et Marchés Agricoles, Ressources et Territoires* comprend l'unité de recherche d'Economie et Sociologie Rurales de l'INRA de Rennes et le département d'Economie Rurale et Gestion d'Agrocampus Ouest.

Adresse :

UMR SMART - INRA, 4 allée Bobierre, CS 61103, 35011 Rennes cedex  
UMR SMART - Agrocampus, 65 rue de Saint Briec, CS 84215, 35042 Rennes cedex

- **LERECO**

Unité de Recherche *Laboratoire d'Etudes et de Recherches en Economie*

Adresse :

LERECO, INRA, Rue de la Géraudière, BP 71627 44316 Nantes Cedex 03

Site internet commun : <http://www.rennes.inra.fr/smart/>

**Liste complète des Working Papers SMART – LERECO :**

<http://www.rennes.inra.fr/smart/Working-Papers-Smart-Lereco>

<http://econpapers.repec.org/paper/raewpaper/>

**The Working Papers SMART – LERECO are produced by UMR SMART and UR LERECO**

- **UMR SMART**

The « Mixed Unit of Research » (UMR1302) *Structures and Markets in Agriculture, Resources and Territories*, is composed of the research unit of Rural Economics and Sociology of INRA Rennes and of the Department of Rural Economics and Management of Agrocampus Ouest.

Address:

UMR SMART - INRA, 4 allée Bobierre, CS 61103, 35011 Rennes cedex, France  
UMR SMART - Agrocampus, 65 rue de Saint Briec, CS 84215, 35042 Rennes cedex, France

- **LERECO**

Research Unit *Economic Studies and Research Lab*

Address:

LERECO, INRA, Rue de la Géraudière, BP 71627 44316 Nantes Cedex 03, France

Common website: [http://www.rennes.inra.fr/smart\\_eng](http://www.rennes.inra.fr/smart_eng)

**Full list of the Working Papers SMART – LERECO:**

[http://www.rennes.inra.fr/smart\\_eng/Working-Papers-Smart-Lereco](http://www.rennes.inra.fr/smart_eng/Working-Papers-Smart-Lereco)

<http://econpapers.repec.org/paper/raewpaper/>

**Contact**

**Working Papers SMART – LERECO**

INRA, UMR SMART

4 allée Adolphe Bobierre, CS 61103

35011 Rennes cedex, France

**Email :** [smart\\_lereco\\_wp@rennes.inra.fr](mailto:smart_lereco_wp@rennes.inra.fr)

---

**2012**

**Working Papers SMART – LERECO**

UMR INRA-Agrocampus Ouest **SMART** (Structures et Marchés Agricoles, Ressources et Territoires)

UR INRA **LERECO** (Laboratoires d'Etudes et de Recherches en Economie)

Rennes, France

---