Techniques d'enquête, décembre 1985
Vol. 11, n° 2, pp. 223–227
Statistique Canada

Utilisation des comptes de l'Hydro dans le modèle d'estimation par régression de la population en Colombie-Britannique[1]

DONALD G. McRAE[2]

RÉSUMÉ

La précision des estimations régionales de la population qui sont établies à l'aide d'un modèle de régression repose foncièrement sur la capacité des indicateurs sélectionnés de traduire avec exactitude les changements démographiques. Par conséquent, il est important de connaître préalablement les caractéristiques des données administratives pouvant servir d'indicateurs de la population dans un modèle de régression. Ce document présente sommairement les avantages et les inconvénients de l'utilisation des comptes résidentiels de l'Hydro dans le modèle d'estimation par régression de la population de la Colombie-Britannique.

MOTS CLÉS: Estimations régionales de la population; méthode de régression; méthode de corrélation des différences; indicateurs de population; comptes de l'Hydro; prestataires d'allocations familiales.

1. INTRODUCTION

Le Bureau central de la statistique produit des estimations postcensitaires de la population pour diverses unités géographiques de la Colombie-Britannique (par ex. les municipalités, les districts sanitaires locaux, les divisions de recensement, les régions de la GRC, et autres). Les estimations courantes de la population sont produites pour ces zones infraprovinciales à l'aide d'une méthode de régression, la méthode de corrélation des différences (MCD). Une description détaillée de cette méthode a été présentée dans des documents antérieurs (Bureau central de la statistique, 1982, McRae 1985). Les données utilisées comme indicateurs de la population sont le nombre de prestataires d'allocations familiales (F), et(ou) le nombre de comptes résidentiels de l'Hydro (H). Les caractéristiques de cette deuxième source de données, les comptes résidentiels de l'Hydro, tels qu'ils sont utilisés dans le modèle de la C.-B., sont examinées dans le reste de cet exposé.

2. SOURCES DE DONNÉES

Les données des comptes résidentiels de l'Hydro à l'intérieur de la province proviennent de neuf administrations et sociétés différentes :

Organisation	% de l'ensemble des comptes de l'Hydro (1985)
(1) British Columbia Hydro	90.8
(2) West Kootenay Power and Light	4.7
(3) Princeton Light and Power Co.	0.2
(4) Ville de Kelowna	0.8
(5) Ville de Penticton	0.8
(6) District de Summerland	0.3
(7) Ville de New Westminster	1.7
(8) Ville de Grand Forks	0.1
(9) Ville de Nelson	0.6

[1] Exposé présenté à la rencontre du Comité fédéral-provincial de la démographie, les 28 et 29 novembre 1985.

[2] D.G. McRae, Bureau central de la statistique, ministère du développement de l'industrie et de la petite entreprise de la Colombie-Britannique, 2e étage, 1405, Douglas Street, Victoria, C.-B., Canada, V8W 3C1.
Les opinions exprimées dans ce document sont celles de l'auteur et ne représentent pas nécessairement le point de vue du gouvernement de la Colombie-Britannique.

Les principaux fournisseurs d'électricité sont la B.C. Hydro et la West Kootenay Power and Light. Les autres administrations achètent l'électricité aux deux grands fournisseurs et la revendent à leurs propres clients (normalement les résidents de la municipalité).

3. PRÉSENTATION DES DONNÉES

Sur les neuf sources de données des comptes résidentiels de l'Hydro, seulement celles qui proviennent de la B.C. Hydro sont ordinolingues, c'est-à-dire lisibles par machine. Les huit autres administrations fournissent les données totalisées par municipalité (urbaine), et un total pour les clients ruraux (non municipaux) s'il y a lieu. La date de référence pour toutes les données est la date de facturation du 31 mai. Dans la plupart des cas, les données peuvent être obtenues dans les 2 à 3 semaines qui suivent la date de facturation.

Les données fournies par la B.C. Hydro se présentent sous deux formes. La première indique le nombre de compteurs résidentiels en date du 31 mai par code de district capital. Un code de district capital (il en existe environ 248 dans la province) est une unité administrative utilisée par la B.C. Hydro, qui correspond à une municipalité là où des municipalités existent. Aux termes d'une entente, les deux principaux fournisseurs d'électricité dans la province paient à chaque municipalité un certain pourcentage des recettes annuelles perçues auprès des clients dans cette municipalité, ce qui tient lieu d'impôt foncier. Par conséquent, ces sociétés, comme la B.C. Hydro, établissent leur système de comptabilité en fonction des clients résidant dans les municipalités visées. En outre, la B.C. Hydro cherche à maintenir une correspondance étroite entre les limites des districts capitaux et celles des districts scolaires.

La deuxième forme de présentation indique pour chacun des compteurs résidentiels, qui sont au nombre d'un million et plus, le code postal de l'adresse de facturation. Ce deuxième fichier de données permet de convertir facilement les comptes de l'Hydro en unités géographiques autres que les municipalités et les districts scolaires, au moyen du code postal.

4. AVANTAGES DES DONNÉES DE L'HYDRO DANS UN MODÈLE DE RÉGRESSION

Des tests empiriques portant sur les deux sources de données différentes, les comptes de l'Hydro (H) et les comptes d'allocations familiales (F), ont été effectués par la production d'estimations de la population pour 1981, avec chacune prise séparément et avec les deux ensemble. Les coefficients de régression utilisés étaient fondés sur les périodes 1971/1976 et 1976/1981, et l'année de base était 1976. Les résultats ont été comparés aux données du recensement de 1981, et les erreurs absolues moyennes en pourcentage (EAMP) ont été calculées (voir tableaux 1 et 2).

Comme on peut voir au tableau 1, les estimations de la population fondées sur les données de l'Hydro produisent, en moyenne, des pourcentages d'erreur inférieurs à ceux des estimations fondées sur les comptes d'allocations familiales. Si on examine de plus près le tableau 1, on constate que l'amélioration de la précision des estimations touche en presque totalité les estimations relatives aux régions d'une population inférieure à 4,000 habitants. Cette observation est confirmée par le tableau 2 où l'on peut voir que, en termes statistiques, il existe une différence importante sur le plan de la précision entre les estimations fondées sur les comptes de l'Hydro et celles fondées sur les comptes d'allocations familiales pour les régions de petite taille (c.-à-d. moins de 4,000 habitants).

On peut évaluer l'effet marginal de l'adjonction d'un autre indicateur de population à la méthode de corrélation des différences en examinant la variation de la précision de l'estimation selon qu'on utilise ou non cet indicateur. D'après les tableaux 1 et 2, il semblerait que l'adjonction des données de l'Hydro améliore du point de vue statistique la précision de

Tableau 1

Comparaison des erreurs d'estimation entre les sources de données pour les municipalités de la Colombie-Britannique - 1981

Source de données	EAMP[3] Ensemble	n	Population ≥ 4000 EAMP	n	Population < 4000 EAMP	n
MCD/H/F	5.53	158	2.99	88	8.72	70
MCD/H	5.16	158	4.04	88	6.58	70
MCD/F	10.46	167	4.57	92	17.69	75

$$\text{EAMP} = \left[\sum_{i=1}^{n} \left| \frac{\hat{Y}_i - Y_i}{Y_i} \right| \right] \div n \times 100$$

où:

Y_i = population du recensement pour la région i
$\hat{Y}_i$ = population estimé pour la région i
n = nombre de régions estimées.

Tableau 2

Test pour déterminer les différences statistiquement significatives entre les erreurs absolues moyennes en pourcentage, selon certaines sources de données - 1981

Source de données	Intervalle de confiance de 95% pour la différence moyenne dans les erreurs absolues en pourcentage		
		Population	
	Ensemble	≥ 4000	< 4000
MCD/H/F - MCD/H	.37 ± .86	−1.05 ± .56[a]	2.14 ± 1.76[a]
MCD/H/F - MCD/F[b]	−4.86 ± 1.55[a]	−1.57 ± .85[a]	−9.00 ± 3.11[a]

[a] Différences statistiquement significatives au niveau de 5% calculées à l'aide d'un test-t bilatéral et d'échantillons appariés, en supposant une distribution normale des moyennes.

[b] Pour apparier les échantillons, sur 167 estimations possibles fondées sur les comptes d'allocations familiales, 158 seulement ont été utilisées. Le nombre des observations est le suivant: ensemble, 158; plus grand que ou égal à 4,000, 88; moins de 4,000, 70.

l'estimation à la fois dans les grandes et les petites régions. Les données des comptes d'allocations familiales, par contre, améliorent la précision pour les régions de grande taille, mais la réduit pour les régions de petite taille, sans produire un effet global statistiquement significatif.

5. INCONVÉNIENTS DES DONNÉES DE L'HYDRO DANS UN MODÈLE DE RÉGRESSION

Un des problèmes qui se posent lorsqu'on utilise les données de l'Hydro dans un modèle d'estimation de la population par régression concerne les logements vacants, ou plus précisément les écarts importants des taux de logements vacants entre l'année de base et l'année d'estimation. Cette lacune a été démontrée dans l'évaluation de 1981 portant sur les localités de la région Peace River-Liard en Colombie-Britannique. Par suite du projet de North-East Coal, les localités de la division de recensement Peace River-Liard en 1981 ont connu une

explosion de la construction lorsque les promoteurs ont construit des logements en prévision d'un afflux de population. Chaque logement, occupé ou non, ou même en construction, devait être muni d'un compteur qui était peut-être peu utilisé, mais qui n'en demeurait pas moins actif, et par le fait même compris dans les calculs. Par conséquent, la variation de la proportion de compteurs par rapport à la population entre 1976 et 1981 a été surévaluée, ce qui a entraîné une surestimation de la population de 1981 pour de nombreuses localités de Peace River-Liard.

Un autre inconvénient de l'utilisation des données de l'Hydro découle de la possibilité d'une conversion, dans le cas des immeubles d'habitation, d'un compteur unique en un réseau de compteurs multiples. Cette situation peut se présenter lorsqu'un ancien immeuble d'appartements, par exemple, qui possédait un seul compteur, est rénové ou transformé en logements à compteurs individuels. Ce changement entraînerait une surestimation de la population dans un modèle de régression, s'il devait survenir à un moment quelconque entre l'année de base et l'année d'estimation.

Enfin, certains problèmes apparaîtront si on utilise les données de l'Hydro pour des régions en mutation, c'est-à-dire où le rapport entre la population et les compteurs résidentiels se modifie. On peut citer comme exemple en Colombie-Britannique le lieu de villégiature de Whistler. Il y a quinze ans cette municipalité était constituée essentiellement de chalets d'hiver construits aux abords d'une pente de ski. Au cours de la dernière décennie, cependant, elle s'est transformée en un lieu résidentiel à longueur d'année. Par conséquent, le nombre de personnes par compteur de l'Hydro, qui à l'origine était très faible par rapport à la moyenne de la Colombie-Britannique, se rapproche maintenant de la norme. Comme pour le cas des logements vacants, l'utilisation des données de l'Hydro pour estimer la population dans un territoire de ce genre donnerait probablement lieu à des erreurs supérieures à la moyenne.

La solution aux trois problèmes mentionnés ci-haut est la suppression des comptes correspondant à une utilisation mensuelle ou bimensuelle faible qui par le fait même représentent des logements considérés vacants. Le Bureau central de la statistique étudie actuellement la possibilité d'appliquer cette mesure dans le cas des données provenant de la B.C. Hydro. Si c'est possible, nous espérons disposer de l'ensemble de données améliorées pour la comparaison avec les résultats du recensement de 1986. Pour l'instant, à titre de solution partielle, les données de l'Hydro pour les secteurs qui en 1981 affichaient un rapport faible ou élevé de personnes par compteur comparativement à la norme provinciale (c.-à-d. moins de 2 ou plus de 5) ne sont pas utilisées.

Une dernière lacune possible est le recours à des sources d'information externes et différentes. Dans le passé, cette situation n'a généralement pas posé de problème. Toutefois, il y a eu quelques rares cas qui ont suscité une remise en question de la qualité des données des compteurs recueillies sur le terrain. Il pourrait s'agir d'une modification des limites d'une municipalité non reflétée dans les données relatives aux compteurs ou de l'adjonction aux données d'un certain type de comptes non résidentiels (par exemple pour les lampadaires). Il importe donc de vérifier soigneusement les données.

6. CONCLUSIONS

Les avantages et les inconvénients de l'utilisation des données de l'Hydro dans le modèle d'estimation par régression de la population en Colombie-Britannique sont les suivants:

Avantages

a) Les données de l'Hydro, lorsqu'elles sont utilisées dans un modèle de régression, produisent une erreur absolue moyenne en pourcentage plus faible que les données sur les allocations familiales, pour les petites régions.

b) Les données sont obtenues de chaque fournisseur sous une forme déjà agrégée par municipalité. Le principal avantage est que les modifications des limites municipales, qui se produisent régulièrement, paraissent dans les données sans entraîner de travail supplémentaire pour le Bureau.
c) La majorité des données peuvent être obtenues sous forme ordinolingue et accompagnées du code postal. Cela permet de les convertir facilement en fonction de régions géographiques autres que les municipalités lorsqu'un triage est effectué par le Fichier principal de conversion des codes postaux du Bureau.
d) Les données peuvent être obtenues gratuitement de chacun des fournisseurs dans un délai relativement court (2 à 3 semaines).

Inconvénients

a) Les écarts dans les taux de logements vacants entre l'année de base et l'année d'estimation produisent une distorsion des estimations.
b) Les immeubles d'habitation qui passent d'un simple compteur à des compteurs multiples parfois entre l'année de base et l'année d'estimation produisent une surestimation.
c) Les régions en mutation, c'est-à-dire qui passent par exemple d'une population saisonnière à une population "stable", introduisent un biais dans les estimations.
d) Les données proviennent de sources externes et différentes, ce qui pourrait causer éventuellement des problèmes sur le plan de la qualité et de la comparabilité des données, et produire également une situation où l'ordre de priorité du programme des estimations de la population du Bureau serait subordonné aux besoins administratifs d'un organisme externe.

BIBLIOGRAPHIE

BUREAU CENTRAL DE LA STATISTIQUE (1982). British Columbia municipal population estimation methodology. Document non publié, Bureau central de la statistique, ministère du Développement de l'industrie et de la petite entreprise, gouvernement de la Colombie-Britannique.

McRAE, D. (1985). A regression approach to small area population estimation. Communication présentée au colloque international sur les statistiques régionales, Ottawa, Canada, 22-24 mai 1985.

McRAE, D. (1982). British Columbia small area estimation model - 1981 municipal and census division evaluation. Document non publié, Bureau central de la statistique, ministère du Développement de l'industrie et de la petite entreprise, gouvernement de la Colombie-Britannique.