Techniques d'enquête, décembre 1990
Vol. 16, n° 2, pp. 321–347
Statistique Canada

Symétrie des flux, en tenant compte de la non-réponse, dans les catégories d'actes criminels déclarés par les victimes

ELIZABETH A. STASNY[1]

RÉSUMÉ

La United States National Crime Survey (enquête nationale sur la criminalité aux États-Unis) est une enquête-ménage de grande envergure ayant pour objet l'établissement d'estimations des pourcentages de personnes ou de ménages victimes d'actes criminels. L'enquête est une enquête par panel avec renouvellement et, selon le plan de sondage, les unités de logement sélectionnées font partie de l'échantillon pendant trois ans et demi, les occupants de ces unités de logements étant interviewés tous les six mois. Comme aussi peu que 25% des personnes interviewées dans le cadre de l'enquête participent à tous les cycles d'interview au cours de la période de trois ans et demi, la non-réponse constitue un des graves problèmes entachant les données longitudinales recueillies à l'aide de la National Crime Survey. De plus, l'occurrence de la non-réponse n'est pas aléatoire, c'est-à-dire qu'elle varie selon que le répondant a été ou non victime d'actes criminels. Le présent article porte sur des modèles d'estimation des flux bruts dans les catégories relatives à deux modes de classement des actes criminels déclarés: selon le nombre d'actes criminels et selon la gravité de l'acte criminel. Dans ces modèles, il y a des mécanismes de non-réponse aléatoire ou non aléatoire et les probabilités se rattachant aux flux bruts peuvent ne pas être assujetties à des contraintes ou être symétriques. Les modèles sont ajustés aux données recueillies dans le cadre de la National Crime Survey à l'aide d'estimateurs du maximum de vraisemblance.

MOTS CLÉS: Données qualitatives; non-réponse dont on n'a pas à tenir compte; enquête longitudinale; National Crime Survey; non-réponse dont il faut tenir compte.

1. INTRODUCTION

La NCS (United States National Crime Survey/enquête nationale sur la criminalité aux États-Unis) est une enquête-ménage de grande envergure réalisée par le U.S. Bureau of the Census pour le compte du Bureau of Justice Statistics. Les données de la NCS servent à établir des estimations trimestrielles des pourcentages de personnes et de ménages victimes d'actes criminels ainsi que des estimations annuelles des taux de criminalité. L'enquête utilise un panel d'unités de logement (UL) avec renouvellement et, selon le plan de sondage, les personnes occupant les UL sélectionnées sont interviewées jusqu'à un maximum de sept fois, à des intervalles de six mois.

Les personnes interviewées dans le cadre de la NCS sont interrogées au sujet des crimes contre la personne ou contre la propriété dont elles ont été victimes au cours des six mois antérieurs. Dans le présent article, nous commençons par étudier les actes criminels déclarés, d'une interview à l'autre, par les ménages (M) occupant les UL sélectionnées. Pour chaque ménage, ces déclarations sont étudiées sous deux aspects: selon le nombre d'actes criminels déclarés (zéro, un et deux ou plus) et selon le genre d'acte criminel déclaré (aucun crime, crime contre la propriété et crime contre la personne).

Comme on ne dispose pas de réponses pour tous les ménages d'un cycle d'interview à l'autre, il nous faut déterminer de quelle façon il convient de traiter les données manquantes. Les données longitudinales recueillies dans le cadre de la NCS soulèvent un grave problème de non-réponse. Ainsi, Fienberg (1980) a noté qu'on ne dispose d'enregistrements d'interview de la

[1] Elizabeth A. Stasny, Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, Ohio 43210, É.-U.

NCS portant sur la totalité de la période de trois ans et demi que pour aussi peu que 25% de l'ensemble des personnes interviewées. De plus, l'occurrence de la non-réponse n'est pas aléatoire, c'est-à-dire qu'elle varie selon que le répondant a été ou non victime d'actes criminels voir, par exemple, Saphire (1984).

Dans cet article, nous élargissons les modèles d'estimation des flux bruts tenant compte de la non-réponse non aléatoire élaborés par Stasny (1986). En particulier, les modèles que nous allons étudier prévoient des matrices de flux symétriques dans les catégories d'actes criminels ainsi qu'une non-réponse parfaitement aléatoire, une non-réponse non aléatoire dont on n'a pas à tenir compte, ou une non-réponse dont il faut tenir compte.

On trouve à la section 2 de l'article une brève description de la NCS et des données longitudinales recueillies dans le cadre de l'enquête. La section 3 expose la forme générale des modèles décrivant la symétrie des matrices de flux bruts avec données manquantes et présente les méthodes itératives permettant d'obtenir les estimateurs du maximum de vraisemblance (EMV) pour les paramètres des modèles. La section 4 porte sur l'ajustement des modèles aux données de la NCS. Enfin, on trouve à la section 5 un énoncé des conclusions qu'on peut tirer de la présente étude ainsi qu'un exposé de certaines pistes de recherche.

2. LA NATIONAL CRIME SURVEY ET SES DONNÉES

2.1 Plan de sondage

L'échantillon de la NCS est un échantillon d'UL à plusieurs degrés stratifié. Mise en oeuvre en juillet 1972 par la Law Enforcement Assistance Administration, l'enquête est réalisée par le Bureau of Justice Statistics depuis décembre 1979. La population cible de la NCS comprend tous les membres âgés de 12 ans ou plus de la population civile occupant des unités de logement, à l'exception des pensionnaires d'établissements institutionnels. L'enquête permet de recueillir des données sur les crimes contre la personne et contre les ménages dont sont victimes les occupants des UL sélectionnées. La NCS porte sur les infractions et les tentatives d'infraction suivantes: les voies de fait, le vol d'automobile ou de véhicule à moteur, le vol avec effraction, le vol simple, le viol et le vol qualifié. Figurent au nombre des actes criminels ne faisant pas partie du champ d'observation de l'enquête: l'enlèvement, le meurtre, le vol à l'étalage et les infractions perpétrées dans les établissements commerciaux.

La NCS est une enquête par panel avec renouvellement pour laquelle chaque UL sélectionnée fait partie de l'échantillon pendant trois ans et demi. Comme les interviews sont réalisées à des intervalles de six mois, chaque occupant des UL sélectionnées peut participer à un total de sept interviews. Toutefois, la première interview réalisée à chaque UL constitue une interview repère (n.d.t. Les données recueillies dans le cadre de cette interview permettent de déterminer si les données obtenues à l'occasion de l'interview suivante portent réellement sur la période de référence) et n'est pas utilisée aux fins de l'établissement d'estimations. Bien qu'un intervalle de six mois s'écoule entre les interviews réalisées à chaque UL, les interviews de la NCS s'échelonnent sur toute l'année. En effet, afin d'assurer une utilisation efficiente des intervieweurs, des interviews sont réalisées chaque mois auprès d'un sixième des UL formant l'échantillon. Comme c'est l'UL qui constitue l'unité d'échantillonnage de la NCS, rien n'est fait pour suivre les personnes qui quittent l'UL au cours de la période de trois ans et demi. On a plutôt choisi d'inclure dans l'univers de l'enquête les nouvelles personnes qui emménagent dans l'UL. Chaque groupe distinct de personnes occupant une UL pendant que cette dernière fait partie de l'échantillon de la NCS est considéré comme un ménage distinct.

Les interviews de la NCS portent sur toutes les personnes de 12 ans ou plus qui occupent l'UL sélectionnée au moment de l'interview. L'interview a pour objet d'interroger les répondants sur les crimes contre la personne ou contre le ménage dont ils ont été victimes au cours des six mois antérieurs. L'interview consiste à poser à chaque ménage répondant une série de six questions de sélection visant à recueillir des renseignements sur les crimes commis contre

le ménage (vol avec effraction, vol simple et vol de véhicule à moteur). Ces questions sont suivies de onze questions de filtrage visant à obtenir de chaque membre du ménage des renseignements sur les crimes contre la personne dont il a été victime (voies de fait, viol et vol qualifié). Enfin, un rapport d'acte criminel est établi pour chaque crime déclaré en réponse aux questions de filtrage.

Pour obtenir de plus amples renseignements sur le plan de sondage et l'historique de la NCS, on peut consulter entre autres les ouvrages du U.S. Deparment of Justice and Bureau of Justice Statistics (1981), Sapmre (1984), Dodge et Skogan (1987) et Montagliani (1987). La NCS utilise un nouveau plan de sondage depuis 1986. On trouvera une description du remaniement de la NCS ainsi que des travaux de recherche s'y rapportant dans Taylor (1987). Toutefois, les données utilisées dans cet article ont été recueillies suivant le plan de sondage intitial de la NCS.

2.2 Les données longitudinales

Les données utilisées dans cet article sont tirées d'un ensemble de données longitudinales de grande taille comprenant tous les renseignements recueillis par l'intermédiaire des interviews régulières de la NCS réalisées de janvier 1975 à juin 1979, à l'exception des interviews réalisées auprès des UL introduites dans l'échantillon par renouvellement en 1979. Afin que les données soient plus simples à traiter, les présents travaux portent uniquement sur un sous-ensemble de données. Ce sous-ensemble a été établi en prenant comme origine choisie au hasard l'enregistrement relatif à la huitième UL dans l'ensemble de données complet, puis en sélectionnant chaque quinzième enregistrement par la suite. L'ensemble de données résultant comporte les enregistrements de la NCS relatifs à 12,432 UL. Comme les UL figurant sur le ficher longitudinal initial sont ordonnées de telle façon que les unités appartenant à la même grappe sont regroupées, l'échantillon systématique de 1 sur 15 ne devrait pas comprendre plus d'une UL appartenant à la même grappe. En conséquence, le présent article ne traite pas du problème des corrélations entre les UL faisant partie d'une même grappe.

2.3 Flux dans les catégories d'actes criminels déclarés

Nous avons utilisé les données hiérarchiques longitudinales afin d'établir, pour les années 1975, 1976, 1977 et 1978, des matrices récapitulatives illustrant les flux au sein des catégories d'actes criminels déclarés entre les première et deuxième interviews réalisées auprès d'un ménage au cours d'une même année. Il convient de noter que, les interviews de la NCS s'échelonnant sur toute l'année, la première interview peut avoir lieu à n'importe quel moment entre les mois de janvier et juin, tandis que la seconde peut prendre place entre les mois de juillet et décembre. Les actes criminels déclarés au cours de la première interview se sont produits entre les mois de juillet et de mai précédents, tandis que les actes déclarés au cours de la seconde se sont produits entre les mois de janvier et novembre. Ainsi, la présente analyse porte uniquement sur les crimes déclarés d'une interview à l'autre. Elle ne saurait, par exemple, aborder les questions relatives aux variations dans les déclarations des répondants à diverses périodes de l'année, si ce n'est de façon très générale.

Il convient de noter qu'un échantillon de ménages de la NCS a participé, pendant la période de collecte des données, à une expérience portant sur la période de référence. Comme les occupants des UL ayant participé à l'expérience devaient déclarer les actes criminels dont ils avaient été victimes pour des périodes de référence autres que la période habituelle de six mois, ces UL n'ont pas été retenues aux fins de cette analyse.

En vue de la présente analyse, chaque ménage interviewé au moins une fois au cours d'une année donnée a été classé en fonction des codes de déclaration et de réponse qui lui ont été attribués au moment des deux interviews. Les actes criminels peuvent avoir été déclarés par n'importe quel membre du ménage et constituer des crimes contre la personne ou contre le ménage. Les analyses dont il est fait état à la section 4 portent sur deux jeux de matrices illustrant les catégories d'actes criminels déclarés. Ces matrices sont présentées à l'annexe I.

Le premier jeu illustre la répartition des ménages selon le nombre d'actes criminels déclarés au cours de chacune des deux interviews réalisées chaque année. Les catégories utilisées sont: aucun crime (aucun acte criminel déclaré), un seul crime (un acte criminel déclaré), crimes multiples (au moins deux actes criminels déclarés) et unités manquantes (le ménage n'a pas répondu ou a été supprimé de l'échantillon par renouvellement). Le deuxième jeu de matrices illustre la répartition des ménages selon le genre de crime déclaré. Les catégories utilisées sont: aucun crime, crime contre la propriété (vol avec effraction, vol simple et vol de véhicule à moteur), crime contre la personne (viol, voies de fait, vol qualifié, vol de sac à l'arraché et vol à la tire) et non-réponse. Ces catégories de genre de crime sont les mêmes que celles utilisées dans le cadre de la NCS. Lorsqu'un même ménage a déclaré de multiples crimes, on utilise le crime le plus grave déclaré aux fins du classement (les crimes contre la personne sont jugés plus graves que les crimes contre la propriété).

On notera le nombre élevé de cas de non-réponse dans les matrices illustrées à l'annexe I. Seulement 50% environ des ménages ayant participé à au moins une des deux interviews se sont prêtés aux deux cycles d'interview. Les modèles présentés dans la prochaine section nous permettront de tenir compte de cette non-réponse tout en étudiant la structure de la matrice sous-jacente des probabilités de flux dans les catégories d'actes criminels déclarés.

3. LES MODÈLES

Nous examinerons dans la présente section la forme générale des modèles qui seront utilisés pour étudier les flux bruts dans les catégories d'actes criminels déclarés dans le cadre de la NCS. Les modèles épousent la forme proposée par Chen et Fienberg (1974) pour les tableaux de contingence comportant des données à classification complète ou partielle.Les modèles pour la non-réponse et le modèle pour la non-réponse aléatoire ont été élaborés par Stasny (1986). Toutefois, le modèle pour la symétrie des flux ne figure pas dans les ouvrages antérieurs de l'auteure. Nous avons choisi de présenter les modèles sous forme générale parce qu'ils peuvent être appliqués à des problèmes autres que l'estimation des flux bruts au sein des catégories d'actes criminels déclarés dans le cadre de la NCS.

3.1 Modèle pour les observations

Considérons des unités d'observation se prêtant à au moins un des deux cycles d'interview d'une enquête. Supposons que, lorsqu'une unité participe à l'enquête, elle est classée dans une de K catégories. Lorsqu'une unité ne participe pas à l'enquête, elle est classée comme manquante. Alors on peut représenter le flux des observations d'une interview à l'autre comme dans le tableau 1.

Tableau 1
Sommaire des observations

		Temps 2				
		1	2	...	K	Manquante
Temps 1	1	x_{11}	x_{12}	...	x_{1K}	x_{1M}
	2	x_{21}	x_{22}	...	x_{2K}	x_{2M}
	⋮	⋮	⋮	⋮	⋮	⋮
	K	x_{K1}	x_{K2}	...	x_{KK}	x_{KM}
	Manquante	x_{M1}	x_{M2}	...	x_{MK}	,

où x_{ij} = nombre d'unités passant de l'état i (code de réponse ou de non-réponse) au temps 1 à l'état j au temps 2.

Nous supposons que chaque unité se trouverait dans une des cellules de la matrice $K \times K$ des catégories de réponses si elle était observée au cours des deux cycles d'interview. Définissons p_{ij} comme la probabilité qu'une unité passe de l'état i au temps 1 à l'état j au temps 2, où i et j prennent les valeurs $1, 2, \ldots, K$. Chaque unité de la cellule (i,j) de la matrice de catégories de réponses a une chance d'être manquante à l'occasion d'un des deux cycles d'interview. Définissons λ_{tij} comme la probabilité qu'une des unités de la cellule (i,j) ne réponde pas au temps t et soit donc classée comme manquante. Alors, les probabilités se rattachant aux observations sont montrées au tableau 2.

Tableau 2
Probabilités se rattachant aux observations

		Temps 2				
		1	2	...	K	Manquante
Temps 1	1	$\{(1 - \lambda_{1ij} - \lambda_{2ij})p_{ij}\}$				$\left\{\sum_{j=1}^{K} p_{ij}\lambda_{2ij}\right\}$
	2					
	⋮					
	K					
Manquante		$\left\{\sum_{i=1}^{K} p_{ij}\lambda_{1ij}\right\}$				

Si on suppose que les probabilités p_{ij} obéissent à une loi multinomiale, la fonction de vraisemblance pour les observations est proportionnelle à

$$\left\{\prod_{i=1}^{K}\prod_{j=1}^{K}\left[p_{ij}(1 - \lambda_{1ij} - \lambda_{2ij})\right]^{x_{ij}}\right\}$$

$$\times \left\{\prod_{i=1}^{K}\left[\sum_{j=1}^{K} p_{ij}\lambda_{2ij}\right]^{x_{iM}}\right\}$$

$$\times \left\{\prod_{j=1}^{K}\left[\sum_{i=1}^{K} p_{ij}\lambda_{1ij}\right]^{x_{Mj}}\right\}.$$

Le nombre des paramètres libres définis ci-dessus est égal à $3K^2 + 2K - 1$, tandis que le nombre de cellules d'observations avec une seule contrainte s'appliquant à la taille de l'échantillon global est seulement de $K^2 + 2K$. Le nombre de paramètres à estimer à l'aide des observations est donc trop élevé et il nous faut réduire le nombre des paramètres compris dans le modèle. Dans la suite de l'article, nous procédons à une telle réduction en considérant deux modèles pour les paramètres p_{ij} et six modèles pour les paramètres λ_{tij}.

3.2 Modèles pour les probabilités p et λ

Nous considérons deux modèles pour estimer les probabilités (p_{ij}) de flux au sein des catégories de réponses: le modèle non assujetti à des contraintes et le modèle des flux symétriques. Suivant le modèle des probabilités de flux non assujetties à des contraintes, il existe une probabilité différente, p_{ij}, pour chaque cellule (i,j) de la matrice de flux. Suivant le modèle

des flux symétriques, $p_{ij} = p_{ji}$ pour $i \neq j$ de telle sorte que la probabilité qu'une unité passe de l'état i au temps 1 à l'état j au temps 2 est la même que la probabilité qu'une unité passe de l'état j au temps 1 à l'état i au temps 2. Notons que la symétrie des probabilités de cellules de la matrice de flux implique une égalité entre les totaux marginaux de ligne et de colonne. Aussi le modèle décrivant la symétrie des probabilités de flux implique-t-il une certaine stabilité de la population puisque le nombre prévu d'unités classées dans une catégorie de réponses donnée est le même au temps 1 et au temps 2.

Selon la définition ci-devant, λ_{tij}, la probabilité que les unités classées dans les catégories de réponses i au temps 1 et j au temps 2 soient manquantes au temps t est fonction du moment auquel la non-réponse se produit et des catégories de réponses dans lesquelles l'unité est classée au temps 1 et au temps 2. Nous considérons, pour estimer ces probabilités, six modèles plus simples que nous exposons ci-après avec les degrés de liberté correspondants suivant les deux modèles pour les p_{ij}:

	d.l. p_{ij} non assujetties à des contraintes	d.l. p_{ij} symétriques
Modèle R: $\lambda_{tij} = \lambda$,	$2K - 1$	$(K^2 + 3K - 2)/2$
Modèle A: $\lambda_{1ij} = \lambda_{1j}$, $\lambda_{2ij} = \lambda_{2i}$,	0	$(K^2 - K)/2$
Modèle B: $\lambda_{tij} = \lambda_t$,	$2K - 2$	$(K^2 + 3K - 4)/2$
Modèle C: $\lambda_{1ij} = \lambda_j$, $\lambda_{2ij} = \lambda_i$,	K	$(K^2 + K)/2$
Modèle D: $\lambda_{1ij} = \lambda_{1i}$, $\lambda_{2ij} = \lambda_{2j}$,	0	$(K^2 - K)/2$
Modèle E: $\lambda_{1ij} = \lambda_i$, $\lambda_{2ij} = \lambda_j$,	K	$(K^2 + K)/2$

Le modèle R est le modèle pour la non-réponse aléatoire. Suivant ce modèle, il n'y a qu'une seule probabilité de non-réponse pour toutes les unités aux deux temps considérés, quelle que soit la catégorie de réponse. Suivant le modèle A, la probabilité qu'une unité soit manquante au temps t est fonction de la valeur de t et de la catégorie de réponses dans laquelle l'unité est classée au moment où elle répond. Notons que si on utilise le modèle A pour les paramètres λ et le modèle non assujetti à des contraintes pour les p_{ij}, alors le modèle est un modèle saturé qui s'ajuste parfaitement aux données. Suivant le modèle B, la probabilité qu'une unité soit manquante au temps t est fonction à uniquement de la valeur de t. Suivant le modèle C, la probabilité qu'une unité soit manquante au temps t est fonction uniquement de la catégorie de réponses dans laquelle l'unité est classée au moment où elle répond. Suivant le modèle D, la probabilité qu'une unité soit manquante au temps t est fonction à la fois de la valeur de t et de la catégorie de réponses dans laquelle l'unité est classée au moment où elle est manquante. Si on utilise le modèle D pour les paramètres et le modèle non assujetti à des contraintes pour les p_{ij}, alors le modèle est un modèle saturé qui s'ajuste parfaitement aux données. Suivant le modèle E, la probabilité qu'une unité soit manquante au temps t est fonction uniquement de la catégorie de réponses dans laquelle l'unité est classée au moment où elle est manquante.

Suivant le modèle R, la non-réponse est dite parfaitement aléatoire. Suivant les modèles A, B et C, on n'a pas à tenir compte de la non-réponse puisque le mécanisme de non-réponse est uniquement fonction des observations. Suivant les modèles D et E, il faut tenir compte de la non-réponse puisque le mécanisme de non-réponse est fonction des données manquantes. Pour plus de renseignements sur les types de non-réponse, voir Little et Rubin (1987).

Nous décrivons dans les deux prochaines sous-sections les méthodes d'ajustement des modèles présentés ci-devant. On peut évaluer l'ajustement des modèles à l'aide de la statistique X^2 de Pearson ou du rapport des vraisemblances G^2. Les deux statistiques suivent des

distributions asymptotiques du χ^2 avec le nombre de degrés de liberté indiqué ci-devant, pour autant que le modèle soit correct. Nous utiliserons ci-après la notation "modèle R-U" pour désigner la combinaison du modèle R pour les paramètres λ et du modèle non assujetti aux contraintes pour les p_{ij}, ainsi que la notation "modèle R-S" pour désigner la combinaison du modèle R pour les paramètres λ et du modèle symétrique pour les p_{ij}. Nous ferons appel à une notation similaire pour désigner la combinaison des modèles A, B, C, D ou E pour les paramètres λ avec un des deux modèles pour les p_{ij}.

3.3 Estimation des paramètres p et λ suivant les modèles R, A, B et C

Les fonctions de vraisemblance pour les huit modèles créés en combinant un des deux modèles pour les p_{ij} au modèle R, A, B ou C pour les λ_{tij} se décomposent en deux facteurs: le premier est une fonction des paramètres p uniquement et le second, une fonction des paramètres λ uniquement. On peut donc obtenir les EMV séparément pour les deux ensembles de paramètres. En outre, les estimations des paramètres p ne varient pas en fonction du modèle utilisé pour les paramètres λ et les estimations des paramètres λ sont indépendantes du modèle utilisé pour les paramètres p.

Chen et Fienberg (1974) exposent une méthode itérative permettant d'obtenir les EMV pour les paramètres p suivant le modèle non assujetti à des contraintes combiné au modèle R, A, B ou C pour les paramètres λ. On trouve les équations relatives à cette méthode à l'annexe II.

Suivant le modèle symétrique pour les paramètres p combiné au modèle R, A, B ou C pour les paramètres λ, le facteur de la fonction de vraisemblance portant uniquement sur les p_{ij} se lit comme suit:

$$\left\{\prod_{i=1}^{k} p_{ii}^{x_{ii}}\right\} \times \left\{\prod_{i=1}^{k}\prod_{j=i+1}^{k} p_{ij}^{x_{ij}}\right\} \times \left\{\prod_{i=2}^{k}\prod_{j=1}^{i-1} p_{ji}^{x_{ij}}\right\}$$

$$\times \left\{\prod_{i=1}^{k} p_{i\cdot}^{x_{iM}}\right\} \times \left\{\prod_{j=1}^{k} p_{j\cdot}^{x_{Mj}}\right\}, \qquad (1)$$

où le point substitué à l'indice inférieur indique que la sommation s'étend à toutes les valeurs de cet indice. On maximise l'équation (1) en tenant compte de la condition que la somme des p_{ij} est égale à un. En général, il faut utiliser une méthode itérative pour obtenir les EMV. Soit $x.. = \sum_{i=1}^{K}\sum_{j=1}^{K} x_{ij}$ le nombre total d'unités observées dans les deux cycles d'interview et $n = x.. + x._M + x_{M\cdot}$, le nombre total d'unités observées dans au moins un des deux cycles d'interview. Alors la méthode itérative utilisée aux fins de l'analyse des données dont il est fait état dans la section 4 est la suivante.

Méthode itérative d'estimation des p_{ij} symétriques suivant les modèles R, A, B et C

1. $p_{ii}^{(0)} = x_{ii}/x..$

   $p_{ij}^{(0)} = (x_{ij} + x_{ji})/2x..$ pour $i \neq j$.

2. $p_{ii}^{(v+1)} = \left[x_{ii} + (x_{iM} + x_{Mi})\, p_{ii}^{(v)}/p_{i\cdot}^{(v)}\right]/n$

   $p_{ij}^{(v+1)} = \left[(x_{ij} + x_{ji}) + (x_{iM} + x_{Mi})\, p_{ij}^{(v)}/p_{i\cdot}^{(v)} + (x_{jM} + x_{Mj})\, p_{ij}^{(v)}/p_{j\cdot}^{(v)}\right]/2n$ pour $i \neq j$.

On répète l'étape 2 pour $\nu = 0, 1, 2, \ldots$ jusqu'à ce que les estimations des paramètres convergent vers le degré d'exactitude prescrit. Les estimations initiales données à l'étape 1 le sont uniquement à titre indicatif. Il est possible d'utiliser d'autres valeurs positives satisfaisant à la contrainte selon laquelle la somme des p_{ij} doit être égale à un.

Chen et Fienberg (1974) exposent également une méthode itérative permettant d'obtenir les EMV pour les paramètres λ suivant le modèle A ainsi que l'estimateur en forme analytique pour les paramètres λ suivant le modèle B. Par ailleurs Stasny (1986) expose une méthode itérative permettant d'obtenir les EMV pour les paramètres λ suivant le modèle C. On trouve les équations relatives à ces méthodes à l'annexe II.

Suivant le modèle R pour les paramètres λ, le facteur de l'équation de vraisemblance portant uniquement sur les paramètres λ est le suivant:

$$\left\{\prod_{i=1}^{K}\prod_{j=1}^{K}(1-2\lambda)^{x_{ij}}\right\} \times \left\{\prod_{i=1}^{K}\lambda^{x_{iM}}\right\} \times \left\{\prod_{j=1}^{K}\lambda^{x_{Mj}}\right\}.$$

L'EMV en forme analytique pour λ est

$$\hat{\lambda} = (x_{\cdot M} + x_{M\cdot})/2n.$$

3.4 Estimation des paramètres p et λ suivant le modèle D

Selon le modèle D-U ou le modèle D-S, les fonctions de vraisemblance pour les observations ne peuvent se décomposer en facteurs et les estimations de tous les paramètres doivent être obtenues simultanément. On trouve dans Stasny (1988) une méthode itérative permettant d'obtenir les EMV suivant le modèle D-U. Les équations relatives à cette méthode sont présentées à l'annexe II. Suivant le modèle D-S, la fonction de vraisemblance pour les observations est la suivante:

$$\left\{\prod_{i=1}^{K} p_{ii}^{x_{ii}}\right\} \times \left\{\prod_{i=1}^{K}\prod_{j=i+1}^{K} p_{ij}^{x_{ij}}\right\} \times \left\{\prod_{i=2}^{K}\prod_{j=i}^{i-1} p_{ji}^{x_{ij}}\right\} \times \left\{\prod_{i=1}^{K}\prod_{j=1}^{K}\left[(1-\lambda_{1i}-\lambda_{2j})\right]^{x_{ij}}\right\}$$

$$\times \left\{\prod_{i=1}^{K}\left[\sum_{j=1}^{K} p_{ij}\lambda_{2j}\right]^{x_{iM}}\right\} \times \left\{\prod_{j=1}^{K}\left[\sum_{i=1}^{K} p_{ji}\lambda_{1i}\right]^{x_{Mj}}\right\}. \qquad (2)$$

On maximise l'équation (2) en tenant compte de la condition que la somme des p_{ij} est égale à un. En général, il faut utiliser une méthode itérative pour obtenir les EMV. La méthode itérative utilisée aux fins de l'analyse des données dont il est fait état dans la section 4 est la suivante.

Méthode itérative d'estimation des paramètres suivant le modèle D-S

1. $p_{ii}^{(0)} = x_{ii}/x..$

   $p_{ij}^{(0)} = (x_{ij} + x_{ji})/2x..$ pour $i \neq j$

   $\lambda_{1i}^{(0)} = x_{M\cdot}/n$

   $\lambda_{2j}^{(0)} = x_{\cdot M}/n.$

$$2.\ p_{ii}^{(\nu+1)} = n^{-1}\left\{x_{ii} + x_{iM}\left[p_{ii}^{(\nu)}\lambda_{2i}^{(\nu)} \Big/ \sum_{h=1}^{K} p_{ih}^{(\nu)}\lambda_{2h}^{(\nu)}\right] + x_{Mi}\left[p_{ii}^{(\nu)}\lambda_{1i}^{(\nu)} \Big/ \sum_{h=1}^{K} p_{ih}^{(\nu)}\lambda_{1h}^{(\nu)}\right]\right\}$$

$$p_{ij}^{(\nu+1)} = (2n)^{-1}\left\{x_{ij} + x_{ji} + x_{iM}\left[p_{ij}^{(\nu)}\lambda_{2j}^{(\nu)} \Big/ \sum_{h=1}^{K} p_{ih}^{(\nu)}\lambda_{2h}^{(\nu)}\right]\right.$$

$$+ x_{jM}\left[p_{ij}^{(\nu)}\lambda_{2i}^{(\nu)} \Big/ \sum_{h=1}^{K} p_{jh}^{(\nu)}\lambda_{2h}^{(\nu)}\right] + x_{Mi}\left[p_{ij}^{(\nu)}\lambda_{1j}^{(\nu)} \Big/ \sum_{h=1}^{K} p_{ih}^{(\nu)}\lambda_{1h}^{(\nu)}\right]$$

$$\left. + x_{Mj}\left[p_{ij}^{(\nu)}\lambda_{1i}^{(\nu)} \Big/ \sum_{h=1}^{K} p_{jh}^{(\nu)}\lambda_{1h}^{(\nu)}\right]\right\} \quad \text{pour} \quad i \neq j$$

$$\lambda_{1i}^{(\nu+1)} = \sum_{j=1}^{K}\left[x_{Mj}p_{ij}^{(\nu)}\lambda_{1i}^{(\nu)} \Big/ \sum_{h=1}^{K} p_{jh}^{(\nu)}\lambda_{1h}^{(\nu)}\right] \Big/ \sum_{j=1}^{K}\left[x_{ij}/\left(1 - \lambda_{1i}^{(\nu)} - \lambda_{2j}^{(\nu)}\right)\right]$$

$$\lambda_{2j}^{(\nu+1)} = \sum_{i=1}^{K}\left[x_{iM}p_{ij}^{(\nu)}\lambda_{2j}^{(\nu)} \Big/ \sum_{h=1}^{K} p_{ih}^{(\nu)}\lambda_{2h}^{(\nu)}\right] \Big/ \sum_{i=1}^{K}\left[x_{ij}/\left(1 - \lambda_{1i}^{(\nu)} - \lambda_{2j}^{(\nu)}\right)\right].$$

On répète l'étape 2 pour $\nu = 0, 1, 2, \ldots$ jusqu'à ce que les estimations des paramètres convergent vers le degré d'exactitude prescrit. Les estimations initiales données à l'étape 1 le sont uniquement à titre indicatif. Il est possible d'utiliser toute autre valeur située entre zéro et un et satisfaisant à la contrainte selon laquelle la somme des p_{ij} doit être égale à un.

3.5 Estimation des paramètres p et λ suivant le modèle E

Selon le modèle E-U ou le modèle E-S, les fonctions de vraisemblance pour les observations ne peuvent se décomposer en facteurs et les estimations de tous les paramètres doivent être obtenues simultanément. Stasny (1988) expose une méthode itérative permettant d'obtenir les EMV suivant le modèle E-U. On trouve les équations relatives à cette méthode à l'annexe II. Suivant le modèle E-S, la fonction de vraisemblance pour les observations est la suivante:

$$\left\{\prod_{i=1}^{K} p_{ii}^{x_{ii}}\right\} \times \left\{\prod_{i=1}^{K}\prod_{j=i+1}^{K} p_{ij}^{x_{ij}}\right\} \times \left\{\prod_{i=2}^{K}\prod_{j=1}^{i-1} p_{ji}^{x_{ij}}\right\} \times \left\{\prod_{i=1}^{K}\prod_{j=1}^{K}\left[(1 - \lambda_i - \lambda_j)\right]^{x_{ij}}\right\}$$

$$\times \left\{\prod_{i=1}^{K}\left[\sum_{j=1}^{K} p_{ij}\lambda_j\right]^{x_{iM}}\right\} \times \left\{\prod_{j=1}^{K}\left[\sum_{i=1}^{K} p_{ji}\lambda_i\right]^{x_{Mj}}\right\}. \qquad (3)$$

On maximise l'équation (3) en tenant compte de la condition que la somme des p_{ij} est égale à un. En général, il faut avoir recours à une méthode itérative pour obtenir les EMV. La méthode itérative utilisée aux fins de l'analyse des données dont il est fait état dans la section 4 est la suivante.

Méthode itérative d'estimation des paramètres suivant le modèle E-S

1. $p_{ii}^{(0)} = x_{ii}/x..$

$$p_{ij}^{(0)} = (x_{ij} + x_{ji})/2x.. \quad \text{pour} \quad i \neq j$$

$$\lambda_i^{(0)} = (x_{M\cdot} + x_{\cdot M})/2n.$$

2. $$p_{ii}^{(\nu+1)} = n^{-1}\left\{x_{ii} + (x_{iM} + x_{Mi})\left[p_{ii}^{(\nu)}\lambda_i^{(\nu)} \Big/ \sum_{h=1}^{K} p_{ih}^{(\nu)}\lambda_h^{(\nu)}\right]\right\}$$

$$p_{ij}^{(\nu+1)} = (2n)^{-1}\left\{x_{ij} + x_{ji} + (x_{iM} + x_{Mi})\left[p_{ij}^{(\nu)}\lambda_j^{(\nu)} \Big/ \sum_{h=1}^{K} p_{ih}^{(\nu)}\lambda_h^{(\nu)}\right]\right.$$

$$\left. + (x_{jM} + x_{Mj})\left[p_{ij}^{(\nu)}\lambda_i^{(\nu)} \Big/ \sum_{h=1}^{K} p_{jh}^{(\nu)}\lambda_h^{(\nu)}\right]\right\} \quad \text{pour} \quad i \neq j$$

$$\lambda_i^{(\nu+1)} = \sum_{j=1}^{K}\left[(x_{jM} + x_{Mj})p_{ji}^{(\nu)}\lambda_i^{(\nu)} \Big/ \sum_{h=1}^{K} p_{jh}^{(\nu)}\lambda_h^{(\nu)}\right]$$

$$\Big/ \sum_{j=1}^{K}\left[(x_{ij} + x_{ji})/(1 - \lambda_i^{(\nu)} - \lambda_j^{(\nu)})\right].$$

On répète l'étape 2 pour $\nu = 0, 1, 2, \ldots$ jusqu'à ce que les estimations des paramètres convergent vers le degré d'exactitude prescrit. Les estimations initiales données à l'étape 1 le sont uniquement à titre indicatif. Il est possible d'utiliser toute autre valeur située entre zéro et un et satisfaisant à la contrainte selon laquelle la somme des p_{ij} doit être égale à un.

4. AJUSTEMENT DES MODÈLES AUX DONNÉES DE LA NCS

Les modèles décrits à la section 3 ont été ajustés aux données de la NCS dont il est question à la section 2. Le lecteur se souviendra que les données de la NCS recueillies pour chaque année de 1975 à 1978 sont classés selon le nombre d'actes criminels déclarés lors de chacune des deux interviews réalisées au cours de l'année et selon le genre d'acte criminel. Comme nous avons utilisé trois catégories de réponses, nous avons $K = 3$. Les erreurs types des estimations des paramètres ont été calculées à partir de la matrice des observations.

Tableau 3a

Estimations des p_{ij} pour les flux dans les catégories de nombre d'actes criminels suivant les modèles R, A, B, et C

		Modèle non assujetti à des contraintes			Modèle symétrique		
		Deuxième interview					
		Aucun Crime	Un seul crime	Crimes multiples	Aucun crime	Un seul crime	Crimes multiples
1975							
Première	Aucun crime	.666 (.0075)	.098 (.0050)	.029 (.0031)	.666 (.0075)	.102 (.0035)	.032 (.0022)
interview	Un seul crime	.106 (.0051)	.029 (.0031)	.014 (.0023)	.102 (.0035)	.029 (.0031)	.012 (.0015)
	Crimes multiples	.036 (.0032)	.011 (.0021)	.012 (.0021)	.032 (.0022)	.012 (.0015)	.012 (.0021)
1976							
Première	Aucun crime	.669 (.0076)	.101 (.0052)	.029 (.0033)	.669 (.0076)	.099 (.0036)	.030 (.0022)
interview	Un seul crime	.098 (.0051)	.034 (.0034)	.014 (.0025)	.099 (.0036)	.034 (.0034)	.014 (.0017)
	Crimes multiples	.031 (.0030)	.014 (.0023)	.011 (.0022)	.030 (.0022)	.014 (.0017)	.010 (.0022)
1977							
Première	Aucun crime	.670 (.0079)	.115 (.0058)	.032 (.0034)	.671 (.0079)	.103 (.0037)	.030 (.0023)
interview	Un seul crime	.092 (.0051)	.026 (.0032)	.016 (.0026)	.103 (.0037)	.026 (.0032)	.016 (.0018)
	Crimes multiples	.028 (.0030)	.016 (.0026)	.006 (.0017)	.030 (.0023)	.016 (.0018)	.006 (.0017)
1978							
Première	Aucun crime	.671 (.0087)	.097 (.0062)	.027 (.0035)	.671 (.0087)	.105 (.0043)	.027 (.0025)
interview	Un seul crime	.111 (.0061)	.032 (.0040)	.009 (.0022)	.105 (.0043)	.032 (.0040)	.010 (.0017)
	Crimes multiples	.027 (.0034)	.013 (.0027)	.013 (.0026)	.027 (.0025)	.010 (.0017)	.013 (.0026)

Nota: Les erreurs types estimées sont indiquées entre parenthèses.

4.1 Estimations des paramètres p selon les modèles R, A, B et C

Selon les modèles R, A, B et C, les estimations des paramètres p comme on l'a déjà dit, ne dépendent pas du mécanisme de non-réponse. Suivant la méthode itérative servant à estimer le paramètre p_{ij} selon le modèle non assujetti à des contraintes et selon le modèle symétrique, le critère utilisé pour mettre fin à l'itération était que les chiffres prévus dans la cellule (i,j) de la matrice de flux, $n\hat{p}_{ij}$, varient d'au plus 0.5 d'une étape à l'autre du processus d'itération.

Dans tous les cas, il y a rapidement eu convergence, après au plus six étapes. Les estimations des paramètres p_{ij} lorsque les ménages sont classés selon le nombre d'actes criminels déclarés sont présentées au tableau 3a, tant pour le modèle non assujetti à des contraintes que pour le modèle symétrique. On trouve au tableau 4a les estimations correspondantes lorsque les ménages sont classés selon le genre d'acte criminel déclaré.

Tableau 3b

Estimations des p_{ij} pour les flux dans les catégories de nombre d'actes criminels suivant le modèle D-S

			Modèle symétrique		
			Deuxième interview		
			Aucun crime	Un seul crime	Crimes multiples
1975					
	Première	Aucun crime	.638 (.0104)	.106 (.0047)	.035 (.0029)
	interview	Un seul crime	.106 (.0047)	.033 (.0039)	.015 (.0019)
		Crimes multiples	.035 (.0029)	.015 (.0019)	.016 (.0027)
1976					
	Première	Aucun crime	.645 (.0100)	.100 (.0045)	.034 (.0029)
	interview	Un seul crime	.100 (.0045)	.037 (.0041)	.017 (.0021)
		Crimes multiples	.034 (.0029)	.017 (.0021)	.015 (.0029)
1977					
	Première	Aucun crime	.642 (.0109)	.106 (.0054)	.033 (.0032)
	interview	Un seul crime	.106 (.0054)	.031 (.0043)	.021 (.0023)
		Crimes multiples	.033 (.0032)	.021 (.0023)	.009 (.0025)
1978					
	Première	Aucun crime	.636 (.0118)	.114 (.0056)	.028 (.0029)
	interview	Un seul crime	.114 (.0056)	.040 (.0051)	.013 (.0021)
		Crimes multiples	.028 (.0029)	.013 (.0021)	.015 (.0030)

Nota: Les erreurs types estimées sont indiquées entre prarenthèses.

Il convient de remarquer, dans les tableaux 3a et 4a, que les matrices de flux des probabilités estimées selon le modèle non assujetti à des contraintes pour les paramètres p_{ij} semblent passablement symétriques. Le modèle pour la symétrie des flux se révèle donc un modèle dont on peut raisonnablement envisager l'utilisation. Notons également que les estimations des paramètres p_{ij} ne semblent pas varier beaucoup au cours de la période de quatre ans. Les ajustements de ces deux modèles pour les paramètres p_{ij} seront étudiés pour chacun des quatre modèles de non-réponse au paragraphe 4.4 ci-après.

Tableau 3c

Estimations des p_{ij} pour les flux dans les catégories de nombre d'actes criminels suivant les modèles E-U et E-S

		Modèle non assujetti à des contraintes			Modèle symétrique		
		Deuxième interview					
		Aucun crime	Un seul crime	Crimes multiples	Aucun crime	Un seul crime	Crimes multiples
1975							
Première	Aucun crime	.639 (.0104)	.102 (.0061)	.031 (.0037)	.639 (.0104)	.106 (.0047)	.035 (.0028)
interview	Un seul crime	.110 (.0061)	.033 (.0039)	.016 (.0026)	.106 (.0047)	.033 (.0039)	.015 (.0019)
	Crimes multiples	.039 (.0039)	.014 (.0025)	.016 (.0027)	.035 (.0028)	.015 (.0019)	.016 (.0027)
1976							
Première	Aucun crime	.645 (.0100)	.103 (.0063)	.032 (.0041)	.645 (.0100)	.101 (.0045)	.033 (.0029)
interview	Un seul crime	.098 (.0057)	.037 (.0041)	.017 (.0030)	.101 (.0045)	.037 (.0041)	.017 (.0021)
	Crimes multiples	.035 (.0037)	.017 (.0027)	.016 (.0029)	.033 (.0029)	.017 (.0021)	.016 (.0029)
1977							
Première	Aucun crime	.636 (.0112)	.124 (.0083)	.037 (.0050)	.642 (.0110)	.106 (.0055)	.033 (.0033)
interview	Un seul crime	.094 (.0060)	.031 (.0043)	.021 (.0031)	.106 (.0055)	.030 (.0043)	.020 (.0023)
	Crimes multiples	.029 (.0036)	.020 (.0031)	.008 (.0024)	.033 (.0033)	.020 (.0023)	.008 (.0025)
1978							
Première	Aucun crime	.639 (.0118)	.106 (.0078)	.029 (.0042)	.637 (.0118)	.112 (.0055)	.028 (.0029)
interview	Un seul crime	.117 (.0070)	.041 (.0051)	.011 (.0026)	.112 (.0055)	.041 (.0051)	.013 (.0021)
	Crimes multiples	.027 (.0037)	.016 (.0032)	.015 (.0030)	.028 (.0029)	.013 (.0021)	.015 (.0030)

Nota: Les erreurs types estimées son indiquées entre parenthèses.

Tableau 4a

Estimations des p_{ij} pour les flux dans les catégories de genre d'actes criminels suivant les modèles, R, A, B, et C

		Modèle non assujetti à des contraintes			Modèle symétrique		
		Deuxième interview					
		Aucun crime	Crime contre la propriété	Crime contre la personne	Aucun crime	Crime contre la propriété	Crime contre la personne
1975							
Première	Aucun crime	.666 (.0075)	.105 (.0053)	.022 (.0026)	.666 (.0075)	.111 (.0037)	.024 (.0018)
interview	Crime contre la propriété	.118 (.0054)	.044 (.0038)	.010 (.0019)	.111 (.0037)	.044 (.0038)	.008 (.0013)
	Crime contre la personne	.025 (.0026)	.007 (.0016)	.004 (.0012)	.024 (.0018)	.008 (.0013)	.004 (.0012)
1976							
Première	Aucun crime	.669 (.0076)	.108 (.0055)	.023 (.0028)	.669 (.0021)	.108 (.0011)	.022 (.0010)
interview	Crime contre la propriété	.108 (.0053)	.047 (.0040)	.010 (.0021)	.108 (.0011)	.047 (.0019)	.011 (.0009)
	Crime contre la personne	.021 (.0025)	.012 (.0021)	.002 (.0011)	.022 (.0010)	.011 (.0009)	.002 (.0012)
1977							
Première	Aucun crime	.670 (.0079)	.128 (.0061)	.019 (.0026)	.671 (.0078)	.115 (.0039)	.018 (.0018)
interview	Crime contre la propriété	.103 (.0053)	.041 (.0039)	.008 (.0018)	.115 (.0039)	.041 (.0040)	.008 (.0014)
	Crime contre la personne	.016 (.0025)	.008 (.0021)	.006 (.0018)	.018 (.0018)	.008 (.0014)	.006 (.0017)
1978							
Première	Aucun crime	.671 (.0087)	.104 (.0064)	.019 (.0031)	.671 (.0088)	.112 (.0044)	.019 (.0021)
interview	Crime contre la propriété	.119 (.0063)	.040 (.0044)	.010 (.0024)	.112 (.0044)	.040 (.0044)	.010 (.0017)
	Crime contre la personne	.019 (.0029)	.011 (.0025)	.006 (.0020)	.019 (.0021)	.010 (.0017)	.006 (.0020)

Nota: Les erreurs types estimées sont indiquées entre parenthèses.

Tableau 4b

Estimations des p_{ij} pour les flux dans les catégories de genre d'acte criminel suivant le modèle D-S

			Modèle symétrique		
			Deuxième interview		
			Aucun crime	Un seul crime	Crimes multiples
1975					
	Première	Aucun crime	.635 (.0101)	.118 (.0046)	.026 (.0026)
	interview	Crime contre la propriété	.118 (.0046)	.052 (.0046)	.011 (.0016)
		Crime contre la personne	.026 (.0026)	.011 (.0016)	.005 (.0016)
1976					
	Première	Aucun crime	.641 (.0098)	.110 (.0046)	.026 (.0028)
	interview	Crime contre la propriété	.110 (.0046)	.052 (.0048)	.015 (.0021)
		Crime contre la personne	.026 (.0028)	.015 (.0021)	.004 (.0019)
1977					
	Première	Aucun crime	.642 (.0104)	.120 (.0052)	.019 (.0024)
	interview	Crime contre la propriété	.120 (.0052)	.050 (.0049)	.011 (.0019)
		Crime contre la personne	.019 (.0024)	.011 (.0019)	.008 (.0022)
1978					
	Première	Aucun crime	.636 (.0117)	.121 (.0057)	.020 (.0025)
	interview	Crime contre la propriété	.121 (.0057)	.049 (.0055)	.012 (.0021)
		Crime contre la personne	.020 (.0025)	.012 (.0021)	.008 (.0025)

Nota: Les erreurs types estimées sont indiquées entre parenthèses.

Tableau 4c

Estimations des p_{ij} pour les flux dans les catégories de genre d'actes criminels déclaré suivant les modèles E-U et E-S

		Modèle non assujetti à des contraintes			Modèle symétrique		
		Deuxième interview					
		Aucun crime	Crime contre la propriété	Crime contre la personne	Aucun crime	Crime contre la propriété	Crime contre la personne
1975							
Première	Aucun crime	.636 (.0100)	.111 (.0062)	.024 (.0034)	.636 (.0101)	.117 (.0046)	.026 (.0026)
interview	Crime contre la propriété	.124 (.0063)	.053 (.0047)	.012 (.0023)	.117 (.0046)	.052 (.0047)	.011 (.0016)
	Crime contre la personne	.027 (.0033)	.009 (.0020)	.005 (.0016)	.026 (.0026)	.011 (.0016)	.005 (.0016)
1976							
Première	Aucun crime	.641 (.0098)	.110 (.0065)	.028 (.0041)	.641 (.0098)	.110 (.0046)	.026 (.0028)
interview	Crime contre la propriété	.110 (.0059)	.051 (.0048)	.014 (.0028)	.110 (.0046)	.052 (.0048)	.015 (.0021)
	Crime contre la personne	.024 (.0033)	.016 (.0028)	.005 (.0019)	.026 (.0028)	.015 (.0021)	.005 (.0019)
1977							
Première	Aucun crime	.636 (.0108)	.138 (.0076)	.023 (.0035)	.641 (.0105)	.121 (.0051)	.019 (.0024)
interview	Crime contre la propriété	.107 (.0060)	.050 (.0048)	.010 (.0022)	.121 (.0051)	.049 (.0048)	.011 (.0018)
	Crime contre la personne	.015 (.0028)	.011 (.0027)	.009 (.0023)	.019 (.0024)	.011 (.0018)	.009 (.0022)
1978							
Première	Aucun crime	.641 (.0117)	.111 (.0078)	.022 (.0040)	.640 (.0117)	.118 (.0056)	.021 (.0026)
interview	Crime contre la propriété	.124 (.0071)	.048 (.0055)	.012 (.0029)	.118 (.0056)	.048 (.0054)	.013 (.0021)
	Crime contre la personne	.020 (.0033)	.014 (.0031)	.009 (.0025)	.021 (.0026)	.013 (.0021)	.008 (.0025)

Nota: Les erreurs types estimées sont indiquées entre parenthèses.

4.2 Estimations des paramètres λ selon les modèles R, A, B et C

Le lecteur se souviendra que les estimations des paramètres λ selon les modèles R, A, B et C sont identiques, peu importe qu'on utilise le modèle non assujetti à des contraintes ou le modèle symétrique pour les paramètres p. Suivant la méthode itérative servant à estimer les paramètres λ selon les modèles A et C, le critère de convergence utilisé était que les estimations des paramètres λ varient d'au plus 0.0005 d'une étape à l'autre. Lorsque la convergence s'est produite après moins de 10,000 étapes, elle a nécessité de 41 à 4150 étapes à partir des estimations initiales des paramètres suggérées à l'annexe II. Dans certains cas, les facteurs de la fonction de vraisemblance pour les observations portant uniquement sur les paramètres λ n'étaient pas classiques. Cela est particulièrement vrai dans le cas des fonctions de vraisemblance pour les données de 1978 suivant les modèles A et C. En pareil cas, on a eu recours à une recherche par quadrillage pour déterminer les points de départ appropriés du processus itératif. Nous avons également effectué une recherche grossière par quadrillage dans tous les cas afin de vérifier si, lorsque la méthode itérative convergeait, elle semblait avoir convergé vers un maximum global plutôt que vers un maximum local.

Les estimations des paramètres λ établies pour le classement selon le nombre d'actes criminels déclarés et le classement selon le genre d'acte criminel déclaré suivant les modèles R, A, B et C sont présentées respectivement dans les tableaux 5, 6, 7 et 8.

Il convient de noter que, suivant les modèles R et B, les estimations des paramètres λ sont les mêmes pour les deux types de classement (selon le nombre d'actes criminels déclarés et selon le genre d'acte criminel déclaré), puisque la probabilité de non-réponse selon ces deux modèles n'est pas fonction de la catégorie de réponse. Suivant les modèles A et C, les estimations des paramètres λ correspondant à la catégorie "aucun crime" sont les mêmes, compte tenu de l'erreur d'arrondi, pour les deux types de classement (selon le nombre d'actes criminels déclarés et selon le genre d'acte criminel déclaré), puisque les ménages n'ayant pas été victimes de crime sont les mêmes dans les deux cas. On notera également que, suivant les modèles A et C, la valeur estimative du paramètre λ, la probabilité estimée d'être un non-répondant, s'accroît généralement en fonction du nombre d'actes criminels déclarés ou de la gravité de l'acte criminel déclaré.

Tableau 5
Estimations des λ suivant le modèle R

	Données classées selon le nombre d'actes criminels déclarés ou selon le genre d'acte criminel déclaré
	$\hat{\lambda}$
1975	.224 (.0035)
1976	.232 (.0035)
1977	.237 (.0036)
1978	.250 (.0040)

Nota: Les erreurs types estimées sont indiquées entre parenthèses.

Tableau 6
Estimations des λ_{1j} et des λ_{2i} suivant le modèle A

	Classement selon le nombre d'actes criminels déclarés						Classement selon le genre d'acte criminel déclaré					
	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{13}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\lambda}_{23}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{13}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\lambda}_{23}$
1975	.208	.272	.327	.221	.234	.275	.208	.280	.322	.220	.246	.246
	(.0062)	(.0159)	(.0261)	(.0064)	(.0147)	(.0242)	(.0062)	(.0151)	(.0321)	(.0064)	(.0139)	(.0303)
1976	.206*	.261*	.397*	.236*	.254*	.267*	.206	.278	.381	.235	.253	.285
	(.0063)	(.0152)	(.0268)	(.0066)	(.0153)	(.0248)	(.0063)	(.0146)	(.0327)	(.0066)	(.0144)	(.0319)
1977	.192	.263	.309	.258	.281	.326	.192	.275	.267	.258	.269	.417
	(.0064)	(.0152)	(.0265)	(.0070)	(.0171)	(.0285)	(.0064)	(.0144)	(.0327)	(.0069)	(.0159)	(.0369)
1978	.207*	.316*	.302*	.269*	.280*	.321*	.207*	.305*	.343*	.269*	.280*	.334*
	(.0072)	(.0182)	(.0308)	(.0079)	(.0176)	(.0300)	(.0072)	(.0174)	(.0364)	(.0079)	(.0166)	(.0362)

Nota: L'astérisque * indique les cas où la fonction de vraisemblance n'est pas classique.
Les erreurs types estimées sont indiquées entre parenthèses.

Tableau 7
Estimations de λ_1 et λ_2 suivant le modèle B

	Classement selon le nombre d'actes criminels déclarés ou selon le genre d'acte criminel déclaré	
	$\hat{\lambda}_1$	$\hat{\lambda}_2$
1975	.223	.226
	(.0058)	(.0058)
1976	.225	.240
	(.0059)	(.0060)
1977	.209	.264
	(.0059)	(.0064)
1978	.227	.273
	(.0067)	(.0071)

Nota: Les erreurs types estimées sont indiquées entre parenthèses.

Tableau 8
Estimations des λ_i suivant le modèle C

	Classement selon le nombre d'actes criminels déclarés			Classement selon le genre d'acte criminel déclaré		
	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$
1975	.214	.252	.300	.214	.262	.284
	(.0039)	(.0118)	(.0199)	(.0039)	(.0109)	(.0262)
1976	.221	.257	.330	.221	.266	.333
	(.0040)	(.0116)	(.0210)	(.0040)	(.0109)	(.0289)
1977	.225	.271	.317	.225*	.273*	.339*
	(.0041)	(.0126)	(.0235)	(.0041)	(.0115)	(.0286)
1978	.237*	.297*	.312*	.237*	.292*	.339*
	(.0046)	(.0139)	(.0236)	(.0046)	(.0130)	(.0299)

Nota: L'astérisque * indique les cas où la fonction de vraisemblance n'est pas classique.
Les erreurs types estimées sont indiquées entre parenthèses.

4.3 Estimations des paramètres suivant les modèles D et E

Comme ils exigent que tous les paramètres soient estimés simultanément, les modèles D et E sont plus difficiles à ajuster que les modèles R, A, B et C. Quel que soit l'ensemble de données de la NCS considéré, les fonctions de vraisemblance suivant les modèles D et E n'étaient pas classiques et il a fallu effectuer une recherche par quadrillage portant sur les valeurs possibles des paramètres λ afin de déterminer des points de départ appropriés pour la méthode itérative. Comme la recherche par quadrillage portant sur les six paramètres λ suivant le modèle D nécessitait beaucoup de temps, nous avons obtenu les estimations des paramètres selon le modèle D-S mais non selon le modèle D-U. Les estimations des paramètres p selon le modèle D-S sont présentées au tableau 3b pour le classement selon le nombre d'actes criminels déclarés et au tableau 4b pour le classement selon le genre d'acte criminel déclaré. Les estimations des paramètres λ suivant le modèle D-S sont présentées au tableau 9 pour les deux types de classement. Les estimations des paramètres p suivant les modèles E-U et E-S sont présentées au tableau 3c pour le classement selon le nombre d'actes criminels déclarés et au tableau 4c pour le classement selon le genre d'acte criminel déclaré. Enfin, les estimations des paramètres λ suivant les modèles E-U et E-S sont présentées au tableau 10 pour les deux types de classement.

Il convient de noter que, suivant les modèles D et E, la valeur estimée de p_{11}, la probabilité qu'une unité reste classée dans la catégorie "aucun crime", est légèrement inférieure à la valeur correspondante suivant les modèles R, A, B et C; les estimations des autres paramètres p suivant les modèles D et E sont toutefois légèrement supérieures aux estimations correspondantes suivant les modèles R, A, B et C. Selon les modèles D et E, la valeur estimative du paramètre λ, la probabilité estimée de non-réponse, s'accroît généralement en fonction du nombre d'actes criminels déclarés ou de la gravité de l'acte criminel déclaré. Lorsque la valeur des estimations décroît en fonction du nombre d'actes criminels déclarés ou de la gravité de l'acte criminel déclaré (pour les données de 1978 suivant le modèle D-S et pour les données de 1978 sur le nombre d'actes criminels déclarés suivant le modèle E-S), la diminution est faible et inférieure à l'erreur type estimée des estimations.

Tableau 9

Estimations de λ_{1i} et de λ_{2j} suivant le modèl D-S

	Classement selon le nombre d'actes criminels déclarés						Classement selon le genre d'acte criminel déclaré					
	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{13}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\lambda}_{23}$	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{13}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\lambda}_{23}$
1975	.210	.246	.319	.194	.321	.387	.208	.264	.319	.192	.339	.372
	(.0085)	(.0303)	(.0368)	(.0085)	(.0282)	(.0362)	(.0084)	(.0249)	(.0523)	(.0085)	(.0235)	(.0507)
1976	.204	.276	.339	.217	.273	.444	.203	.280	.383	.215	.297	.453
	(.0083)	(.0274)	(.0344)	(.0084)	(.0291)	(.0331)	(.0083)	(.0244)	(.0443)	(.0084)	(.0255)	(.0416)
1977	.175	.307	.380	.249	.298	.374	.175	.304	.438	.248	.315	.341
	(.0086)	(.0301)	(.0403)	(.0089)	(.0326)	(.0439)	(.0086)	(.0243)	(.0424)	(.0089)	(.0259)	(.0491)
1978	.211	.278	.290	.236	.413	.384	.211	.276	.293	.236	.411	.391
	(.0094)	(.0282)	(.0433)	(.0099)	(.0261)	(.0443)	(.0094)	(.0264)	(.0563)	(.0098)	(.0246)	(.0567)

Nota: Les erreurs types estimées sont indiquées entre parenthèses.

Tableau 10

Estimations de λ_i suivant le modèle E

	Classement selon le nombre d'actes criminels déclarés			Classement selon le genre d'acte criminel déclaré		
	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$
	p_{ij} non assujettis à des contraintes					
1975	.202 (.0060)	.285 (.0235)	.348 (.0262)	.201 (.0058)	.302 (.0180)	.336 (.0418)
1976	.211 (.0057)	.275 (.0226)	.387 (.0232)	.209 (.0056)	.286 (.0193)	.419 (.0327)
1977	.210 (.0063)	.315 (.0259)	.372 (.0351)	.209 (.0061)	.318 (.0183)	.394 (.0295)
1978	.224 (.0065)	.340 (.0208)	.342 (.0296)	.225 (.0065)	.326 (.0203)	.385 (.0333)
	p_{ij} symétriques					
1975	.202 (.0060)	.285 (.0235)	.351 (.0258)	.201 (.0059)	.301 (.0180)	.341 (.0408)
1976	.211 (.0057)	.274 (.0223)	.389 (.0229)	.209 (.0056)	.287 (.0191)	.418 (.0327)
1977	.213 (.0061)	.301 (.0267)	.376 (.0339)	.213 (.0060)	.309 (.0190)	.391 (.0302)
1978	.224 (.0065)	.343 (.0204)	.338 (.0298)	.225 (.0065)	.329 (.0199)	.379 (.0339)

Nota: Les erreurs types estimées sont indiquées entre parenthèses.

4.4 Ajustement des modèles

Le tableau 11 indique les valeurs de X^2 et G^2 ainsi que le nombre de degrés de liberté s'y rapportant pour l'ensemble des douze modèles (y compris le modèle D-U qui doit s'ajuster parfaitement aux données) et pour les deux types de classement des données. On notera que les modèles ont été ajustés pour illustrer les méthodes élaborées dans le présent article et que nous n'avons pas tenu compte de toute la complexité du plan de sondage. Bien que l'existence de grappes ne pose aucun problème dans notre sous-échantillon de données de la NCS, nous préférerions, dans une analyse plus approfondie, ajuster les modèles aux données des diverses strates séparément puis combiner les estimations relatives au strates afin d'obtenir des estimations pour l'ensemble de la population.

À l'évidence, ni le modèle R, modèle pour la non-réponse aléatoire, ni le modèle B, selon lequel la probabilité de non-réponse est fonction uniquement de la valeur de t, ne s'ajuste bien aux données que ce soit selon le modèle des p_{ij} non assujetties à des contraintes ou selon le modèle symétrique des p_{ij}.

Les modèles C-U et C-S s'ajustent assez bien aux données de 1975 et raisonnablement bien aux données de 1976. Comme le modèle C-S s'ajuste raisonnablement bien aux données et qu'il est plus parcimonieux, nous le préférons au modèle C-U. Selon le modèle C, la probabilité de non-réponse est fonction uniquement de la catégorie dans laquelle l'unité est classée au moment de l'interview à laquelle le ménage a répondu et non du moment où l'interview a eu lieu. Aussi le modèle C est-il le modèle de symétrie des probabilités de non-réponse dans le cadre des deux

Tableau 11
Ajustement des modèles

	Classement selon le nombre d'actes criminels déclarés				Classement selon le genre d'acte criminel déclaré			
	p_{ij} non assujetties à des contraintes		p_{ij} symétriques		p_{ij} non assujetties à des contraintes		p_{ij} symétriques	
	X^2	G^2	X^2	G^2	X^2	G^2	X^2	G^2
Modèle R	(d.l. = 5)		(d.l. = 8)		(d.l. = 5)		(d.l. = 8)	
1975	42.7	41.2	45.9	45.6	38.2	36.9	42.0	41.5
1976	70.2	67.1	69.7	67.7	57.7	55.9	58.3	56.4
1977	74.2	75.2	83.9	85.3	85.4	84.8	94.8	95.3
1978	61.7	62.7	64.9	66.3	63.2	64.1	65.5	66.8
Modèle A	(d.l. = 0)		(d.l. = 3)		(d.l. = 0)		(d.l. = 3)	
1975	0.0	0.0	4.4	4.4	0.0	0.0	4.6	4.6
1976	0.0	0.0	0.6	0.6	0.0	0.0	0.5	0.5
1977	0.0	0.0	10.1	10.1	0.0	0.0	10.5	10.5
1978	0.0	0.0	3.7	3.7	0.0	0.0	2.7	2.7
Modèle B	(d.l. = 4)		(d.l. = 7)		(d.l. = 4)		(d.l. = 7)	
1975	42.7	41.1	45.9	45.5	38.2	36.9	42.0	41.5
1976	69.1	64.5	68.5	65.1	56.2	53.3	56.9	53.8
1977	47.1	45.4	58.7	55.5	57.0	54.9	68.4	65.4
1978	47.6	46.0	50.1	49.6	49.1	47.4	50.7	50.1
Modèle C	(d.l. = 3)		(d.l. = 6)		(d.l. = 3)		(d.l. = 6)	
1975	6.9	6.9	11.3	11.3	7.4	7.4	12.0	12.0
1976	21.2	21.3	21.8	21.9	15.1	15.1	15.6	15.6
1977	38.1	38.3	48.2	48.4	45.6	45.7	56.0	56.3
1978	31.1	31.1	34.7	34.8	29.9	30.0	32.6	32.7
Modèle D	(d.l. = 0)		(d.l. = 3)		(d.l. = 0)		(d.l. = 3)	
1975	0.0	0.0	5.0	5.0	0.0	0.0	5.6	5.6
1976	0.0	0.0	15.3	15.3	0.0	0.0	11.6	11.6
1977	0.0	0.0	11.5	11.5	0.0	0.0	18.0	18.0
1978	0.0	0.0	10.2	10.2	0.0	0.0	9.9	9.8
Modèle E	(d.l. = 3)		(d.l. = 6)		(d.l. = 3)		(d.l. = 6)	
1975	7.0	7.0	11.3	11.3	7.3	7.3	12.0	12.0
1976	21.0	21.1	21.8	21.9	14.8	14.9	15.6	15.6
1977	33.0	33.0	48.2	48.4	39.5	39.5	56.0	56.3
1978	32.0	32.1	34.6	34.8	30.9	31.0	32.6	32.7

Nota: $\chi^2_{.99}(3) = 11.34$, $\chi^2_{.99}(4) = 13.28$, $\chi^2_{.99}(5) = 15.09$, $\chi^2_{.99}(6) = 16.81$, $\chi^2_{.99}(7) = 18.48$, et $\chi^2_{.99}(8) = 20.09$.

cycles d'interview. Lorsque le modèle C est combiné avec le modèle symétrique pour les paramètres p, nous obtenons des fréquences prévues de cellule symétriques pour les flux d'observations. Notons que, selon les observations montrées dans l'annexe 1, le nombre de cas de non-réponse est beaucoup plus élevé au deuxième qu'au premier cycle d'interview des années 1977 et 1978. C'est en raison de cet écart entre les taux de non-réponse que le modèle C s'ajuste moins bien aux données de 1977 et de 1978.

L'ajustement des modèles E-U et E-S à l'échantillon d'observations est presque identique à celui des modèles C-U et C-S respectivement. Cela n'a rien de surprenant quand on sait que les interprétations des modèles sont assez semblables. Suivant le modèle C, la non-réponse est fonction de la catégorie de réponse dans laquelle est classé le ménage lorsqu'il répond, tandis que suivant le modèle E, elle est fonction de la catégorie de réponse dans laquelle est classé le ménage lorsqu'il ne répond pas. Comme l'ajustement de ces deux modèles est semblable,

il nous est impossible d'opter en faveur de l'un ou de l'autre en nous fondant uniquement sur les observations. Logiquement, le modèle E semble plus réaliste, puisque nous pouvons nous attendre à ce que la non-réponse varie selon que le ménage a été ou non victime d'actes criminels. Comme les deux modèles s'ajustent aux données de façon semblable, il se peut que l'état (victime ou non-victime) dans lequel se trouve le ménage au moment où il répond constitue, en règle générale, un bon indicateur de l'état dans lequel il se trouve au moment où il ne répond pas. Si cette hypothèse se confirmait, nous préférerions utiliser le modèle C puisqu'il est plus facile à ajuster que le modèle E.

Le modèle A-S, suivant lequel la non-réponse est fonction du moment où l'interview a lieu et de l'état dans lequel se trouve le ménage lorsqu'il répond, s'ajuste très bien aux données de 1975, 1976 et 1978 et raisonnablement bien aux données de 1977. L'ajustement du modèle D-S est semblable à celui du modèle A-S, sauf dans le cas des données de 1976 auxquelles il s'ajuste beaucoup mieux que le modèle A-S. Encore une fois, il nous est impossible d'opter pour le modèle A ou le modèle D en nous fondant uniquement sur les observations. (Les modèles A-U et D-U s'ajustent parfaitement aux données.) De façon générale, nous sommes très satisfaits de l'ajustement du modèle A-S tant aux données sur le nombre d'actes criminels qu'aux données sur le genre d'acte criminel, et ce pour les quatre années considérées. Comme le modèle A s'ajuste raisonnablement bien à toutes les données, nous devons conclure que la non-réponse observée dans le cadre de la NCS varie selon que le ménage a été ou non victime d'actes criminels.

Il convient de noter que, dans la majorité des cas, l'ajustement des modèles mesuré à l'aide de X^2 et G^2 varie très peu lorsqu'on utilise le modèle des p_{ij} symétriques plutôt que celui des p_{ij} non assujetties à des contraintes. Comme le modèle symétrique, plus parcimonieux, nous permet de gagner 3 degrés de liberté, nous le préférons au modèle non assujetti à des contraintes. Le fait que nous choisissions le modèle symétrique pour les probabilités de flux indique que le nombre d'actes criminels déclarés dans le cadre des deux cycles annuels d'interview de la NCS est relativement stable. Cette stabilité découle du fait que la symétrie des probabilités de flux sous-jacentes implique l'égalité des totaux marginaux. Aussi, le nombre de ménages ne déclarant aucun crime, déclarant un crime ou déclarant plus d'un crime demeure à peu près le même du premier au deuxième cycle d'interview d'une année. Il en va de même du nombre de ménages n'ayant été victimes d'aucun crime, ayant été victimes d'un crime contre la propriété ou ayant été victimes d'un crime contre la personne.

5. CONCLUSIONS ET PISTES DE RECHERCHE

Nous avons vu que, une fois combiné avec un modèle suivant lequel la non-réponse est fonction à la fois de la valeur de t et de l'état (victime ou non-victime) dans lequel se trouve le ménage, le modèle décrivant la symétrie des matrices de flux dans les catégories d'actes criminels déclarés s'ajuste bien aux sommaires de données de la NCS. De plus, ce modèle s'ajuste aux données peu importe que les ménages soient classés selon le nombre d'actes criminels déclarés ou selon le genre d'acte criminel déclaré.

Bien sûr, les présents travaux ne constituent qu'une tentative initiale d'étudier la non-réponse et les flux dans les catégories d'actes criminels déclarés dans le cadre de la NCS. Ainsi, nous avons souligné que les probabilités estimées de flux symétriques dans les catégories n'ont pas semblé fluctuer beaucoup au cours de la période de quatre ans allant de 1975 à 1978, tandis que les probabilités estimées de non-réponse ont semblé afficher une certaine variation. Il pourrait se révéler intéressant d'ajuster aux données de la NCS un modèle décrivant à la fois des probabilités de flux constantes et des probabilités de non-réponse variant en fonction de la valeur de t. Si les probabilités de non-réponse fluctuent réellement en fonction du moment où l'interview a eu lieu, non seulement d'une année à l'autre mais d'un cycle d'interview à l'autre, il est essentiel d'essayer de découvrir la cause de ces fluctuations.

Dans le présent article, toutes les données manquantes ont été traitées de la même façon. De fait, les données peuvent être manquantes parce qu'une UL a été supprimée de l'échantillon par renouvellement, parce qu'un ménage a emménagé dans l'UL sélectionnée ou l'a quittée, parce que personne n'était sur place, parce que le ménage a refusé de répondre ou pour quelque autre raison. On peut raisonnablement supposer que la non-réponse découlant du fait qu'une UL a été supprimée de l'échantillon par renouvellement constitue une non-réponse aléatoire, mais que les autres types de non-réponse ne sont pas aléatoires. À cet égard, Stasny (1988) présente des modèles pour différents types de non-réponse qui peuvent être combinés aux modèles décrivant la symétrie des flux présentés dans cet article. En outre, les modèles que nous avons étudiés ne permettent pas de tenir compte des ménages qui ne participent à aucun des deux cycles d'interview. Comme il existe sûrement de tels ménages, il pourrait être intéressant d'étudier les modèles markoviens, comme ceux qu'on trouve dans Stasny (1987), qui permettent de tenir compte de la non-réponse aux deux cycles d'interview.

Surtout, il serait sûrement utile d'étudier des sommaires de données plus naturels que ceux auxquels nous avons eu recours. Alors que nous avons utilisé des sommaires portant sur chaque cycle d'interview, il serait plus révélateur d'utiliser des sommaires mensuels ou trimestriels. En pareil cas, il faudrait prendre en considération la nature complexe du calendrier d'interviews de la NCS et en tenir compte dans les modèles. Ainsi, le code de réponse attribué à un ménage resterait inchangé pour la totalité de la période de référence de six mois sur laquelle porte une interview. L'élaboration de modèles tenant compte de ce paramètre constitue à n'en pas douter une piste de recherche qu'il importe d'explorer.

REMERCIEMENTS

Les données de la National Crime Survey sur lesquelles se fonde le présent article ont été mises à notre disposition par le Consortium interuniversitaire pour la recherche politique et sociale. Ces données ont été initialement recueillies par la United States Law Enforcement Assitance Administration. L'ensemble de données longitudinales utilisé a été créé par le Bureau of Justice Statistics, à partir des fichiers de données trimestrielles à grande diffusion. La présente recherche a été rendue possible en partie grâce à une subvention du Bureau of Justice Statistics, U.S. Department of Justice, et du Commitee on Law and Justice Statistics, American Statistical Association, qui a permis à l'auteure d'assister aux deux ateliers sur la conception et l'utilisation de la National Crime Survey. Les opinions exprimées dans le cadre du présent article sont entièrement celles de l'auteure. L'auteure remercie sincèrement les membres du comité de lecture pour leurs commentaires relatifs à une version antérieure du mémoire.

ANNEXE I

Les observations

	Classement selon le nombre d'actes criminels déclarés			
	Deuxième interview			
	Aucun crime	Un seul crime	Crimes multiples	Non-réponse
1975 - **Première interview**				
Aucun crime	1963	256	67	901
Un crime	306	73	31	179
Crimes multiples	95	26	24	83
Non-réponse	866	193	91	
1976 - **Première interview**				
Aucun crime	1884	257	53	951
Un crime	266	84	24	186
Crimes multiples	82	34	18	75
Non-réponse	831	197	106	
1977 - **Première interview**				
Aucun crime	1742	260	66	994
Un crime	228	56	31	177
Crimes multiples	63	31	10	76
Non-réponse	716	194	79	
1978 - **Première interview**				
Aucun crime	1370	157	45	831
Un crime	222	50	14	165
Crimes multiples	50	18	19	66
Non-réponse	651	174	57	

	Classement selon le genre d'acte criminel déclaré			
	Deuxième interview			
	Aucun crime	Crime contre la propriété	Crime contre la personne	Non-réponse
1975 - **Première interview**				
Aucun crime	1963	271	52	901
Crime contre la propriété	331	107	22	217
Crime contre la personne	70	17	8	45
Non-réponse	866	225	59	
1976 - **Première interview**				
Aucun crime	1884	266	44	951
Crime contre la propriété	295	111	19	211
Crime contre la personne	53	26	4	50
Non-réponse	831	235	68	
1977 - **Première interview**				
Aucun crime	1742	283	43	994
Crime contre la propriété	262	89	18	194
Crime contre la personne	29	12	9	59
Non-réponse	716	231	42	
1978 - **Première interview**				
Aucun crime	1370	173	29	831
Crime contre la propriété	238	64	14	184
Crime contre la personne	34	15	8	47
Non-réponse	651	184	47	

ANNEXE II: Méthodes d'obtention des EMV des paramètres p et λ

Notons que $x.. = \sum_{i=1}^{K} \sum_{j=1}^{K} x_{ij}$ est le nombre total d'unités participant aux deux cycles d'interview et $n = x.. + x._M + x_{M.}$ est la taille de l'échantillon global. Les valeurs de départ utilisées ci-dessous le sont uniquement à titre indicatif. On peut utiliser comme valeurs initiales pour les estimations des paramètres p toutes autres valeurs positives dont la somme est égale à 1 et, comme valeurs initiales pour les paramètres λ, toutes autres valeurs situées entre zéro et un.

EMV pour les p_{ij} non assujetties à des contraintes suivant les modèles R, A, B et C

1. $p_{ij}^{(0)} = x_{ij}/x..$

2. $p_{ij}^{(\nu+1)} = \left[x_{ij} + x_{iM} p_{ij}^{(\nu)}/p_{i.}^{(\nu)} + x_{Mj} p_{ij}^{(\nu)}/p_{.j}^{(\nu)}\right]/n.$

On répète l'étape 2 pour $\nu = 0, 1, 2, \ldots$ jusqu'à ce que les estimations des paramètres p_{ij} convergent vers le degré d'exactitude prescrit.

EMV pour les λ suivant le modèle A

1. $\lambda_{1j}^{(0)} = x_{M.}/n$ et $\lambda_{2i}^{(0)} = x._M/n.$

2. a) $\lambda_{1j}^{(\nu+1)} = x_{Mj} \Big/ \sum_{i=1}^{K} \left[x_{ij}/\left(1 - \lambda_{1j}^{(\nu)} - \lambda_{2i}^{(\nu)}\right)\right]$

b) $\lambda_{2i}^{(\nu+1)} = x_{iM} \Big/ \sum_{j=1}^{K} \left[x_{ij}/\left(1 - \lambda_{1j}^{(\nu)} - \lambda_{2i}^{(\nu)}\right)\right].$

On répète l'étape 2 pour $\nu = 0, 1, 2, \ldots$ jusqu'à ce que les estimations des paramètres λ convergent vers le degré d'exactitude prescrit. Si $x_{hM} > \sum_{j=1}^{K} x_{hj}$ ou $x_{Mh} > \sum_{i=1}^{K} x_{ih}$ pour certaines valeurs de h, de telle sorte que, parmi toutes les unités ayant été classées dans une catégorie donnée à l'occasion d'un cycle d'interview, le nombre d'unités n'ayant pas participé à l'autre cycle d'interview est plus élevé que celui des unités y ayant participé, alors la valeur des estimations des paramètres correspondants se situera à l'extérieur de la plage des valeurs prescrites (0 à 1) à l'une ou l'autre étape et il faudra utiliser d'autres formules à la place de celles exposées ci-devant (voir Chen et Fienberg 1974). Si pour certaines valeurs de j $x_{Mj} > \sum_{i=1}^{K} x_{ij}$, il faut, pour ces valeurs, remplacer l'étape 2a) ci-devant par

$$\lambda_{1j}^{(\nu+1)} = 1 - \lambda_{2h}^{(\nu)} - \left(\lambda_{1j}^{(\nu)}/x_{Mj}\right)\left\{\sum_{i=1}^{K}\left[x_{ij}/\left(1 - \lambda_{1j}^{(\nu)} - \lambda_{2i}^{(\nu)}\right)\right]\right\}\left(1 - \lambda_{1j}^{(\nu)} - \lambda_{2h}^{(\nu)}\right),$$

où h est choisi à chaque étape de la méthode itérative de façon à ce que $\lambda_{2h}^{(\nu)} \geq \lambda_{2i}^{(\nu)}$ pour tous les $i = 1, 2, \ldots K$. Si pour certaines valeurs de i, $x_{iM} > \sum_{j=1}^{K} x_{ij}$, il faut, pour ces valeurs, remplacer l'étape 2b) ci-devant par

$$\lambda_{2i}^{(\nu+1)} = 1 - \lambda_{1h}^{(\nu)} - \left(\lambda_{2i}^{(\nu)}/x_{iM}\right)\left\{\sum_{j=1}^{K}\left[x_{ij}/\left(1 - \lambda_{1j}^{(\nu)} - \lambda_{2i}^{(\nu)}\right)\right]\right\}\left(1 - \lambda_{1h}^{(\nu)} - \lambda_{2i}^{(\nu)}\right),$$

où h est choisi à chaque étape de la méthode de façon à ce que $\lambda_{1h}^{(\nu)} \geq \lambda_{1j}^{(\nu)}$ pour tous les $j = 1, 2, \ldots K$.

EMV pour les λ suivant le modèle B

$$\hat{\lambda}_1 = x_{M\cdot}/n \quad \text{et} \quad \hat{\lambda}_2 = x_{\cdot M}/n.$$

EMV pour les λ suivant le modèle C

1. $\lambda_i^{(0)} = (x_{iM} + x_{Mi})/2n.$

2. $$\lambda_i^{(\nu+1)} = (x_{iM} + x_{Mi}) \Big/ \left\{ \sum_{j=1}^{K} \left[(x_{ij} + x_{ji}) / \left(1 - \lambda_i^{(\nu)} - \lambda_j^{(\nu)}\right) \right] \right\}.$$

On répète l'étape 2 pour $\nu = 0, 1, 2, \ldots$ jusqu'à ce que les estimations des paramètres λ convergent vers le degré d'exactitude prescrit. Si $x_{Mi} + x_{iM} > \sum_{j=1}^{K}(x_{ij} + x_{ji})$ pour certaines valeurs de i, il faut, tout comme dans le cas du modèle A, utiliser une autre formule à la place de l'étape 2 ci-devant. En pareil cas, on remplace l'étape 2 par

$$\lambda_i^{(\nu+1)} = 1 - \lambda_h^{(\nu)} - \left[\lambda_i^{(\nu)} / (x_{iM} + x_{Mi})\right] \left\{ \sum_{j=1}^{K} \left[(x_{ij} + x_{ji}) / \left(1 - \lambda_i^{(\nu)} - \lambda_j^{(\nu)}\right) \right] \right\} \left(1 - \lambda_h^{(\nu)} - \lambda_i^{(\nu)}\right),$$

où h est choisi à chaque étape de la méthode itérative de façon à ce que $\lambda_h^{(\nu)} \geq \lambda_j^{(\nu)}$ pour tous les $j = 1, 2, \ldots K$.

EMV pour les paramètres suivant le modèle D-U

1. $p_{ij}^{(0)} = x_{ij}/x..,\quad \lambda_{1i}^{(0)} = x_{M\cdot}/n, \quad \text{et} \quad \lambda_{2j}^{(0)} = x_{\cdot M}/n.$

2. $$p_{ij}^{(\nu+1)} = n^{-1}\left\{ x_{ij} + x_{iM}\left[p_{ij}^{(\nu)}\lambda_{2j}^{(\nu)} \Big/ \sum_{h=1}^{K} p_{ih}^{(\nu)}\lambda_{2h}^{(\nu)} \right] + x_{Mj}\left[p_{ij}^{(\nu)}\lambda_{1i}^{(\nu)} \Big/ \sum_{h=1}^{K} p_{hj}^{(\nu)}\lambda_{1h}^{(\nu)} \right] \right\}$$

$$\lambda_{1i}^{(\nu+1)} = \sum_{j=1}^{K} \left[x_{Mj} p_{ij}^{(\nu)} \lambda_{1i}^{(\nu)} \Big/ \sum_{h=1}^{K} p_{hj}^{(\nu)}\lambda_{1h}^{(\nu)} \right] \Big/ \sum_{j=1}^{K} \left[x_{ij} / \left(1 - \lambda_{1i}^{(\nu)} - \lambda_{2j}^{(\nu)}\right) \right]$$

$$\lambda_{2j}^{(\nu+1)} = \sum_{i=1}^{K} \left[x_{iM} p_{ij}^{(\nu)} \lambda_{2j}^{(\nu)} \Big/ \sum_{h=1}^{K} p_{ih}^{(\nu)}\lambda_{2h}^{(\nu)} \right] \Big/ \sum_{i=1}^{K} \left[x_{ij} / \left(1 - \lambda_{1i}^{(\nu)} - \lambda_{2j}^{(\nu)}\right) \right].$$

On répète l'étape 2 pour $\upsilon = 0, 1, 2, \ldots$ jusqu'à ce que les estimations des paramètres λ convergent vers le degré d'exactitude prescrit.

EMV pour les paramètres suivant le modèle E-U

1. $p_{ij}^{(0)} = x_{ij}/x..$ et $\lambda_i^{(0)} = (x_{M\cdot} + x_{\cdot M})/2n.$

2. $$p_{ij}^{(\nu+1)} = n^{-1}\left\{x_{ij} + x_{iM}\left[p_{ij}^{(\nu)}\lambda_j^{(\nu)} \Big/ \sum_{h=1}^{K} p_{ih}^{(\nu)}\lambda_h^{(\nu)}\right] + x_{Mj}\left[p_{ij}^{(\nu)}\lambda_i^{(\nu)} \Big/ \sum_{h=1}^{K} p_{hj}^{(\nu)}\lambda_h^{(\nu)}\right]\right\}$$

$$\lambda_i^{(\nu+1)} = \left\{\sum_{j=1}^{K} x_{jM}\left[p_{ji}^{(\nu)}\lambda_i^{(\nu)} \Big/ \sum_{h=1}^{K} p_{jh}^{(\nu)}\lambda_h^{(\nu)}\right] + x_{Mj}\left[p_{ij}^{(\nu)}\lambda_i^{(\nu)} \Big/ \sum_{h=1}^{K} p_{hj}^{(\nu)}\lambda_h^{(\nu)}\right]\right\}$$

$$\times \left\{\sum_{j=1}^{K} (x_{ij} + x_{ji})/\left(1 - \lambda_i^{(\nu)} - \lambda_j^{(\nu)}\right)\right\}^{-1}.$$

On répète l'étape 2 pour $\nu = 0, 1, 2, \ldots$ jusqu'à ce que les estimations des paramètres λ convergent vers le degré d'exactitude prescrit.

BIBLIOGRAPHIE

CHEN, T., et FIENBERG, S.E. (1974). Two-Dimensional Contingency Tables With Both Completely and Partially Cross-Classified Data. *Biometrics*, 30, 629-642.

DODGE, R.W., et SKOGAN, W.G. (1987). The History, Organization and Utilization of the National Crime Survey. Mémoire présenté au Workshop on the National Crime Survey, du 6 au 17 juillet 1987, University of Maryland.

FIENBERG, S.E. (1980). The Measurement of Crime Victimization: Prospects for Panel Analysis of a Panel Survey. *The Statistician*, 29, 313-350.

LITTLE, R.J.A., et RUBIN, D.B. (1987). *Statistical Analysis with Missing Data*, New York: Wiley & Sons.

MONTAGLIANI, H. (1987). NCS Design. Mémoire présenté au Workshop on the National Crime Survey, du 6 au 17 juillet 1987, University of Maryland.

SAPHIRE, D.G. (1984). *Estimation of Victimization Prevalence Using Data from the National Crime Survey*. Lecture Notes in Statistics, Vol. 23. New York: Springer-Verlag.

STASNY, E.A. (1986). Estimating Gross Flows Using Panel Data With Nonresponse: An Example From the Canadian Labour Force Survey. *Journal of the American Statistical Association*, 81, 42-47.

STASNY, E.A. (1987). Some Markov-Chain Models for Nonresponse in Estimating Gross Labor Force Flows. *Journal of Official Statistics*, 3, 359-373.

STASNY, E.A. (1988). Modeling Non-Ignorable Non-Response in Categorical Panel Data With an Example in Estimating Gross Labor-Force Flows. *Journal of Business and Economic Statistics*, 6, 207-219.

TAYLOR, B.T. (1987). Redesign of the National Crime Survey. Mémoire présenté au Workshop on the National Crime Survey, du 6 au 17 juillet 1987, University of Maryland.

U.S. DEPARTMENT OF JUSTICE et BUREAU OF JUSTICE STATISTICS (1981). *The National Crime Survey: Working Papers, Volume I: Current and Historical Perspectives*. Washington, DC.