Techniques d'enquête, décembre 1990
Vol. 16, n° 2, pp. 297–306
Statistique Canada

Estimation des coefficients de corrélation de panel pour l'Enquête sur la population active du Canada

HYUNSHIK LEE[1]

RÉSUMÉ

L'enquête sur la population active du Canada repose sur un plan avec renouvellement de panel. À chaque mois, on renouvelle un sixième de l'échantillon global et on en conserve les cinq sixièmes. Ainsi, une fois qu'un panel est introduit dans l'échantillon, il y demeure 6 mois avant d'en être exclu. Cette caractéristique du plan de sondage ainsi que le mode de sélection des panels font que les estimations de panel pour le même mois ou des mois différents sont corrélées. La corrélation entre deux estimations de panel est désignée comme la corrélation de panel. Nous définissons trois types de corrélation de panel dans cet article: (1) la corrélation (désignée par ρ) entre des estimations de la même caractéristique tirées du même panel à des mois différents; (2) la corrélation (désignée par γ) entre des estimations de la même caractéristique tirées de panels géographiquement rapprochés l'un de l'autre à des mois différents; (3) la corrélation (désignée par τ) entre des estimations de caractéristiques différentes tirées du même panel dans le même mois ou dans des mois différents. En deuxième lieu, nous décrivons des méthodes permettant d'estimer les corrélations de panel et calculons les coefficients de corrélation estimés pour certaines variables en nous servant de données pour 1980-1981 et 1985-1987; nous terminons par une analyse des résultats.

MOTS CLÉS: Enquête par panel; renouvellement; méthode de Taylor.

1. INTRODUCTION

L'enquête sur la population active (EPA) est une enquête permanente menée auprès des ménages à chaque mois et qui repose sur un plan avec renouvellement de panel. L'échantillon de l'EPA est constitué de six panels de même taille qui, à tour de rôle, font place à un nouveau panel à chaque mois. Le nouveau panel passe six mois dans l'échantillon avant d'en être supprimé. (Pour une description détaillée de la méthodologie de l'EPA, voir Platek et Singh (1976) et Singh et coll. (1990).) Par conséquent, il existe une forte corrélation entre les estimations tirées du même panel, donc des mêmes unités d'échantillonnage, à des mois différents. En outre, les panels qui sont supprimés de l'échantillon sont remplacés habituellement par un panel voisin. La proximité géographique des panels fait aussi que les estimations respectives sont corrélées. On appelle ce genre de corrélations des corrélations de panel. Dans cet article, nous voyons comment on peut estimer ces corrélations et présentons les coefficients de corrélation estimés pour certaines variables. Cette étude avait pour but initialement d'analyser la méthode d'estimation composite; ses résultats peuvent néanmoins s'appliquer à toute situation où intervient la corrélation de panel.

Notre article se divise comme suit. Dans la section 2, nous présentons les définitions, les symboles et les hypothèses nécessaires. Dans la section suivante, nous décrivons les méthodes d'estimation pertinentes. Enfin, la section 4 renferme les résultats de l'estimation ainsi qu'une analyse.

[1] H. Lee, Division des méthodes d'enquêtes-entreprises, Statistique Canada, 11ième étage, Immeuble R.H. Coats, Parc Tunney, Ottawa, Ontario, K1A 0T6.

2. DÉFINITION DES COEFFICIENTS DE CORRÉLATION DE PANEL

Pour définir divers types de corrélation de panel, il faut tout d'abord définir la notion de "panels communs" et de "panel antécédent". Un panel est identifié par un numéro qui indique depuis combien de mois ce panel est présent dans l'échantillon. Ainsi, Panel 1 au mois m devient Panel 2 au mois $m + 1$, Panel 3 au mois $m + 2$, et ainsi de suite. On utilise souvent le terme groupe de renouvellement pour désigner un panel, peu importe depuis combien de mois ce panel est présent dans l'échantillon. Par exemple, le groupe de renouvellement 1 introduit dans l'échantillon en janvier est toujours identifié comme le groupe 1 jusqu'à ce qu'il soit supprimé de l'échantillon, en juillet. Ainsi, Panel 1 en janvier désigne le groupe de renouvellement 1 et Panel 2 en février désigne le même groupe de renouvellement, qui en est à son deuxième mois dans l'échantillon, et ainsi de suite.

Deux panels qui représentent le même groupe de renouvellement à des mois différents sont désignés comme des *panels communs*. Lorsqu'un groupe de renouvellement est supprimé de l'échantillon, il est remplacé le plus souvent par un groupe formé de ménages voisins et auquel on attribue le même numéro de renouvellement. Le panel associé au groupe de renouvellement que l'on supprime est le *panel antécédent* à celui qui est associé au nouveau groupe. Ainsi, dans l'exemple ci-dessus, Panel 6 en juin, qui représente le groupe de renouvellement 1, est le panel antécédent au Panel 1 de juillet. Le tableau 1 montre schématiquement les panels communs et les panels antécédents pour les mois donnés m et $m - j$.

Puisqu'on peut identifier chaque panel par deux éléments (mois et numéro de panel), désignons par P(mois, numéro de panel) un panel quelconque. Alors, P$(m, 4)$ et P$(m - 1, 3)$, par exemple, sont des panels communs à 1 mois d'intervalle. De même, P$(m, 4)$ et P$(m - 2, 2)$, sont des panels communs à 2 mois d'intervalle. On désigne par ρ_j le coefficient de corrélation des valeurs estimées d'une caractéristique établies à partir de panels communs à j mois d'intervalle. Évidemment, il ne peut exister de panels communs à plus de 5 mois d'intervalle; la valeur de l'indice j ne peut donc excéder 5. Nous supposons que ρ_j est indépendant de m et du numéro de panel. En revanche, il est fonction de j et varie selon les caractéristiques.

On désigne par γ_j le coefficient de corrélation des valeurs estimées d'une caractéristique établies à partir d'un panel donné et de son panel antécédent, séparés par j mois d'intervalle. Dans ce cas toutefois, la valeur de j peut atteindre 11; autrement dit, γ_{11} est le dernier coefficient de corrélation de la série et il désigne la corrélation entre les estimations de P$(m, 6)$ et celles de P$(m - 11, 1)$. Nous supposons là aussi que γ est indépendant de m et du numéro de panel. En revanche, comme le coefficient précédent, il est fonction de j et varie selon les caractéristiques.

Le troisième type de corrélation de panel est la corrélation entre les valeurs estimées de deux caractéristiques différentes, établies à partir de panels communs à j mois d'intervalle; ce type de corrélation est désignée par τ_j. Dans ce cas, la valeur de l'indice j peut varier de 0 à 5 et nous posons les mêmes hypothèses que pour ρ et γ.

Tableau 1

Panels communs et panels antécédents pour les mois m et $m - j$

m	$m-1$	$m-2$	$m-3$	$m-4$	$m-5$	$m-6$	$m-7$	$m-8$	$m-9$	$m-10$	$m-11$
1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))	((3))	((2))
2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))	((3))
3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))
4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))
5	4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))
6	5	4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)

Nota: Les parenthèses simples désignent des *panels antécédents* et les doubles, des *seconds panels antécédents*.

Voici donc les définitions formelles de ρ, de γ et de τ:

Soit $y_{m,l}$ la valeur estimée d'une caractéristique d'intérêt, établie à partir du panel $P(m,l)$ de l'EPA. Nous supposons que $V(y_{m,l}) = \sigma_y^2$, quels que soient m et l. Alors, ρ_j est défini au moyen de l'équation

$$\text{Cov}(y_{m,l},y_{m-j,l-j}) = \rho_j\sigma_y^2, \quad 1 \leq j \leq 5, \quad j < l \leq 6,$$

et γ_j, au moyen de l'équation

$$\text{Cov}(y_{m,l},y_{m-j,6+l-j}) = \gamma_j\sigma_y^2,$$

où $1 \leq l \leq j$ si $1 \leq j \leq 6$ et $j - 5 \leq l \leq 6$ si $7 \leq j \leq 11$.

Il serait normal de supposer que ρ_j et γ_j diminuent lorsque j augmente et que ρ_j est plus grand que γ_j étant donné que le premier désigne une corrélation qui met en rapport les mêmes ménages tandis que le second désigne une corrélation qui met en rapport des ménages voisins les uns des autres. Nous pouvons aussi définir de la même manière la corrélation entre un panel donné et le panel qui précède son panel antécédent (appelé *second panel antécédent* et identifié par des parenthèses doubles dans le tableau 1) en désignant cette corrélation par δ; nous avons ainsi $\delta_7, \delta_8, \ldots, \delta_{17}$. Les valeurs de δ seront moins élevées que celles de γ_j mais pourraient s'en rapprocher sensiblement, pour le même indice, à cause de la proximité géographique du panel antécédent et du second panel antécédent. Toutefois, compte tenu du temps et des ressources dont nous disposons, nous n'étudierons pas ici les coefficients de corrélation δ.

Nous supposons que $\text{Cov}(y_{m,l},y_{m,l'}) = 0$ si $l \neq l'$ et que $\text{Cov}(y_{m,l},y_{m-j,l'}) = 0$ si $P(m - j,l')$ et $P(m,l)$ ne sont pas des panels communs et si le premier n'est pas non plus le panel antécédent du second.

Afin de définir le coefficient de corrélation τ, posons $x_{m,l}$ comme la valeur estimée d'une autre caractéristique, établie à partir du panel $P(m,l)$ de l'EPA, et posons comme hypothèse que $V(x_{m,l}) = \sigma_x^2$ est indépendante de m et de l. Alors, le coefficient de corrélation τ peut être déterminé à l'aide de l'équation

$$\text{Cov}(y_{m,l},y_{m-j,l-j}) = \tau_j\sigma_x\sigma_y, \quad 0 \leq j \leq 5, \quad j < l \leq 6.$$

3. ESTIMATION DES COEFFICIENTS DE CORRÉLATION DE PANEL

Comme nous disposions d'un programme d'ordinateur pour l'estimation de la variance, la méthode décrite ci-dessous a été élaborée de manière que nous puissions utiliser ce programme sans devoir y apporter trop de modifications. La méthode utilisée en l'occurrence est la méthode généralisée de Keyfitz (Choudhry et Lee 1987; Lee 1989a), mieux connue sous le nom de méthode de Taylor. Le programme peut calculer des estimations de la variance de combinaisons linéaires de valeurs estimées mensuelles.

Pour estimer les coefficients de corrélation voulus à l'aide du programme mentionné plus haut, nous utilisons l'équation fondamentale suivante:

$$\text{Cov}(A,B) = \frac{V(A) + V(B) - V(A - B)}{2}. \tag{1}$$

$V(A - B), V(A)$ et $V(B)$ peuvent être déterminées à l'aide du programme et $\text{Cov}(A,B)$ peut être calculée au moyen de l'expression ci-dessus. On trouvera dans Kish (1965) une expression pour $V(A - B)$ dont on peut déduire l'équation (1).

3.1 Estimation des coefficients de corrélation ρ

Soit $A = \sum_{l=2}^{6} y_{m,l}$ et $B = \sum_{l=1}^{5} y_{m-1,l}$. On obtient les valeurs A et B en faisant abstraction, dans le premier cas, du Panel 1 et, dans le second cas, du Panel 6. Notons que les panels éliminés ne sont pas des panels communs alors que tous les autres le sont. À l'aide du programme mentionné plus haut, nous calculons les valeurs estimées de $V(A - B)$, $V(A)$ et de $V(B)$, puis nous nous servons de l'équation (1) pour estimer $\text{Cov}(A,B)$. D'après les hypothèses posées dans la section 2, il est facile de constater que

$$\text{Cov}(A,B) = 5\rho_1\sigma_y^2,$$

$$V(A) = V(B) = 5\sigma_y^2,$$

et, par conséquent,

$$\rho_1 = \frac{\text{Cov}(A,B)}{\sqrt{V(A)V(B)}}. \tag{2}$$

Il ne reste plus qu'à utiliser des valeurs estimées de $\text{Cov}(A,B)$, $V(A)$ et $V(B)$ pour obtenir la valeur estimée de ρ_1. On peut estimer ρ_2, ρ_3 et ρ_4 de la même façon en posant $A = \sum_{l=j+1}^{6} y_{m,l}$ et $B = \sum_{l=1}^{6-j} y_{m-j,l}$, $j = 2, 3, 4$. Cependant, lorsqu'il s'agit d'estimer ρ_5 de cette manière, des difficultés surgissent. En effet, lorsqu'on élimine tous les panels non communs pour les mois m et $m - 5$, il ne reste plus qu'un panel pour chaque mois et cela complique l'estimation de la variance pour ce qui a trait aux unités autoreprésentatives (UAR). Les UAR représentent les grandes agglomérations et chacune de ces unités fait l'objet d'un échantillonnage indépendant. Le problème ne se pose pas pour les unités non autoreprésentatives (UNAR), qui représentent les régions situées à l'extérieur des UAR, notamment les régions rurales et les petits centres urbains. Chaque unité primaire d'échantillonnage (UPÉ) qui sert d'échantillon répété pour l'estimation de la variance compte tous les groupes de renouvellement; par conséquent, même si on élimine 5 panels non communs, il y a toujours un panel de représenté dans l'UPÉ, de sorte qu'il est possible de calculer la variance. Dans les UAR toutefois, les groupes de renouvellement tiennent lieu d'échantillons répétés et s'il ne reste plus qu'un groupe, il n'y a donc qu'un échantillon répété par strate et on ne peut, par conséquent, calculer la variance de la manière habituelle. Nous avons donc calculé $\hat{\rho}_5$ par prédiction en utilisant le modèle de régression non linéaire $\rho = a + bt + ce^{-t}$, $t = 1, \ldots, 4$. Dans la sous-section 4.1, nous verrons une autre façon d'estimer ρ_5.

3.2 Estimation des coefficients de corrélation γ

On constate facilement que $\text{Cov}(A,B) = (5\rho_1 + \gamma_1)\sigma_y^2$ si $A = \sum_{l=1}^{6} y_{m,l}$ et $B = \sum_{l=1}^{6} y_{m-1,l}$. En règle générale,

$$\text{Cov}(A,B) = \{(6 - j)\rho_j + j\gamma_j\}\sigma_y^2,$$

où

$$A = \sum_{l=1}^{6} y_{m,l},$$

$$B = \sum_{l=1}^{6} y_{m-j,l}, \quad j = 1, \ldots, 4.$$

On peut alors calculer la valeur estimée de γ_j à l'aide de l'équation suivante:

$$\gamma_j = \frac{1}{j}\left[6\frac{\text{Cov}(A,B)}{\sqrt{V(A)V(B)}} - (6-j)\rho_j\right], \tag{3}$$

en ayant recours à des valeurs estimées des éléments du membre de droite. On peut aussi estimer directement les coefficients de corrélation γ, y compris γ_5, en se servant de l'équation

$$\gamma_j = \frac{\text{Cov}(A_j,B_j)}{\sqrt{V(A_j)V(B_j)}}, \tag{4}$$

où $A_j = \sum_{l=1}^{j} y_{m,l}$ et $B_j = \sum_{l=7-j}^{6} y_{m-j,l}$, $j = 2, \ldots, 5$. Dans la section 4, nous comparons les deux méthodes au moyen de données empiriques.

L'équation (4) permet de calculer d'autres coefficients γ (γ_j, $j = 6, \ldots, 10$); dans ce cas,

$$A_j = \sum_{l=j-5}^{6} y_{m,l},$$

$$B_j = \sum_{l=1}^{12-j} y_{m-j,l}.$$

Il n'existe pas de méthode simple pour estimer directement ou indirectement γ_{11} . Les valeurs estimées $\hat{\gamma}_5$ et $\hat{\gamma}_{11}$ ont toutes deux été prédites au moyen d'un modèle linéaire logarithmique, $\gamma = \exp(a + bt)$, $t = 1, \ldots, 4, 6, \ldots, 10$.

3.3 Estimation des coefficients de corrélation τ

On peut estimer τ de la même manière que ρ en remplaçant simplement $y_{m,l}$ par $x_{m,l}$. Soit $A = \sum_{l=j+1}^{6} x_{m,l}$ et $B = \sum_{l=1}^{6-j} y_{m-j,l}$, $j = 0, 1, \ldots, 4$. Nous avons alors

$$\text{Cov}(A,B) = (6-j)\tau_j\sigma_x\sigma_y,$$

$$V(A) = (6-j)\sigma_x^2,$$

$$V(B) = (6-j)\sigma_y^2,$$

ce qui donne

$$\tau_j = \frac{\text{Cov}(A,B)}{\sqrt{V(A)V(B)}}, \quad j = 0, 1, \ldots, 4. \tag{5}$$

L'équation ci-dessus permet d'estimer tous les τ sauf τ_5, qui est prédit au moyen d'un modèle linéaire logarithmique, $\tau = \exp(a + bt)$, $t = 1, \ldots, 4$.

4. RÉSULTATS ET ANALYSE

En nous servant des méthodes décrites dans la section précédente, nous avons établi les valeurs estimées de ρ et γ à partir des données de l'EPA de 1980-1981 et de 1985-1987 pour 5 caractéristiques: actif (IN LF), occupé (EMP), occupé en agriculture (EMP AG), occupé dans d'autres secteurs (EMP NON-AG), en chômage (UNEMP). Avec les données de 1980-1981, nous avons estimé les corrélations de panel pour trois provinces seulement: la Nouvelle-Écosse (N.-É.), l'Ontario (ONT) et la Colombie-Britannique (C.-B.). Toutefois, avec des données plus récentes (mars 1985 – février 1987), nous avons établi des estimations pour toutes les provinces. Nous avons aussi ajouté quatre autres caractéristiques, soit les personnes occupées et les personnes en chômage pour deux groupes d'âge, 15-24 et 25+ (EMP 15-24, EMP 25+, UNEMP 15-24, UNEMP 25+). En ce qui concerne l'estimation de τ, nous nous sommes servis uniquement des données de 1985-1987 pour les quatre dernières caractéristiques et pour les provinces de N.-É., d'ONT et d'Alberta (ALT).

Nous ne présentons et ne commentons ici qu'une partie des résultats. Le lecteur trouvera tous les résultats dans Lee (1989b).

4.1 Valeurs estimées des coefficients de corrélation ρ

Le tableau 2 donne les valeurs estimées des coefficients de corrélation ρ. Bien que les données de 1985-1987 aient permis de calculer des valeurs estimées de ρ pour les cinq caractéristiques initiales (IN LF, EMP, EMP AG, EMP NON-AG, UNEMP) et pour toutes les provinces, les deux séries d'estimations qui figurent dans le tableau 2 à des fins de comparaison ne portent que sur trois provinces (N.-É., ONT et C.-B.). Ce tableau contient aussi les estimations de ρ pour les quatre autres caractéristiques (EMP 15-24, EMP 25+, UNEMP 15-24, UNEMP 25+) pour la N.-É. et l'Ontario.

Tableau 2

Valeurs estimées du coefficient de corrélation ρ (données de 1980-1981 et de 1985-1987)

Prov.	Caractéristique	Données de 1980-1981					Données de 1985-1987				
		$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	$\hat{\rho}_5$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	$\hat{\rho}_5$
N.-É.	IN LF	0.862	0.797	0.744	0.679	0.622	0.845	0.769	0.730	0.696	0.670
	EMP	0.866	0.783	0.714	0.651	0.590	0.863	0.768	0.713	0.686	0.660
	EMP AG	0.913	0.837	0.756	0.678	0.598	0.912	0.867	0.825	0.802	0.773
	EMP NON-AG	0.865	0.774	0.710	0.649	0.594	0.873	0.779	0.724	0.697	0.670
	UNEMP	0.590	0.455	0.333	0.243	0.145	0.703	0.546	0.426	0.415	0.375
	EMP 15-24						0.773	0.632	0.556	0.495	0.446
	EMP 25+						0.878	0.800	0.754	0.729	0.705
	UNEMP 15-24						0.618	0.454	0.364	0.300	0.246
	UNEMP 25+						0.695	0.554	0.443	0.440	0.406
ONT	IN LF	0.843	0.782	0.717	0.674	0.622	0.846	0.781	0.732	0.681	0.635
	EMP	0.852	0.779	0.709	0.664	0.611	0.853	0.771	0.706	0.648	0.592
	EMP AG	0.955	0.926	0.901	0.861	0.827	0.962	0.948	0.944	0.937	0.934
	EMP NON-AG	0.861	0.791	0.724	0.678	0.625	0.866	0.795	0.746	0.701	0.660
	UNEMP	0.580	0.445	0.334	0.286	0.222	0.579	0.436	0.328	0.291	0.238
	EMP 15-24						0.747	0.605	0.500	0.429	0.356
	EMP 25+						0.888	0.824	0.777	0.732	0.691
	UNEMP 15-24						0.468	0.339	0.257	0.219	0.178
	UNEMP 25+						0.622	0.468	0.365	0.313	0.256
C.-B.	IN LF	0.849	0.767	0.705	0.665	0.622	0.817	0.753	0.701	0.647	0.597
	EMP	0.835	0.755	0.695	0.651	0.607	0.851	0.770	0.711	0.651	0.597
	EMP AG	0.896	0.809	0.733	0.656	0.582	0.938	0.886	0.847	0.828	0.805
	EMP NON-AG	0.855	0.769	0.715	0.661	0.616	0.857	0.784	0.730	0.679	0.632
	UNEMP	0.516	0.407	0.334	0.320	0.294	0.634	0.524	0.459	0.363	0.290

Comme prévu, les valeurs de ρ sont généralement élevées puisqu'elles mesurent la corrélation entre des estimations de panels communs. Elles sont le plus élevées pour EMP AG et le moins élevées pour UNEMP. C'est comme si la valeur de ρ indiquait le degré de mobilité d'un sous-groupe particulier de la population active. Par exemple, la valeur élevée de ρ pour la caractéristique EMP AG dénote une faible mobilité de la population active en agriculture, tandis que la faible valeur de ρ pour UNEMP indique une grande mobilité des personnes en chômage. La différence de mobilité de la population active entre les deux groupes d'âge étudiés est également très visible. Les plus jeunes (15-24) sont plus mobiles que leurs aînés (25+).

Les données du tableau 2 font ressortir clairement la tendance décroissante des valeurs de ρ. Cette tendance est très bien décrite par un modèle de régression non linéaire, $\rho_t = a + bt + ce^{-t}$. Les valeurs de R^2 (corrélation multiple) se rapprochent de 1 (> 0.98). Les valeurs prédites de ρ_5 semblent donc très satisfaisantes. Par ailleurs, Lee (1989a et 1989b) détermine $\hat{\rho}_5$ en extrapolant $\hat{\rho}_3$ et $\hat{\rho}_4$. La différence entre les valeurs prédites et les valeurs extrapolées est toutefois très mince, étant inférieure à 0.01 pour toutes les caractéristiques sauf UNEMP, UNEMP 15-24 et UNEMP 25+, où elle n'excède pas 0.03.

4.2 Valeurs estimées des coefficients de corrélation γ

Comme nous l'avons mentionné dans la sous-section 3.2, il y a deux façons d'estimer γ_2, γ_3 et γ_4, soit par l'équation (3) ou l'équation (4). Nous désignerons la première comme la Méthode 1 et la seconde comme la Méthode 2. Seule la Méthode 1 peut servir à estimer γ_1, tandis que γ_5 ne peut être estimé directement que par la Méthode 2. Dans le tableau 3, nous comparons les deux méthodes à l'aide de données empiriques. Les valeurs de $\hat{\gamma}_5$ indiquées pour la Méthode 1 ont été prédites au moyen d'un modèle linéaire logarithmique. Le tableau montre que les deux méthodes ont donné des résultats assez différents. On voit très bien que les coefficients calculés à l'aide de la Méthode 2 suivent une tendance à la hausse qui est contraire à notre intuition, tandis que les coefficients calculés par la Méthode 1 sont plus acceptables. En outre, si nous comparons les valeurs estimées du tableau 3 à $\hat{\gamma}_1$, qui ne peut être calculée que par la Méthode 1, cette dernière semble produire des résultats plus acceptables que la Méthode 2. Nous avons donc opté pour la Méthode 1. Néanmoins, dans des circonstances normales, les deux méthodes devraient être équivalentes et produire des résultats comparables. Il semble, dans ce cas-ci, que les données réelles ne soient pas parfaitement conformes aux hypothèses que nous avons posées pour définir les équations.

Tableau 3

Comparaison de valeurs estimées de γ_2, γ_3, γ_4 et γ_5 calculées à l'aide de méthodes différentes (Ontario, 1980-1981)

Caractéristique	Méthode	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$
IN LF	1	0.141	0.128	0.133	0.135
	2	0.107	0.105	0.116	0.120
EMP	1	0.136	0.142	0.142	0.147
	2	0.100	0.115	0.126	0.133
EMP AG	1	0.483	0.474	0.486	0.451
	2	0.321	0.370	0.407	0.448
EMP NON-AG	1	0.150	0.147	0.157	0.163
	2	0.117	0.134	0.145	0.149
UNEMP	1	0.074	0.076	0.063	0.080
	2	0.043	0.056	0.046	0.043

Nota: Les méthodes 1 et 2 correspondent aux équations (3) et (4) respectivement (section 3).

Tableau 4A

Valeurs estimées du coefficient de corrélation γ
(données de 1980-1981)

Prov.	Caractéristique	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$	$\hat{\gamma}_8$	$\hat{\gamma}_9$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{11}$
N.-É.	IN LF	0.288	0.263	0.265	0.250	0.236	0.233	0.211	0.199	0.193	0.167	0.164
	EMP	0.262	0.219	0.228	0.226	0.219	0.239	0.210	0.200	0.188	0.161	0.172
	EMP AG	0.351	0.308	0.283	0.237	0.205	0.190	0.141	0.113	0.063	0.021	0.007
	EMP NON-AG	0.238	0.187	0.189	0.180	0.164	0.151	0.123	0.121	0.136	0.091	0.086
	UNEMP	0.106	0.176	0.091	0.097	0.091	0.076	0.066	0.063	0.066	0.032	0.031
ONT	IN LF	0.161	0.141	0.128	0.133	0.135	0.136	0.125	0.127	0.124	0.122	0.117
	EMP	0.164	0.136	0.142	0.142	0.147	0.149	0.148	0.150	0.153	0.141	0.146
	EMP AG	0.477	0.483	0.474	0.486	0.451	0.474	0.459	0.429	0.394	0.323	0.368
	EMP NON-AG	0.184	0.150	0.147	0.157	0.163	0.167	0.166	0.169	0.174	0.156	0.165
	UNEMP	0.141	0.074	0.076	0.063	0.080	0.051	0.045	0.060	0.077	0.136	0.074
C.-B.	IN LF	0.177	0.137	0.117	0.119	0.119	0.112	0.101	0.112	0.094	0.066	0.070
	EMP	0.211	0.146	0.133	0.107	0.101	0.083	0.050	0.068	0.058	−0.033	−0.015
	EMP AG	0.380	0.311	0.301	0.272	0.241	0.216	0.198	0.170	0.122	0.078	0.071
	EMP NON-AG	0.207	0.166	0.161	0.129	0.108	0.093	0.069	0.038	0.023	−0.004	−0.020
	UNEMP	0.126	0.125	0.114	0.103	0.091	0.076	0.062	0.092	0.032	0.040	0.031

Tableau 4B

Valeurs estimées du coefficient de corrélation γ
(données de 1985-1987)

Prov.	Caractéristique	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$	$\hat{\gamma}_8$	$\hat{\gamma}_9$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{11}$
N.-É.	IN LF	0.250	0.238	0.247	0.230	0.216	0.204	0.181	0.196	0.189	0.162	0.160
	EMP	0.170	0.183	0.205	0.196	0.185	0.157	0.158	0.194	0.198	0.219	0.198
	EMP AG	0.326	0.296	0.246	0.245	0.265	0.267	0.234	0.217	0.259	0.269	0.231
	EMP NON-AG	0.146	0.168	0.199	0.201	0.178	0.153	0.152	0.189	0.199	0.216	0.201
	UNEMP	0.233	0.267	0.241	0.211	0.206	0.168	0.171	0.176	0.157	0.187	0.147
	EMP 15-24	0.107	0.127	0.140	0.133	0.112	0.105	0.099	0.107	0.090	0.074	0.082
	EMP 25+	0.088	0.075	0.117	0.108	0.100	0.099	0.090	0.103	0.099	0.137	0.118
	UNEMP 15-24	0.051	0.080	0.042	0.024	0.054	0.061	0.079	0.081	0.058	0.011	0.049
	UNEMP 25+	0.155	0.129	0.177	0.171	0.148	0.159	0.158	0.127	0.102	0.134	0.124
ONT	IN LF	0.162	0.138	0.141	0.134	0.132	0.135	0.127	0.116	0.111	0.103	0.101
	EMP	0.114	0.122	0.121	0.122	0.117	0.124	0.119	0.108	0.110	0.112	0.111
	EMP AG	0.508	0.518	0.553	0.561	0.571	0.569	0.582	0.617	0.668	0.650	0.672
	EMP NON-AG	0.133	0.140	0.132	0.140	0.157	0.156	0.168	0.182	0.204	0.205	0.210
	UNEMP	0.030	0.047	0.055	0.047	0.043	0.048	0.039	0.030	0.039	0.048	0.041
	EMP 15-24	0.012	−0.006	0.018	0.031	0.017	0.023	0.011	0.011	0.016	0.044	0.029
	EMP 25+	0.354	0.358	0.349	0.343	0.319	0.312	0.298	0.285	0.276	0.240	0.246
	UNEMP 15-24	0.068	0.039	0.038	0.058	0.033	0.026	0.008	0.018	0.011	−0.002	−0.006
	UNEMP 25+	0.052	0.054	0.033	0.017	0.034	0.033	0.026	0.018	0.021	0.044	0.022
C.-B.	IN LF	0.103	0.095	0.113	0.103	0.090	0.090	0.091	0.083	0.078	0.030	0.055
	EMP	0.125	0.100	0.112	0.111	0.116	0.135	0.123	0.121	0.118	0.095	0.114
	EMP AG	0.394	0.443	0.426	0.401	0.396	0.400	0.401	0.381	0.347	0.334	0.345
	EMP NON-AG	0.080	0.067	0.076	0.072	0.091	0.109	0.111	0.118	0.112	0.106	0.124
	UNEMP	0.096	0.086	0.084	0.080	0.083	0.097	0.068	0.074	0.068	0.083	0.071

Les valeurs estimées des coefficients de corrélation γ figurent dans les tableaux 4A et 4B. Comme on pouvait s'y attendre, les valeurs de γ sont beaucoup moins élevées que celles de ρ mais elles reflètent, comme ρ, des différences de mobilité entre certains sous-groupes de la population active.

La tendance générale des $\hat{\gamma}$ est assez bizarre, surtout en ce qui concerne les estimations établies à l'aide des données de 1985-1987. Pour environ 25% des cas qui forment la tableau 4B - un cas équivaut à une ligne du tableau, on note une tendance à la hausse des valeurs estimées. La droite de régression linéaire logarithmique présente alors une pente positive mais assez faible. En outre, dans la plupart de ces cas, la valeur de R^2 est peu élevée, ce qui indique la faible précision de l'ajustement obtenu à l'aide du modèle linéaire logarithmique. Cela ne signifie pas pour autant qu'il existe d'autres modèles qui pourraient nous permettre d'obtenir un meilleur ajustement. Tout ce que nous pouvons déduire de ces observations, c'est qu'aucune tendance claire ne ressort de ce tableau. Dans environ la moitié des cas pour lesquels on note une tendance à la baisse, la valeur de R^2 est supérieure à 0.5.

En ce qui concerne les estimations de $\hat{\gamma}$ établies à l'aide des données de 1980-1981, la situation est tout à fait différente. On note un seul cas où les valeurs estimées suivent une tendance à la hausse; de plus, la valeur de R^2 est supérieure à 0.5 pour la plupart des cas. De fait, pour ce qui a trait à la tendance, les estimations relatives à la N.-É. et à la C.-B. semblent plus conformes que celles pour l'ONT.

4.3 Valeurs estimées des coefficients de corrélation τ

Le tableau 5 renferme les valeurs estimées de τ, établies à l'aide des données de 1985-1987, pour toutes les combinaisons possibles de EMP 15-24 (désignée par x_1), EMP 25+ (x_2), UNEMP 15-24 (x_3) et UNEMP 25+ (x_4). La corrélation entre x_1 et x_2 est essentiellement positive, tout comme celle entre x_3 et x_4. Pour toutes les autres combinaisons, elle est essentiellement négative. En ce qui a trait au degré de corrélation, seuls les coefficients relatifs à (x_1, x_3) et à (x_2, x_4) sont sensiblement différents de zéro, les autres étant proches de zéro. Ces observations semblent confirmer ce que nous savons déjà du transfert de membres de la population active du même groupe d'âge entre le sous-groupe des personnes occupées et celui des personnes en chômage. Lorsque le nombre des personnes occupées augmente, celui des personnes en chômage diminue et vice-versa. La tendance est évidemment à la hausse dans ces cas.

Tableau 5

Valeurs estimées du coefficient de corrélation τ
x_1: EMP 15-24, x_2: EMP 25+, x_3: UNEMP 15-24, x_4: UNEMP 25+,
(données de 1985-1987)

Province	Caractéristique	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_3$	$\hat{\tau}_4$	$\hat{\tau}_5$
N.-É.	(x_1,x_2)	0.150	0.140	0.148	0.181	0.187	0.196
	(x_1,x_3)	−0.440	−0.275	−0.187	−0.135	−0.039	0.126
	(x_1,x_4)	−0.036	−0.040	−0.043	−0.015	0.024	0.022
	(x_2,x_3)	−0.029	−0.037	−0.078	−0.049	−0.016	−0.038
	(x_2,x_4)	−0.437	−0.374	−0.276	−0.182	−0.231	−0.094
	(x_3,x_4)	0.136	0.127	0.094	0.055	0.049	0.020
ONT	(x_1,x_2)	0.092	0.070	0.055	0.040	0.028	0.010
	(x_1,x_3)	−0.420	−0.267	−0.205	−0.161	−0.145	−0.010
	(x_1,x_4)	−0.065	−0.056	−0.053	−0.036	−0.028	−0.019
	(x_2,x_3)	−0.061	−0.054	−0.054	−0.042	−0.089	−0.074
	(x_2,x_4)	−0.392	−0.303	−0.230	−0.187	−0.181	−0.077
	(x_3,x_4)	0.058	0.043	0.022	0.013	0.022	0.001

L'ajustement s'est fait à l'aide d'un modèle linéaire logarithmique, qui a servi à prédire les valeurs de τ_5. Il semble satisfaisant sauf pour ce qui a trait à la corrélation entre x_2 et x_3, pour laquelle la valeur de R^2 est très faible dans les deux provinces (N.-É. et ONT).

4.4 Conclusions

L'estimation de coefficients de corrélation à l'aide de données d'enquête complexe est un problème délicat, non à cause de la difficulté des formules – en fait, les formules utilisées dans cette étude sont élémentaires – mais plutôt à cause des nombreuses contraintes auxquelles est soumise l'application de ces formules. Si nous n'avions pas posé les hypothèses de la section 3, il aurait été impossible d'estimer les corrélations de panel à l'aide du programme dont nous disposions. Par ailleurs, ces hypothèses devraient être compatibles avec les données réelles auxquelles sont appliquées les formules. Dans cette analyse, il semble que les hypothèses posées ne cadrent pas tout à fait avec les données réelles, vu la divergence des résultats obtenus à l'aide des équations (3) et (4) (tableau 3). Néanmoins, nous ne jugeons pas que les estimations sont aberrantes.

Les résultats de cette étude ont servi de façon concluante à comparer divers estimateurs composites pour l'EPA (Kumar et Lee 1983). Récemment, Binder et Dick (1990) ont proposé une méthode d'analyse des modèles saisonniers ARMMI qui tient compte des erreurs d'enquête. Ils ont appliqué cette méthode aux données de l'EPA en se servant des corrélations de panel estimées. Toutefois, si les résultats obtenus à l'aide de ces corrélations dépendent largement de la précision des estimations, ils doivent être interprétés avec prudence.

REMERCIEMENTS

L'auteur tient à remercier les arbitres anonymes ainsi que le rédacteur en chef, M.P. Singh, et la rédactrice adjointe L. Mach pour leurs précieux commentaires. Il remercie également S. Kumar et Y. Bélanger, de Statistique Canada, qui ont contribué, par leurs remarques constructives, à améliorer la version antérieure de cet article.

BIBLIOGRAPHIE

CHOUDHRY, G.H., et LEE, H. (1987). Estimation de la variance pour l'enquête sur la population active du Canada. *Techniques d'enquête*, 13, 157-172.

BINDER, D.A., et DICK, J.P. (1990). Méthode pour l'analyse des modèles ARMMI. *Techniques d'enquête*, ce numéro.

KISH, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.

KUMAR, S., et LEE, H. (1983). Évaluation de l'application d'estimateurs composites à l'enquête sur la population active du Canada. *Techniques d'enquête*, 9, 196-221.

LEE, H. (1989a). Variance estimation methodology and general purpose variance estimation system for the Labour Force Survey. Methodology Working Paper Series, SSMD-89-022 E, Statistique Canada.

LEE, H. (1989b). Estimation of panel correlations for the Canadian Labour Force Survey. Methodology Working Paper Series, SSMD-89-023 E, Statistique Canada.

PLATEK, R., et SINGH, M.P. (1976). *Méthodologie de l'enquête sur la population active du Canada.* N° 71-526 au catalogue, Statistique Canada.

SINGH, M.P., DREW, J.D., GAMBINO, J., et MAYDA, F. (1990). *Méthodologie de l'enquête sur la population active du Canada.* N° 71-526 au catalogue, Statistique Canada.