Survey Methodology, December 1985
Vol. 11, No. 2, pp. 197–201
Statistics Canada

The Use of Hydro Accounts in the British Columbia Regression Based Population Estimation Model[1]

DONALD G. McRAE[2]

ABSTRACT

The accuracy of small area population estimates derived from a regression based model is heavily dependent on the ability of the indicator data selected to accurately reflect population change. Hence, prior knowledge as to the characteristics of the administrative data used as potential population indicators in a regression model is important. This report summarizes the strengths and weaknesses associated with the use of residential hydro accounts in the British Columbia regression based population estimation model.

KEY WORDS: Small area population estimates; Regression method; Difference-correlation method; Population indicators; Hydro accounts; Family allowance recipients.

1. INTRODUCTION

The Central Statistics Bureau produces post-censal population estimates for a variety of geographic units within the Province of British Columbia including municipalites, local health areas, census divisions and RCMP regions among others. Current population estimates are produced for these sub-provincial areas by means of a regression approach, specifically the Difference-Correlation Method (DCM).

A detailed description of this methodology is given in earlier papers (Central Statistics Bureau 1982, McRae 1985). The data used as indicators of population are the number of family allowance recipients (F), and/or the number of residential hydro accounts (H). The characteristics of this second data source, residential hydro accounts, relative to the British Columbia model will be examined over the remainder of this paper.

2. DATA SOURCES

Residential hydro accounts data within British Columbia are obtained from nine different organizations. These are:

Organization	% of Total Hydro Accounts (1985)
(1) British Columbia Hydro	90.9
(2) West Kootenay Power and Light	4.7
(3) Princeton Light and Power Co.	0.2
(4) City of Kelowna	0.8
(5) City of Penticton	0.8
(6) District of Summerland	0.3
(7) City of New Westminster	1.7
(8) City of Grand Forks	0.1
(9) City of Nelson	0.6

[1] Presented at the meeting of the Federal-Provincial Committee on Demography, Ottawa, November 28-29, 1985.

[2] D.G. McRae, Central Statistics Bureau, Ministry of Industry and Small Business Development, Government of British Columbia, 2nd Floor, 1405 Douglas Street, Victoria, British Columbia, Canada V8W 3C1.

The views expressed in this paper are those of the author and do not necessarily represent the views of the Government of British Columbia.

The major suppliers of residential electrical power are British Columbia Hydro and West Kootenay Power and Light. The other organizations purchase power from the two major suppliers, and retail this electricity to their own customers (usually the residents of the municipality).

3. DATA FORMAT

Of the nine sources of residential hydro accounts data, only that provided by British Columbia Hydro is in machine readable form. The other eight organizations provide the data totalled by municipality (urban), along with a total of any rural (non-municipal) customers. The reference date for all data is the May 31 billing file, and in most cases the data can be obtained within 2 to 3 weeks of the billing date.

Data provided by British Columbia Hydro is in two formats. The first shows the number of residential meters as of May 31 by Capital District Code. A Capital District Code, of which there are approximately 248 in the Province, is an administrative unit used by British Columbia Hydro and corresponds to a municipality where municipalities exist. By agreement, both the major power suppliers in the Province pay each municipality a certain percentage of the annual revenue collected from the residential customers in that municipality in lieu of property taxes. As a result, power companies such as British Columbia Hydro, design their accounting systems to correspond to customers within municipal boundaries. In addition, British Columbia Hydro attempts to maintain a close correspondence between Capital District boundaries and school district boundaries.

The second format provides for each of the one million plus residential meters the postal code of billing address. This second data file allows the easy translation of hydro meters to geographic units other than municipalities and school districts via the postal code.

4. STRENGTHS OF THE HYDRO DATA IN A REGRESSION MODEL

Empirical tests of the two different data sources, hydro meters (H) and family allowance accounts (F), were conducted by producing 1981 population estimates with each separately and together. The regression coefficients used were derived from the pooled 1971/76 and 1976/81 periods, and the base year was 1976. The results were compared with the 1981 Census, and the Average Absolute Percent Errors (AAPE) were calculated. The results are given in Tables 1 and 2.

As can be seen in Table 1, population estimates based on hydro data produce, on average, lower percentage errors than the family allowance based estimates. Closer examination of Table 1 reveals that the improvement in estimation accuracy lies almost entirely with the estimates for areas with population less than 4000. This observation is reinforced in Table 2 where it is shown that, statistically speaking, there is a significant difference in the estimation accuracy between the hydro and family allowance based estimates for areas less than 4000 population.

The marginal effect of adding another population indicator to the Difference-Correlation Method can be judged by examining the change in estimation accuracy with and without the additional indicator. It would appear from Tables 1 and 2 that the inclusion of hydro data statistically improves the estimation accuracy in both large and small areas. Family allowance data, on the other hand, improves the accuracy for larger areas but reduces the estimation accuracy for smaller areas, with no statistically significant effect overall.

Table 1

Comparison of Estimation Errors Among Data Sources for British Columbia Municipalities - 1981

Data Source	AAPE Overall	n	Population ≥ 4000 AAPE	n	Population < 4000 AAPE	n
DCM/H/F	5.53	158	2.99	88	8.72	70
DCM/H	5.16	158	4.04	88	6.58	70
DCM/F	10.46	167	4.57	92	17.69	75

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

where:

Y_i = census population for region i

$\hat{Y}_i$ = estimated population for region i

n = number of areas estimated.

Table 2

Test for Statistically Significant Differences Between the Average Absolute Percent Errors for Selected Data Sources - 1981

Data Source	95% Confidence Interval for the Average Difference in Asbolute Percent Errors: Overall	Population ≥ 4000	Population < 4000
DCM/H/F − DCM/H	.37 ± .86	−1.05 ± .56[a]	2.14 ± 1.76[a]
DCM/H/F − DCM/F[b]	−4.86 ± 1.55[a]	−1.57 ± .85[a]	−9.00 ± 3.11[a]

[a] Statistically significant differences at the 5% level utilizing a two tailed T-test, paired samples and assuming normally distributed means.

[b] In order to pair the samples only 158 of the possible 167 family allowance base estimates were used. The number of observations were: overall, 158; greater than or equal to 4000, 88; less than 4000, 70.

5. WEAKNESSES OF HYDRO DATA IN A REGRESSION MODEL

One problem encountered when using hydro data in a regression model for population estimates is that of vacant dwellings, or more accurately, significant differences in the rate of vacant dwellings between the base and estimating years. This weakness of the data was demonstrated by the 1981 evaluation of the communities in the Peace River-Liard region of British Columbia (McRae 1982). As a result of the North-East Coal project, the communities of the Peace River-Liard Census Division in 1981 were experiencing a building boom as developers contructed dwellings in anticipation of a population influx. Each dwelling, occupied or not (or even under construction) would require a meter, which may have had low usage, but was still active and hence counted. As a result, the change in share of meters from 1976 to 1981 was overstated relative to population, producing overestimates of the 1981 population for many of Peace-River Liard communities.

Another weakness of the hydro data is the potential for a change-over of multiple dwelling units from single to multiple meters. This may occur when an older apartment building, for example, serviced by a single meter is remodelled or replaced with individually metered units. This problem would produce an overestimate of population in a regression model if it were to occur sometime between the base and estimating years.

Finally, some problems will result if the hydro data is used for areas that have a changing nature, or in other words, a changing relationship between population and residential meters. One example of this in B.C. is the resort community of Whistler. Fifteen years ago this municipality was largely a collection of winter cabins on a ski hill. However, over the last decade this area has been shifting to a year round residence basis. Consequently, the number of persons per hydro meter, which was originally very low relative to the B.C. average, is moving toward the norm. Like the vacancy problem, the use of hydro data to estimate the population for such a community would likely result in above average errors.

The solution to all three of the problems mentioned above is to remove accounts that have a low monthly or bi-monthly usage, and hence are assumed to be vacant. The feasibility of this procedure is currently being examined by the Central Statistics Bureau in relation to the data obtained from B.C. Hydro. If possible, we hope to have the improved data set available for calibration against the 1986 Census. Currently, as a partial solution hydro data for areas that had in 1981 a low or high ratio of persons per meter relative to the provincial norm (i.e. less than 2 or greater than 5) are not used.

A final potential weakness of the hydro data is the reliance on external and different organizations for the information. In the past, this situation generally has not proven to be a problem. However, there have been some rare cases that have called into question the quality of the meter data collected in the field. Such a case may be a boundary change of a municipality not being reflected in the meter data, or the addition of some types of non-residential accounts (such as lamp standards) to the data. As a result, careful monitoring of the data is important.

6. CONCLUSIONS

The following strengths and weaknesses are associated with the use of hydro data in the British Columbia regression based population estimation model.

Strengths:

(a) The hydro data, when used in a regression model, produces a lower average absolute percent error than family allowance data for small areas.

(b) The data is obtained from each supplier in a format that is already aggregated to municipalities. The major advantage of this is that changes in municipal boundaries, which occur regularly, are reflected in the data with no additional work on the part of the Bureau.

(c) The majority of the data can be obtained in machine readable form along with the postal code. This allows the easy translation of the data to geographic regions other than municipalities when sorted by the Bureau's postal code Translation Master File.

(d) The data can be obtained free of charge from each of the suppliers within a relatively short time period (2 to 3 weeks).

Weaknesses:

(a) Differential vacancy rates between the base and estimating years will bias the estimates.

(b) Dwelling units (such as apartments) that change from a single to multiple meter sometime between the base and estimating years will bias the estimates upwards.

(c) Areas with a changing nature, such as from a seasonal to "stable" population, will introduce bias into the estimates.

(d) The data is obtained from external and different organizations. This potentially could cause problems in terms of data quality and comparability, as well as producing a situation in which the priorities of the Bureau's population estimates program are subservient to the administrative needs of an external organization.

REFERENCES

CENTRAL STATISTICS BUREAU (1982). British Columbia municipal population estimation methodology. Unpublished Report, Central Statistics Bureau, Ministry of Industry and Small Business Development, Government of British Columbia.

McRAE, D. (1985). A regression approach to small area population estimation. Paper submitted to the International Symposium on Small Area Statistics, Ottawa, Canada.

McRAE, D. (1982). British Columbia small area estimation model - 1981 municipal and census division evaluation. Unpublished Report, Central Statistics Bureau, Ministry of Industry and Small Business Development, Government of British Columbia.