

**Une méthode de réconciliation des données d'enquêtes pour évaluer la dynamique spatiale de l'emploi au Canada, 1987-2008 : Application à l'Enquête sur la population active (EPA)**

**A Method for Adjusting Survey Data to Estimate the Regional Dynamics of Employment, Canada 1987-2008: Application to the Labour Force Survey (LFS)**

Jean Dubé

Volume 87, Number 1, March 2011

URI: <https://id.erudit.org/iderudit/1006337ar>

DOI: <https://doi.org/10.7202/1006337ar>

[See table of contents](#)

Publisher(s)

HEC Montréal

ISSN

0001-771X (print)

1710-3991 (digital)

[Explore this journal](#)

Cite this article

Dubé, J. (2011). Une méthode de réconciliation des données d'enquêtes pour évaluer la dynamique spatiale de l'emploi au Canada, 1987-2008 : Application à l'Enquête sur la population active (EPA). *L'Actualité économique*, 87(1), 9–29. <https://doi.org/10.7202/1006337ar>

Article abstract

This paper presents an application of an estimation method for calculating coherent spatial employment distributions for Canada over time. The cross-entropy minimization method allows for data reconciliation when disaggregation causes divergence in total distributions. Using Statistics Canada Labour Force Survey data by economic region and industry class, spatial employment distribution (rectangular) matrices were obtained over the entire time period (1987-2008). The paper proposes a procedure for reconciling survey data in cases where confidentiality and rounding rules are applied in order to obtain a coherent spatial-temporal data set even where the original data are incomplete.

## UNE MÉTHODE DE RÉCONCILIATION DES DONNÉES D'ENQUÊTES POUR ÉVALUER LA DYNAMIQUE SPATIALE DE L'EMPLOI AU CANADA, 1987-2008 : APPLICATION À L'ENQUÊTE SUR LA POPULATION ACTIVE (EPA)\*

Jean DUBÉ

*Département Sociétés, Territoires et Développement  
Université du Québec à Rimouski (UQAR)  
jean\_dube@uqar.qc.ca*

**RÉSUMÉ** – Cet article présente une application d'une méthode d'estimation de la répartition spatiale et temporelle de l'emploi régional au Canada. La méthode de minimisation de l'entropie croisée permet de réconcilier des données régionales qui sont, *a priori*, divergentes des totaux agrégés. À partir des données d'emploi provenant de l'Enquête sur la population active en fonction des régions économiques et des secteurs productifs, des matrices (rectangulaires) de répartitions spatiales de l'emploi sont obtenues pour l'ensemble des années disponibles (1987-2008). L'article montre comment il est possible de réconcilier les données d'enquêtes faisant état de règles d'arrondis et de règles de divulgation dans le but d'obtenir une source de données spatio-temporelle permettant d'utiliser, au meilleur des connaissances, des données incomplètes.

**ABSTRACT** – *A Method for Adjusting Survey Data to Estimate the Regional Dynamics of Employment, Canada 1987-2008: Application to the Labour Force Survey (LFS).* This paper presents an application of an estimation method for calculating coherent spatial employment distributions for Canada over time. The cross-entropy minimization method allows for data reconciliation when disaggregation causes divergence in total distributions. Using Statistics Canada Labour Force Survey data by economic region and industry class, spatial employment distribution (rectangular) matrices were obtained over the entire time period (1987-2008). The paper proposes a procedure for reconciling survey data in cases

---

\* Cette recherche a été subventionnée par le Fonds québécois de recherche sur la société et la culture (FQRSC).

L'auteur tient à remercier monsieur Pierre Desgagnés du ministère des Transports du Québec (MTQ) avec lequel il a développé une collaboration sur un sujet semblable et qui lui a donné l'idée de développer cet article à saveur appliquée. L'auteur tient également à remercier Mario Polèse, André Lemelin et un évaluateur anonyme pour leurs commentaires qui ont permis d'améliorer la qualité de l'article.

where confidentiality and rounding rules are applied in order to obtain a coherent spatial-temporal data set even where the original data are incomplete.

## INTRODUCTION

La science régionale et l'économie géographique s'intéressent à la compréhension des conséquences économiques de l'espace en étudiant la répartition spatiale et la localisation des activités économiques. Cette branche s'attarde, en outre, à l'étude des comportements à un niveau de désagrégation géographique important. Une source de données intéressante pour étudier la situation régionale de l'emploi à l'échelle des régions économiques canadiennes entre les recensements repose sur l'Enquête sur la population active (EPA) de Statistique Canada. Or, l'EPA comprend certaines règles de divulgation/confidentialité et/ou d'arrondissements qui peuvent rendre les données disponibles incomplètes pour un certain niveau de désagrégation. Les règles de divulgation ont notamment pour effet d'introduire une distorsion entre les totaux selon que les entités géographiques sont agrégées ou désagrégées. Bien que de peu d'importance au niveau géographique agrégé, ces restrictions compliquent l'étude des phénomènes de croissance de l'emploi sectoriel à l'échelle des petites régions par le fait que peu de données et d'informations existent pour ce niveau de désagrégation.

En plus de ce problème s'ajoute celui de la variation constante de l'échantillonnage dans les régions économiques de faible taille qui a pour effet d'introduire une source de variation supplémentaire dans les données. La caractérisation de l'échantillonnage a pour conséquence de créer des séries temporelles plus volatiles lorsque le niveau de désagrégation est important. Une façon simple de retirer certains mouvements irréguliers des séries de données chronologiques consiste à utiliser une méthode de lissage telle que, par exemple, un modèle à moyenne mobile intégrée autorégressif (ARIMA<sup>1</sup>). Par contre, l'application de cette technique sur des séries de données indépendantes a pour effet de ne pas respecter les contraintes naturelles posées par les données d'enquête, à savoir le respect des totaux, provinciaux et nationaux.

Certaines méthodes ont été développées afin de pallier le manque de données ou encore l'incompatibilité de celles-ci à l'échelle désagrégée lorsque comparées aux totaux agrégés. Une de ces approches, la méthode de minimisation de l'entropie croisée (MinXEnt), permet de régler simultanément les deux lacunes à partir d'un problème d'optimisation reposant sur la construction d'une matrice de répartition des données. La méthode MinXEnt permet d'obtenir une nouvelle distribution spatiale d'une variable d'intérêt de sorte qu'elle assure un minimum de divergence avec la distribution d'origine tout en respectant un certain nombre de contraintes. Celles-ci représentent, notamment, les règles de divulgation et d'arrondissements dans la publication des données, ainsi que la somme des variables, à l'échelle désagrégée, qui doit être égale à la somme à l'échelle agrégée.

---

1. La méthode X-12-ARIMA est d'ailleurs retenue par Statistique Canada et le U.S. Census Bureau afin de produire des données désaisonnalisées (Hood et Feldpausch, 2003).

L'approche MinXEnt est retenue et appliquée pour la répartition de l'emploi au Canada par région économique et par secteur productif entre 1987 et 2008. Dans ce contexte, la méthode permet d'obtenir une nouvelle base de données, à la fois spatiale et temporelle, qui respecte les contraintes d'agrégation ainsi que les contraintes de publication des données. Cette nouvelle base de données permet, sans être jugée optimale, d'étudier certains phénomènes économiques spatialisés à un niveau de désagrégation important, autant au niveau géographique qu'au niveau sectoriel. La méthode peut être appliquée à d'autres sources de données en appliquant des contraintes similaires, de nouvelles contraintes jugées pertinentes ou encore d'autres contraintes plus générales afin de dériver une nouvelle source de données comportant moins de limites.

L'article est divisé en quatre parties. La première présente l'Enquête sur la population active : les caractéristiques d'échantillonnage, les règles de divulgation et de suppression des données ainsi que les données utilisées pour obtenir une nouvelle répartition désagrégée cohérente de l'emploi au Canada. La deuxième partie présente formellement la méthode de la minimisation de l'entropie croisée dans le but d'obtenir une nouvelle estimation de la répartition sectorielle de l'emploi régional au Canada. La troisième partie présente les résultats d'estimation et met en relief les différences entre les données obtenues et celles de l'Enquête sur la population active. Finalement, une brève conclusion clôture l'article.

## 1. ENQUÊTE SUR LA POPULATION ACTIVE (EPA)

### 1.1 *Caractérisation de l'EPA*

L'Enquête sur la population active (EPA) est menée chaque mois auprès d'environ 54 000 ménages canadiens dans le but de dresser le portrait détaillé du marché du travail (Statistique Canada, 2009). Cette enquête permet de calculer certains indicateurs fondamentaux tels que le taux de participation au marché du travail, le taux d'emploi et le taux de chômage, par province, mais également par région économique et par région métropolitaine de recensement. Elle permet également de connaître la distribution spatiale de l'emploi en fonction des classes industrielles et constitue la principale source de données permettant de suivre l'évolution de l'emploi au Canada entre les recensements.

L'EPA consiste en une entrevue pour un échantillon de la population en âge de travailler (15 ans et plus). Un ménage ciblé est suivi pendant 6 mois et chaque mois, un sixième de l'échantillon est renouvelé (chevauchement) afin d'assurer un taux de réponse intéressant tout en contrôlant pour les coûts de l'enquête. Un échantillon de ménages est sélectionné à partir d'une méthode d'échantillonnage à deux degrés (Statistique Canada, 2008). En première étape, un échantillon de régions géographiques, désignées par le terme unité primaire d'échantillonnage (UPE), est sélectionné. Par la suite, pour chacune des UPE, un échantillon de

ménages est sélectionné. Comme toute enquête, l'EPA demeure sensible aux erreurs d'échantillonnages<sup>2</sup> et aux erreurs non reliées à l'échantillonnage<sup>3</sup>.

Étant donné le renouvellement du sixième de l'échantillon à chaque mois et la méthode d'échantillonnage en deux étapes, les données peuvent montrer une volatilité temporelle considérable pour de petites régions économiques et fragiliser l'analyse chronologique des données à courte fréquence, à plus forte raison si l'analyse régionale est effectuée au niveau des secteurs productifs. Pour cette raison, les analyses à l'échelle des régions économiques sont souvent effectuées avec des moyennes annuelles ou encore en utilisant des séries désaisonnalisées. Malgré qu'elles permettent de diminuer l'impact de la volatilité des séries, ces approches n'éliminent pas la totalité des problèmes liés à la volatilité et encore moins ceux liés à l'incompatibilité des données agrégées et désagrégées.

L'EPA est astreinte à certaines règles de publication des données selon la province. Premièrement, les données publiées sont arrondies à la centaine la plus près. Deuxièmement, certaines valeurs sont supprimées de la base de données lorsque jugées trop peu nombreuses ou susceptibles de révéler certaines informations et contrevenir à la règle de divulgation. La publication des valeurs dépend ainsi d'un certain seuil de divulgation qui varie en fonction des provinces (tableau 1). Ainsi, les données publiées sont susceptibles de montrer certains comportements hasardeux tels que la présence de « *faux zéros* ». Pour certains secteurs productifs, le nombre d'emplois est nul pour une année donnée pour ensuite se situer à un nombre supérieur du seuil de divulgation l'année suivante et redevenir nul par la suite.

TABLEAU 1

SEUIL MINIMAL DE DIVULGATION DES DONNÉES D'EMPLOI DE L'EPA PAR PROVINCE  
(EN DEÇÀ DE CES SEUILS, LES DONNÉES NE SONT PAS PUBLIÉES.)

<b>Province</b>	<b>Emplois</b>	<b>Province</b>	<b>Emplois</b>
Terre-Neuve	500	Ontario	1 500
Île-du-Prince-Édouard	200	Manitoba	500
Nouvelle-Écosse	500	Saskatchewan	500
Nouveau-Brunswick	500	Alberta	1 500
Québec	1 500	Colombie-Britannique	1 500

SOURCE : Statistique Canada (2009).

2. L'estimation est différente de ce que l'on peut obtenir lorsque l'ensemble de la population est interviewé.

3. Les personnes interviewées peuvent ne pas donner la réponse exacte ou peuvent encore approximer la réponse. Il s'agit de toutes les erreurs humaines reliées au sondage.

Ces quelques effets non désirables ne sont pas sans importance lorsque le sujet d'intérêt est la croissance de l'emploi par secteur productif au niveau des régions économiques. La troncature des données, de par un nombre considérable de valeurs *zéro*, implique une mauvaise évaluation des taux de croissance et, par conséquent, risque d'introduire des éléments perturbateurs supplémentaires dans un modèle statistique. De plus, ces règles introduisent des divergences entre le nombre total d'emplois selon le niveau de désagrégation, géographique et industriel, retenu. Dans ces cas, il devient souhaitable d'obtenir une meilleure estimation du nombre d'emplois qui respecte les contraintes de répartitions agrégées de l'emploi, à la fois par secteur productif et par province.

### 1.2 Données de l'EPA, 1987-2008

Les données d'emploi de l'EPA peuvent être obtenues par région économique, par secteur productif ou par région économique en fonction des secteurs productifs à partir de 1987<sup>4</sup>. L'EPA permet de diviser le Canada (à l'exception des territoires) en 69 régions (économiques) et en 16 secteurs productifs, inspirés du système de classification des industries d'Amérique du Nord (SCIAN<sup>5</sup> – tableau 2). Le total de l'emploi est comparable au total provincial selon que celui-ci est ventilé par secteur productif ou par région économique (tableau 3). Quelques écarts subsistent, mais ceux-ci sont très faibles et sont majoritairement la résultante d'arrondissements (centaine près), bien qu'il existe quelques exceptions où les écarts sont de 200 à 300 emplois.

Par contre, de par les contraintes de publication des données, la somme totale des emplois par secteur productif et par région économique donne un portrait différent du nombre d'emplois total agrégé par province et pour le pays (tableau 4). Les écarts entre le nombre total d'emplois par province par rapport au nombre total d'emplois pour le Canada peuvent être causés par les règles d'arrondissements. Par contre, les écarts plus grands notés au niveau des secteurs productifs par région économique ne peuvent provenir uniquement des règles d'arrondissements. Ceux-ci sont particulièrement marqués pour les provinces du Québec, de l'Alberta et de la Colombie-Britannique, où la règle de divulgation minimale est fixée à 1 500 emplois par région économique et secteur productif.

Il devient donc nécessaire de reconstituer les données d'emploi au niveau des secteurs productifs par région économique si on s'intéresse à la situation régionale de l'emploi par branches productives. La méthode de minimisation de l'entropie croisée (MinXEnt) permet d'estimer une distribution spatiale de l'emploi qui soit cohérente avec les informations émanant de l'EPA. Elle s'appuie sur des fondements théoriques rigoureux tout en faisant preuve de transparence et en recourant au principe d'objectivité scientifique (Kapur et Kesavan, 1992).

---

4. Par contre, les données régionales ne sont pas disponibles pour les régions économiques de la Colombie-Britannique avant 1995.

5. <http://www.statcan.gc.ca/subjects-sujets/standard-norme/naics-scian/2002/naics-scian021-fra.htm>.

TABLEAU 2

LISTE DES SECTEURS INDUSTRIELS SELON L'EPA  
ET CORRESPONDANCE POSSIBLE AVEC LE SCIAN

Secteurs industriels de l'EPA	Secteurs industriels du SCIAN	Code SCIAN
• Agriculture	• Agriculture, foresterie, pêche et chasse	11
• Foresterie, pêche, mines et extraction de pétrole et de gaz	• Extraction minière et extraction de pétrole et de gaz	21
• Services publics	• Services publics	22
• Construction	• Construction	23
• Fabrication	• Fabrication	31-33
• Commerce	• Commerce de gros	41
	• Commerce de détail	44-45
• Transport et entreposage	• Transport et entreposage	48-49
• Finance, assurances, immobilier et location	• Finances et assurances	52
	• Services immobiliers et services de location et de location à bail	53
• Services professionnels, scientifiques et techniques	• Services professionnels, scientifiques et techniques	54
• Services aux entreprises, services relatifs aux bâtiments et autres services de soutien	• Gestion de sociétés et d'entreprises	55
	• Services administratifs, services de soutien, services de gestion des déchets et services d'assainissement	56
• Services d'enseignement	• Services d'enseignement	61
• Soins de santé et assistance sociale	• Soins de santé et assistance sociale	62
• Information culture et loisirs	• Industrie de l'information et industrie culturelle	51
	• Arts, spectacles et loisirs	71
• Hébergement et services de restauration	• Hébergement et services de restauration	72
• Autres services	• Autres services, sauf les administrations publiques	81
• Administrations publiques	• Administrations publiques	91

TABLEAU 3

NOMBRE TOTAL D'EMPLOI (EN MILLIERS) PAR PROVINCE SELON LA MÉTHODE D'AGRÉGATION

	Terre-Neuve			Île-du-Prince-Édouard			Nouvelle-Écosse			Nouveau-Brunswick		
	Par SP	Par RE	Total	Par SP	Par RE	Total	Par SP	Par RE	Total	Par SP	Par RE	Total
1987	190,2	190,4	190,3	53,3	53,4	53,4	359,2	358,9	359,0	280,0	280,3	280,3
1988	199,8	199,7	199,7	54,6	54,6	54,6	373,8	373,7	373,7	291,1	291,1	291,0
1989	206,4	206,4	206,3	55,0	55,0	55,0	381,5	381,6	381,6	296,8	296,8	296,8
1990	207,0	206,8	206,9	55,0	55,1	55,1	385,4	385,2	385,3	300,3	300,2	300,3
1991	204,6	204,7	204,7	53,4	53,4	53,4	380,6	380,7	380,6	295,2	295,3	295,2
1992	195,1	194,9	194,9	53,6	53,7	53,7	368,8	368,8	368,9	296,8	297,0	296,9
1993	193,8	193,7	193,8	54,5	54,6	54,6	366,4	366,4	366,5	299,8	299,9	299,9
1994	193,7	193,5	193,5	55,7	55,7	55,7	372,5	372,5	372,6	298,6	298,7	298,6
1995	194,5	194,4	194,4	57,2	57,2	57,2	375,8	376,0	375,9	307,4	307,4	307,5
1996	187,3	187,5	187,5	58,9	58,9	58,9	376,8	376,9	376,9	305,6	305,5	305,5
1997	188,3	188,3	188,3	58,6	58,8	58,8	381,9	381,9	382,0	309,1	309,1	309,1
1998	192,6	192,3	192,4	59,5	59,6	59,6	395,3	395,2	395,3	315,3	315,3	315,2



TABLEAU 3 (suite)

	Terre-Neuve			Île-du-Prince-Édouard			Nouvelle-Écosse			Nouveau-Brunswick		
	Par SP	Par RE	Total	Par SP	Par RE	Total	Par SP	Par RE	Total	Par SP	Par RE	Total
1999	200,9	201,0	201,0	60,3	60,2	60,2	404,1	404,0	404,0	325,4	325,4	325,5
2000	198,2	198,0	198,0	62,4	62,7	62,7	411,4	411,4	411,4	331,1	331,2	331,2
2001	203,8	203,7	203,8	63,6	63,6	63,6	415,2	415,3	415,2	330,0	330,1	330,1
2002	207,3	207,1	207,2	64,6	64,7	64,7	423,0	422,9	422,9	343,1	343,1	343,1
2003	212,1	212,3	212,3	66,0	66,1	66,1	431,4	431,1	431,2	343,1	343,1	343,1
2004	214,3	214,4	214,3	66,9	66,9	66,9	442,0	442,2	442,2	350,2	350,1	350,1
2005	213,9	214,0	214,1	68,2	68,2	68,2	443,1	443,2	443,1	350,5	350,5	350,5
2006	215,6	215,6	215,7	68,6	68,6	68,6	441,8	441,8	441,8	355,4	355,5	355,4
2007	217,2	217,1	217,1	69,5	69,3	69,3	447,6	447,5	447,6	362,7	362,9	362,8
2008	220,4	220,4	220,3	70,1	70,2	70,2	453,2	453,1	453,2	366,0	366,1	366,2

NOTE : SP = secteur productif; RE = région économique

TABLEAU 3 (suite)

	Québec			Ontario			Manitoba			Saskatchewan		
	Par SP	Par RE	Total	Par SP	Par RE	Total	Par SP	Par RE	Total	Par SP	Par RE	Total
1987	3 022,1	3 022,1	3 022,1	4 895,4	4 895,5	4 895,6	505,3	505,3	505,2	461,9	461,8	461,9
1988	3 081,5	3 081,5	3 081,4	5 083,0	5 083,1	5 083,1	506,3	506,3	506,3	463,0	462,8	462,8
1989	3 128,8	3 128,8	3 128,8	5 198,8	5 199,1	5 199,0	512,6	512,7	512,7	456,3	456,4	456,4
1990	3 140,2	3 140,4	3 140,3	5 194,2	5 194,1	5 194,1	513,7	513,8	513,8	454,1	454,2	454,2
1991	3 084,5	3 084,4	3 084,4	5 017,2	5 017,1	5 017,1	506,7	506,7	506,8	453,3	453,4	453,3
1992	3 038,4	3 038,5	3 038,6	4 932,7	4 932,7	4 932,9	500,0	500,0	499,9	448,2	448,1	448,0
1993	3 030,9	3 030,9	3 030,9	4 938,2	4 938,1	4 938,0	503,6	503,7	503,8	448,4	448,6	448,5
1994	3 094,6	3 095,0	3 094,8	5 013,6	5 013,8	5 013,6	507,7	507,6	507,7	454,6	454,4	454,5
1995	3 135,3	3 135,4	3 135,3	5 100,1	5 099,9	5 100,0	516,6	516,4	516,5	458,2	458,0	458,0
1996	3 129,8	3 129,6	3 129,8	5 167,2	5 167,1	5 167,1	517,3	517,4	517,4	456,9	456,9	456,8
1997	3 172,8	3 173,0	3 172,8	5 291,2	5 291,5	5 291,4	525,5	525,6	525,6	466,2	466,2	466,2
1998	3 257,2	3 257,4	3 257,5	5 453,2	5 453,3	5 453,3	534,2	534,2	534,2	470,5	470,5	470,5

TABLEAU 3 (suite)

	Québec			Ontario			Manitoba			Saskatchewan		
	Par SP	Par RE	Total	Par SP	Par RE	Total	Par SP	Par RE	Total	Par SP	Par RE	Total
1999	3 328,1	3 327,8	3 328,1	5 636,7	5 636,8	5 636,7	541,4	541,4	541,4	471,5	471,5	471,6
2000	3 402,9	3 402,8	3 402,8	5 817,0	5 817,1	5 817,1	552,3	552,2	552,3	473,7	473,5	473,5
2001	3 440,1	3 440,3	3 440,2	5 926,0	5 926,3	5 926,2	554,3	554,4	554,3	460,4	460,3	460,3
2002	3 569,9	3 570,1	3 569,9	6 031,0	6 031,4	6 031,4	567,2	567,1	567,2	468,6	468,3	468,3
2003	3 628,9	3 629,0	3 628,8	6 213,1	6 213,1	6 213,2	570,2	570,1	570,3	476,0	476,1	476,1
2004	3 680,3	3 680,4	3 680,5	6 316,6	6 316,5	6 316,5	576,6	576,6	576,6	479,8	479,6	479,7
2005	3 717,2	3 717,3	3 717,3	6 397,7	6 397,9	6 397,7	580,3	580,4	580,3	483,4	483,4	483,5
2006	3 765,4	3 765,4	3 765,4	6 492,7	6 492,8	6 492,7	587,0	587,0	587,0	491,5	491,6	491,6
2007	3 851,6	3 851,6	3 851,7	6 593,9	6 593,8	6 593,8	596,6	596,5	596,5	502,0	501,7	501,8
2008	3 881,7	3 881,8	3 881,7	6 687,4	6 687,2	6 687,3	606,7	606,6	606,7	513,0	512,7	512,7

NOTE : SP = secteur productif; RE = région économique

TABLEAU 3 (suite)

	Alberta			Colombie-Britannique		
	Par SP	Par RE	Total	Par SP	Par RE	Total
1987	1 187,5	1 187,7	1 187,7	1 377,6	–	1 377,7
1988	1 222,3	1 222,3	1 222,3	1 434,7	–	1 434,6
1989	1 251,3	1 251,3	1 251,3	1 508,3	–	1 508,3
1990	1 276,9	1 276,9	1 276,8	1 559,6	–	1 559,6
1991	1 284,5	1 284,4	1 284,4	1 577,6	–	1 577,5
1992	1 280,0	1 280,0	1 280,0	1 617,3	–	1 617,2
1993	1 288,9	1 288,8	1 288,7	1 668,0	–	1 668,0
1994	1 324,8	1 324,6	1 324,5	1 743,1	–	1 743,2
1995	1 364,9	1 364,9	1 364,9	1 785,8	1 785,7	1 785,6
1996	1 405,0	1 405,0	1 405,1	1 816,4	1 816,4	1 816,4
1997	1 451,3	1 451,5	1 451,4	1 860,4	1 860,5	1 860,5
1998	1 509,9	1 509,9	1 509,9	1 858,4	1 858,4	1 858,4
1999	1 543,7	1 543,9	1 544,0	1 894,3	1 894,4	1 894,4
2000	1 583,8	1 583,8	1 584,0	1 931,2	1 931,3	1 931,3
2001	1 630,9	1 630,9	1 630,9	1 921,5	1 921,4	1 921,6
2002	1 670,8	1 670,7	1 670,8	1 965,1	1 964,9	1 965,0
2003	1 716,7	1 716,6	1 716,7	2 014,5	2 014,6	2 014,7
2004	1 757,4	1 757,4	1 757,5	2 062,7	2 062,7	2 062,7
2005	1 784,3	1 784,3	1 784,4	2 130,5	2 130,6	2 130,5
2006	1 870,6	1 870,7	1 870,7	2 195,5	2 195,4	2 195,5
2007	1 959,4	1 959,5	1 959,4	2 266,4	2 266,3	2 266,3
2008	2 013,2	2 013,2	2 013,3	2 314,2	2 314,2	2 314,3

NOTE : SP = secteur productif; RE = région économique

TABLEAU 4

NOMBRE TOTAL D'EMPLOIS (EN MILLIERS) PAR SECTEUR PRODUCTIF  
SELON LA DÉSAGRÉGATION GÉOGRAPHIQUE RETENUE

	Régions économiques (1)	Provinces (2)	Canada (3)	Écart : (3) - (1)	Écart : (3) - (2)
1987	12 262,4	12 332,5	12 333,1	70,7	0,6
1988	12 649,5	12 710,1	12 709,6	60,1	-0,5
1989	12 939,9	12 995,8	12 996,1	56,2	0,3
1990	13 032,8	13 086,4	13 086,3	53,5	-0,1
1991	12 799,6	12 857,6	12 857,5	57,9	-0,1
1992	12 678,1	12 730,9	12 730,8	52,7	-0,1
1993	12 743,8	12 792,5	12 792,7	48,9	0,2
1994	13 003,5	13 058,9	13 058,8	55,3	-0,1
1995	13 228,1	13 295,8	13 295,5	67,4	-0,3
1996	13 366,0	13 421,2	13 421,3	55,3	0,1
1997	13 640,3	13 705,3	13 705,9	65,6	0,6
1998	13 985,7	14 046,1	14 046,1	60,4	0,0
1999	14 339,5	14 406,4	14 406,6	67,1	0,2
2000	14 709,3	14 764,0	14 764,1	54,8	0,1
2001	14 884,1	14 945,8	14 946,2	62,1	0,4
2002	15 237,4	15 310,6	15 310,5	73,1	-0,1
2003	15 611,3	15 672,0	15 672,2	60,9	0,2
2004	15 889,0	15 946,8	15 946,9	57,9	0,1
2005	16 113,5	16 169,1	16 169,5	56,0	0,4
2006	16 428,3	16 484,1	16 484,3	56,0	0,2
2007	16 815,8	16 866,9	16 866,3	50,5	-0,6
2008	17 069,6	17 125,9	17 125,9	56,3	0,0

## 2. MÉTHODE

### 2.1 Minimisation de l'entropie croisée<sup>6</sup>

Le problème de minimisation de l'entropie croisée (MinXEnt) permet de généraliser la méthode d'ajustement de matrice (Macgill, 1977), bien connue pour équilibrer les matrices d'entrées-sorties. Plus concrètement, la méthode MinXEnt consiste à ajuster une matrice *a priori* – en l'occurrence la matrice de répartition sectorielle de l'emploi régional – de façon à respecter une information qui est imposée comme contrainte et qui, dans le cas présent, est donnée par la répartition agrégée de l'emploi sectoriel dans les provinces canadiennes.

L'application de la méthode, visant à minimiser la divergence entre la matrice de répartition *a priori* et celle obtenue *a posteriori*, exige la définition d'une mesure de différence. Celle-ci est donnée par la mesure de l'entropie croisée de Kullback et Leibler (1951) qui consiste à minimiser la divergence entre les éléments d'origine (ou *a priori*<sup>7</sup> –  $q_{ij}$ ) et la nouvelle distribution (ou *a posteriori*<sup>8</sup> –  $p_{ij}$ ) (équation 1).

$$\text{Min } I(p|q) = \sum_i \sum_j p_{ij} \log \left( \frac{p_{ij}}{q_{ij}} \right). \quad (1)$$

Dans le cas actuel, les éléments généraux de la distribution *a priori* représentent le nombre d'emplois publié (EPA) pour le secteur productif  $j$  de la région économique  $i$  alors que les éléments généraux de la distribution *a posteriori* représentent le nombre d'emplois estimé pour le secteur productif  $j$  de la région économique  $i$ . La forme de la fonction objectif assure que la nouvelle distribution obtenue (le résultat) est unique puisqu'elle est convexe (Kullback, 1959; Kapur et Kesavan, 1992).

La fonction objectif est minimisée en tenant compte d'un certain nombre de contraintes. Premièrement, puisque la fonction n'est pas définie pour des valeurs nulles *a priori*, il est nécessaire d'effectuer une transformation additive<sup>9</sup> pour chacun des éléments si certains sont nuls. Deuxièmement, une contrainte naturelle est posée afin d'assurer la non-négativité des éléments (équation 2). Le système tient également compte des contraintes provenant des règles de publication des données de l'EPA dont : *i*) le respect des seuils de divulgation pour chacune des régions économiques,  $s_i$ , lorsque la valeur initiale est nulle ( $q_{ij} = 0$ ) (équation 3); *ii*) le respect de la règle d'arrondissement des données initiales – soit à la centaine

6. La méthode est présentée de façon rapide dans le présent article. Cependant, le lecteur intéressé par plus de détails peut consulter Theil (1967, 1971), Dubé (2003), Dubé et Dupéré (2004) et Dubé et Lemelin (2005).

7. Ces éléments ne respectent pas nécessairement les contraintes naturelles et celles posées par les règles imposées à divulgation des données.

8. La distribution *a posteriori* respecte les contraintes imposées.

9. Il s'agit d'ajouter un terme constant (valeur positive), mais très petit, à chacune des cellules.

près (0,1) (équation 4); *iii*) le total marginal du nombre d'emplois par secteur productif,  $\bar{q}_{.j}$  (équation 5); et *iv*) le total marginal par région économique,  $\bar{q}_{.i}$  (équation 6).

$$p_{ij} \geq 0, \quad (2)$$

$$(p_{ij} - q_{ij}) \leq s_i \quad \forall q_{ij} = 0, \quad (3)$$

$$|q_{ij} - p_{ij}| \leq 0,1 \quad \forall q_{ij} \neq 0, \quad (4)$$

$$\sum_{i=1}^I p_{ij} = \bar{q}_{.j} \quad (5)$$

et

$$\sum_{j=1}^J p_{ij} = \bar{q}_{.i} . \quad (6)$$

Les contraintes fixent l'ordre de grandeur de la répartition totale alors que l'ajustement porte sur la structure de la matrice. L'optimum répond au principe scientifique de Laplace qui stipule que lorsqu'aucune information supplémentaire ne permet de croire que des contraintes additionnelles peuvent être posées au problème d'optimisation, alors la nouvelle distribution obtenue est celle qui a le plus de chance de se réaliser.

L'application de la méthode MinXEnt repose sur la construction de matrices rectangulaires de données, une par année, qui exposent la répartition spatiale de l'emploi dans les régions économiques en fonction des secteurs productifs. Cette matrice de répartition, de dimension  $I \times J$ , provient des données désagrégées de l'EPA.

### 3. EXAMEN DES RÉSULTATS

Puisque les totaux marginaux peuvent varier entre le total des lignes (régions économiques) et des colonnes (secteurs productifs) étant donné les règles d'arrondissements, il est nécessaire d'effectuer certains compromis sur l'imposition des contraintes des totaux marginaux. Les règles de publications font en sorte qu'aucune des deux contraintes n'est totalement compatible avec le total provincial de l'emploi et qu'aucun des totaux provinciaux n'est compatible avec le total national. Ainsi, à moins d'ajuster proportionnellement les totaux marginaux en se basant sur le total national<sup>10</sup>, il est nécessaire d'effectuer un choix quant à la contrainte serrante à imposer au système sur les totaux marginaux. Puisque les données de répartition de l'emploi par secteur productif par province sont dispo-

---

10. Cet ajustement est plutôt arbitraire et c'est en partie pourquoi les contraintes sont fixées à partir d'une distribution marginale dont les chiffres sont supposés « plus fiables ».

nibles pour l'ensemble des années<sup>11</sup>, elles sont identifiées comme contraintes serrantes des totaux marginaux. Les totaux marginaux des régions économiques peuvent varier à la centaine près<sup>12</sup>. De plus, puisque la fonction objectif n'est pas déterminée pour des valeurs nulles *a priori*, une valeur de départ unitaire<sup>13</sup> est ajoutée à tous les éléments généraux  $q_{ij}$ .

Les nouvelles matrices de répartition de l'emploi, en fonction des secteurs productifs par région économique, permettent de prendre compte des règles d'arrondissements et de divulgation tout en respectant les contraintes imposées. Elles permettent d'assurer une cohérence des résultats tout en minimisant la « distance » ou « l'écart » entre la distribution de l'emploi publiée par l'EPA et la nouvelle distribution estimée. En tout, 22 matrices de distribution spatiale de l'emploi, une par année, sont obtenues<sup>14</sup>. Les résultats sont ensuite compilés dans une nouvelle base de données permettant de dresser le portrait de la répartition spatio-temporelle de l'emploi au Canada entre 1987 et 2008 selon la définition de l'EPA.

Les résultats sont imposants en termes de volume (chaque matrice est composée de 69 lignes et 16 colonnes), ce qui rend difficile la présentation détaillée des résultats<sup>15</sup>. Sans présenter formellement l'ensemble des tableaux de données, certains constats généraux émanent. Premièrement, étant donné les contraintes imposées au système, il s'avère que les nouvelles distributions permettent d'obtenir des résultats, par année et par province, qui sont cohérents avec les données provinciales de la répartition de l'emploi par secteur productif selon l'EPA. Deuxièmement, il est maintenant possible d'obtenir une estimation du nombre d'emplois pour chacun des secteurs productifs dans chacune des régions économiques, ce qui n'est pas le cas avec les données originales étant donné les règles de divulgation. Troisièmement, la méthode permet d'obtenir une estimation plus détaillée (et plus probable) du nombre d'emplois. Puisque les totaux marginaux sont supérieurs aux totaux de la répartition désagrégée de l'emploi compte tenu des restrictions imposées à leur divulgation, les estimations sont, règle générale, plus élevées que les données publiées. À titre d'exemple, le nombre d'emplois relié au secteur de l'agriculture (secteur 11) en 1987 pour la région économique de Toronto est de 7 500 selon l'EPA, alors que l'application de la méthode MinXEnt suggère que ce nombre est de 7 525.

---

11. Contrairement aux données par région économique qui ne sont pas disponibles pour la Colombie-Britannique entre 1987 et 1994.

12. Par contre, certains écarts sont plus élevés pour quelques années (1997, 2000 et 2007). Dans ces cas, la règle d'arrondissement est augmentée à deux centaines (0,2) (1997 et 2007) et à trois centaines (0,3) (2000). Cette règle est également relâchée pour les années antérieures à 1995 puisque les données ne sont pas disponibles pour les régions économiques de la Colombie-Britannique et que les valeurs initiales d'emploi utilisées sont celles de 1995, soit la première observation disponible.

13. La valeur unitaire indique que les valeurs de départ, pour les cellules où aucune information n'est disponible, sont fixées à un millier d'emplois.

14. Le problème d'optimisation est effectué à partir du logiciel GAMS. Les personnes intéressées à consulter le programme peuvent communiquer directement avec l'auteur.

15. Les résultats détaillés sont disponibles auprès de l'auteur sur demande.



De plus, la méthode fait en sorte qu'il est maintenant possible, à partir des données de répartition de 1995<sup>16</sup>, d'obtenir une répartition du nombre d'emplois pour les régions économiques de la Colombie-Britannique entre 1987 et 1994 malgré que ces données ne soient pas disponibles *a priori* (graphique 1). Bien qu'il soit plutôt difficile d'établir la précision et la fiabilité de ces estimations, elles demeurent néanmoins intéressantes dans l'optique où certaines informations partielles sont disponibles pour établir les contraintes agrégées.

L'ensemble des corrections apportées aux estimations publiées du nombre d'emplois procure des séries chronologiques continues. Par exemple, le secteur économique des « services d'utilité publique (22) » pour les régions économiques de la Côte-Nord de la Nouvelle-Écosse (graphique 2) et de Parklands Nord (graphique 3) font maintenant état d'un nombre d'emploi non nul à chaque année. Le même constat s'applique au secteur économique « extraction minière et extraction de pétrole et de gaz (21) » pour la région économique de Montréal (graphique 4). D'une part, ces estimations permettent de calculer des taux de croissance pour chacune des années, calculs qui étaient jusque-là impossibles pour plusieurs secteurs productifs de plusieurs régions économiques. D'autre part, les estimations introduisent néanmoins des marges d'erreurs possiblement importantes dans les calculs des taux de croissance, bien que moins grandes que dans les séries originales.

Au final, l'application de la méthode MinXEnt permet d'obtenir une nouvelle base de données de type *panel* (ou *pseudo-panel*) qui respecte certaines contraintes provinciales et qui élimine les problèmes de ruptures et de discontinuités des séries chronologiques liées aux règles d'arrondissements ainsi que ceux de divulgation. À défaut de pouvoir utiliser les données « réelles »<sup>17</sup> de l'emploi régional sectoriel, il s'agit d'une estimation intéressante, complète et moins mauvaise. Cette approche peut facilement être extrapolée à plusieurs autres applications, ce qui en fait un outil intéressant pour les économistes régionaux. Par contre, le problème de la volatilité des taux de croissance calculés demeure entier étant donné deux réalités : *i*) la structure de l'échantillon de l'EPA et *ii*) la résolution strictement spatiale du problème d'optimisation, par opposition à une approche spatio-temporelle.

Une grande partie de la volatilité des séries provient de la méthode d'échantillonnage à deux degrés. Celle-ci a une incidence marquée lorsque les régions

---

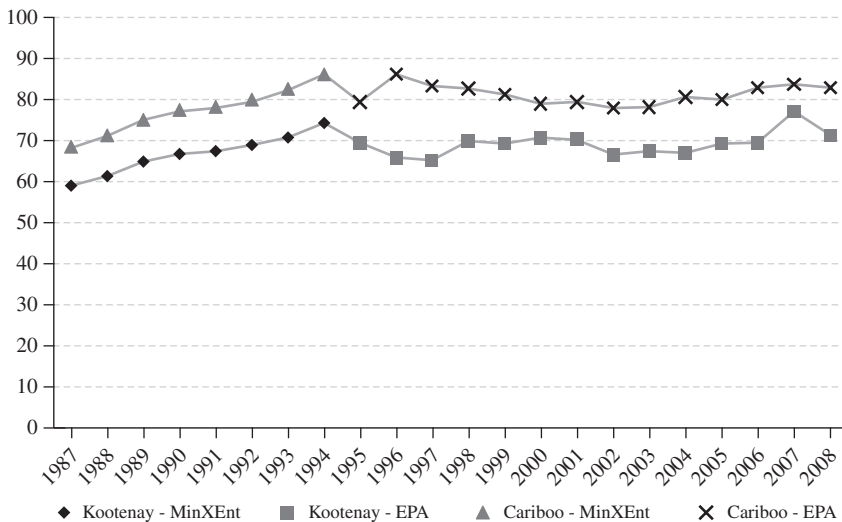
16. L'approche suppose que la distribution de l'emploi en 1995 reflète bien la distribution de l'emploi pour les années manquantes antérieures.

17. La donnée réelle d'emploi n'est pas disponible à moins d'effectuer un recensement de l'emploi. Malgré tout, les données de l'EPA ont l'avantage d'être rajustées tous les cinq ans afin que les estimations reflètent la réalité des recensements.

Une répartition de l'emploi par secteur productif serait fort probablement plus juste avec les données de l'Enquête sur la rémunération et les heures de travail (EERH) qui est basée sur des fichiers administratifs. Cependant, l'EERH a le désavantage d'exclure, *a priori*, les travailleurs autonomes, diminuant ainsi le nombre d'emplois.

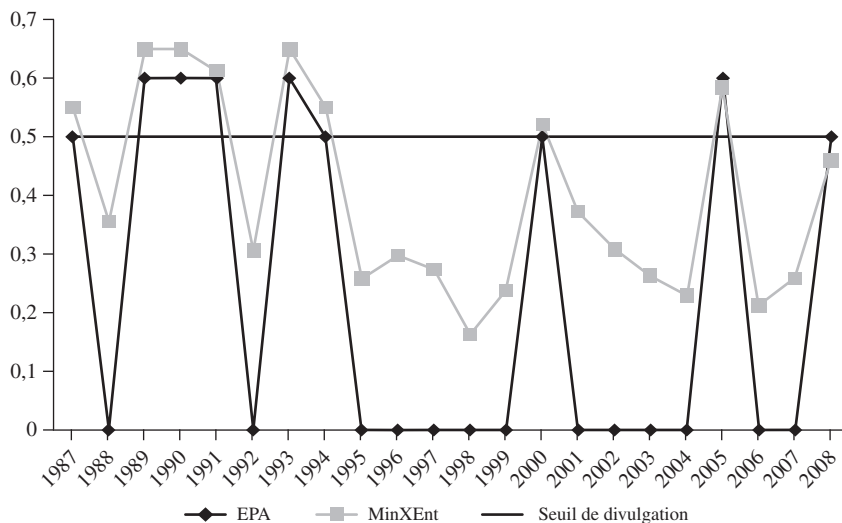
GRAPHIQUE 1

ÉVOLUTION DU NOMBRE TOTAL D'EMPLOIS (EN MILLIERS)  
 POUR DEUX RÉGIONS ÉCONOMIQUES DE LA COLOMBIE-BRITANNIQUE  
 (TOUS LES SECTEURS PRODUCTIFS), 1987-2008.



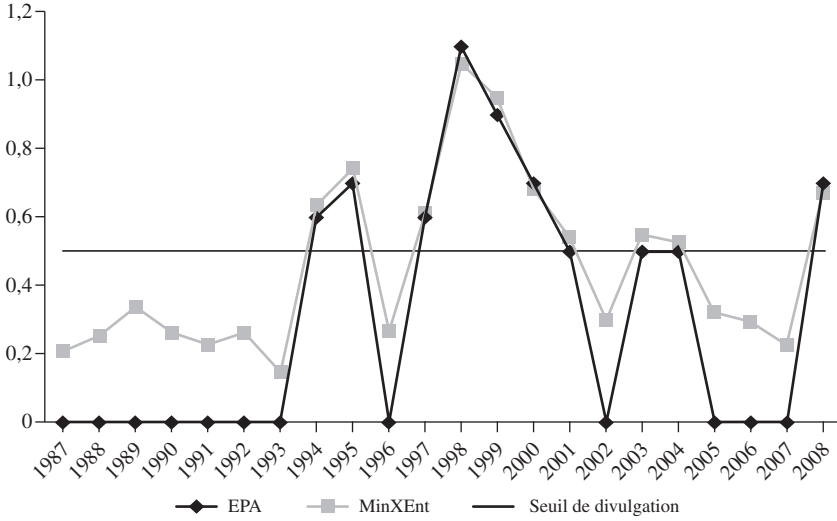
GRAPHIQUE 2

ÉVOLUTION DU NOMBRE D'EMPLOIS (EN MILLIERS) DU SECTEUR PRODUCTIF  
 « SERVICES D'UTILITÉ PUBLIQUE – 22 » POUR LA RÉGION ÉCONOMIQUE  
 DE LA CÔTE-NORD (NOUVELLE-ÉCOSSE), 1987-2008.



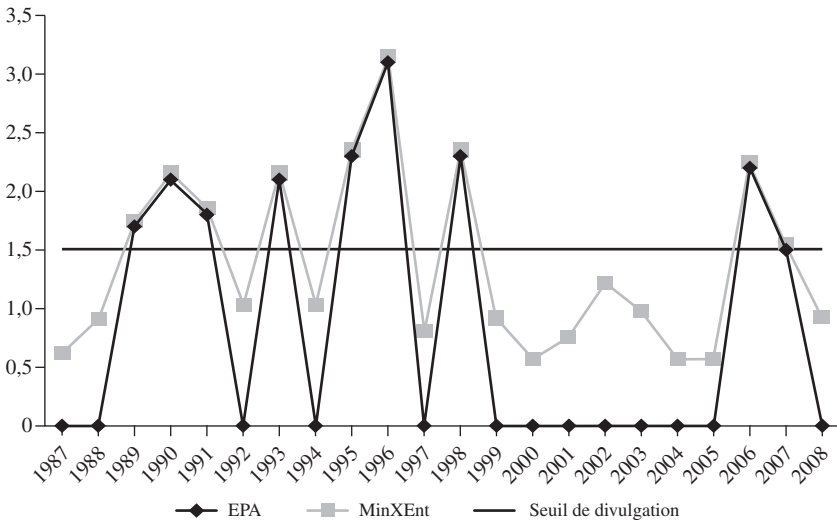
GRAPHIQUE 3

ÉVOLUTION DU NOMBRE D'EMPLOI (EN MILLIERS) DU SECTEUR PRODUCTIF « SERVICES D'UTILITÉ PUBLIQUE – 22 » POUR LA RÉGION ÉCONOMIQUE DE PARKLANDS NORD (MANITOBA), 1987-2008.



GRAPHIQUE 4

ÉVOLUTION DU NOMBRE D'EMPLOI (EN MILLIERS) DU SECTEUR PRODUCTIF « EXTRACTION MINIÈRE ET EXTRACTION DE PÉTROLE ET DE GAZ – 21 » POUR LA RÉGION ÉCONOMIQUE DE MONTRÉAL, 1987-2008.



économiques s'étendent sur de grandes surfaces dont la densité de la population est faible et dont la répartition de l'emploi par secteur productif n'est pas uniformément distribuée sur le territoire. Une autre partie de la volatilité peut provenir d'un manque de spécification des contraintes par rapport à de réelles valeurs nulles dans certains secteurs productifs de quelques régions économiques. Cette possibilité aurait ainsi pour effet de faussement répartir une partie de la distribution vers d'autres cellules, accentuant ainsi la volatilité temporelle des séries.

Finalement, l'application de la méthode spatiale dans un contexte spatio-temporel s'avère incomplète et peut ainsi générer une autre source de volatilité des séries temporelles. Une application plus globale, permettant de considérer à la fois les dimensions temps et espace, serait une approche intéressante dans l'optique où elle permettrait de traiter directement cette partie du problème de volatilité et possiblement une autre partie liée à la structure de l'échantillon de l'EPA. Par contre, cette éventualité s'avère nettement plus complexe puisque la forme de la matrice à estimer est à trois dimensions (régions économiques, secteurs productifs et temps). La solution optimale passe par le développement d'une approche spatio-temporelle intégrée et une possible redéfinition des totaux marginaux afin de régler à la fois les problèmes de divulgation spatiale de certaines données et le problème de variabilité temporelle causé par une composition fortement changeante de l'échantillon à petite échelle.

## CONCLUSION

Les données d'emploi provenant de l'Enquête sur la population active (EPA) permettent d'étudier l'évolution de l'emploi sectoriel au Canada pour un ensemble prédéterminé de régions géographiques, en l'occurrence les 69 régions économiques définies par Statistique Canada. Cependant, certaines faiblesses sont inhérentes à cette source de données. La première vient de toutes les données d'enquête dont l'échantillonnage change chaque mois : les séries chronologiques à faible fréquence (mensuelle, trimestrielle) issues de telles enquêtes sont nécessairement plus volatiles. La seconde vient du fait que le nombre d'emplois par secteur productif par région économique est différent du nombre total d'emplois par secteur productif par province. Des règles de divulgation et d'arrondissements font en sorte d'introduire des problèmes de ruptures et de discontinuités des séries chronologiques produisant des incohérences à des géographies plus fines. La troisième, issue de la seconde, vient d'une présence importante de valeurs *zéro*, fragilisant ainsi le calcul des taux de croissance de l'emploi pour certains secteurs productifs régionaux.

Une façon de contourner certaines faiblesses consiste à appliquer la méthode de minimisation de l'entropie croisée (MinXEnt) qui permet, à la lumière des données existantes, d'obtenir une nouvelle distribution spatiale de l'emploi par secteur productif et par région économique qui satisfait aux différentes contraintes fixées. Cette méthode permet d'obtenir, pour chacune des années, une distribution de l'emploi qui satisfait les contraintes de répartitions marginales de

l'emploi par secteur productif par province en plus d'éliminer le problème des données manquantes. L'application de la méthode MinXEnt a également l'avantage de fournir une estimation du nombre d'emplois pour certaines régions économiques où ces données ne sont pas disponibles pour certaines années (Colombie-Britannique). Au final, l'article démontre qu'il est possible, à partir des données de sondage irrécyclables au départ, d'obtenir une nouvelle distribution de l'emploi qui soit cohérente avec les estimations globales. En d'autres termes, la méthode a l'avantage de permettre un raccord des séries chronologiques des données régionales de la moins mauvaise manière possible.

Cependant, certaines questions demeurent en suspens. La méthode MinXEnt ne peut régler la totalité des problèmes liés à la volatilité des données de l'emploi par secteur productif et région économique provenant de l'EPA. Une application plus globale permettant de considérer à la fois les dimensions temps et espace serait une approche intéressante, mais qui reste à développer, puisqu'elle permettrait d'attaquer à la fois les problèmes de divulgation et d'arrondissements des données ainsi que la structure et la rotation du plan d'échantillonnage.

#### BIBLIOGRAPHIE

- DUBÉ, J. (2003), « Estimation des flux d'échanges interrégionaux par la méthode de minimisation de l'entropie croisée », Mémoire de maîtrise codirigé par A. Lemelin et B. Decaluwé, Québec : Université Laval, 106 p.
- DUBÉ, J. et M. DUPÉRÉ (2006), « Une méthode de répartition pour déduire les flux d'échanges économiques », *Bulletin économique du transport au Québec*, ministère des Transports du Québec, 37 : 24-28.
- DUBÉ, J. et M. DUPÉRÉ (2004), *Expérimentation de la méthode d'entropie croisée pour l'estimation d'échanges économiques à partir de flux de transport*, Collection *Études et recherches en transport*, ministère des Transports du Québec, 38 p.
- DUBÉ, J. et A. LEMELIN (2005), « Estimation expérimentale des flux d'échanges interrégionaux par la méthode de minimisation de l'entropie croisée », *Revue canadienne de sciences régionales*, 28(3) : 513-534.
- HOOD, C.C. et R.M. FELDPAUSCH (2003), *Quelques propriétés des diagnostics de désaisonnalisation du modèle X-12-ARIMA*, no. 11-522-XIF au catalogue, Statistique Canada, 12 p.
- KAPUR, J.N. et H.K. KESAVAN (1992), *Entropy Optimization Principles with Applications*, Academic Press, Inc., San Diego, CA, 405 p.
- KULLBACK, S. (1959), *Information Theory and Statistics*, John Wiley & Sons Inc., Canada, 395 p.
- KULLBACK, S. et R. LEIBLER (1951), « On Information and Sufficiency », *Annals of Mathematical Statistics*, 22 : 79-86.
- MACGILL, S.M. (1977), « Theoretical Properties of Biproportionnal Matrix Adjustment », *Environment and Planning A*, 9 : 687-701.

- STATISTIQUE CANADA (2009), *Produits et services de l'Enquête sur la population active*, no. 71-544-X au catalogue, Statistique Canada, 19 p.
- STATISTIQUE CANADA (2008), *Méthodologie de l'Enquête sur la population active au Canada*, no. 71-526-X au catalogue, Statistique Canada, 122 p.
- THEIL, H. (1967), *Economics and Information Theory*, vol. 7 dans la série « Studies in Mathematical and Managerial Economics », Rand McNally and Company, Chicago, 488 p.
- THEIL, H. (1971), *Principles of Econometrics*, A Wiley/Hamilton Publications, John Wiley & Sons, Inc., 736 p.