BRITISH
COLUMBIA

BC STATS

Ministry of Finance and
Corporate Relations

Generalized Estimation System (GES)

Small Area Population Estimation Methodology

Population Section,
BC Stats,
Ministry of Finance and Corporate Relations,
Government of British Columbia,

December, 1998

TABLE OF CONTENTS

Generalized Estimation System (GES)
Small Area Population Estimation Methodology

SUMMARY

The **Generalized Estimation System (GES)** refers to the procedures and methods by which current British Columbia small area population estimates are prepared. A Regression Approach, specifically the **Difference-Correlation Method (DCM)**, is the primary method underlying the sub-provincial population estimates. A secondary method, known as **Proportional Allocation (PA)**, is also used to estimate the population for certain classes of areas. Both these methods use information derived from a set of indicators obtained from administrative files that are symptomatic of regional population changes. Shifts in these indicators are examined and correlated with known changes in regional populations as recorded by the Census of Canada. Once a relationship is derived, population estimates are made by observing changes in the symptomatic indicators and computing the consequent changes in population.

The symptomatic indicators used for small area population estimates are:

1. The number of residential electrical (hydro) connections as of July 1.
2. The number of Old Age Security (OAS) recipients aged 65 and over as of July 1.

In essence, the British Columbia small area population estimation model works as follows: Beginning with the most recent federal census (in this case the 1996 Census of Canada), each region's share of provincial population is adjusted up or down according to the current share of the provincial total of a weighted combination of residential hydro connections and/or Old Age Security recipients. Estimates of the population living in municipalities along with that portion of the population living outside the municipality but within the regional district (i.e., unorganized area), are controlled at the provincial level by a British Columbia population estimate prepared by Statistics Canada. Regional district population estimates are derived by summing the municipal and unorganized area population estimates. Local health areas are also controlled at the provincial level, and in order to ensure consistency, the local health area population estimates within each regional district are then tied to the regional district population estimates.

The key to deriving a small area population estimate using a Regression Approach is the identification of the relationship between population change and symptomatic indicator change. This is accomplished by constructing an equation, via least squares linear regression, that relates past changes in regional populations to changes in the indicators symptomatic of population. The current estimation model utilizes regression coefficients derived from pooled 1986/91 and 1991/96 data.

As noted above, regression-based estimates are supplemented in a number of specific cases with Proportional Allocation-based estimates. Empirical evidence indicated that for the smaller communities (less than 4,000 population), more accurate population estimates could be obtained by using a Proportional Allocation-based approach as opposed to the Regression approach.

The accuracy of both the developed municipal and regional district, and local health area population estimation models was assessed by comparing the 1996 estimated values with the 1996 Census populations adjusted for net census undercount and incompletely enumerated Indian Reserves.[1] The Average Absolute Percent Error (AAPE) was used as the primary summary measure of estimation accuracy. Summary expressions of the differences between two sets of population figures may take many different forms, each giving slightly different information. Other summary statistics used were the median absolute percent error, frequency distributions of percent error and the Index of Dissimilarity.[2] The performance of the model is given in Tables 1 through 5.

Due to the nature of the methodology, the errors associated with the estimates are expected to increase with time. Since the errors represented in Tables 1 through 5 are calculated on the basis of a five year change (1991-96), they are expected to over-represent the error associated with the estimates during the first, second, third and fourth years subsequent to a census.

1 It is estimated that there was a net undercount (i.e., those not counted, minus those counted more than once, or counted when they should not be counted) of 145,579 persons in British Columbia in the 1996 Census. Of this total, it is estimated that 3,136 persons living on 19 Indian Reserves were incompletely enumerated during the 1996 Census. See Section 4 for a summary of census adjustments.

2 Simply defined, if the resulting population estimates were used to allocate funds, the Index of Dissimilarity gives the percentage of all funds allocated on a per capita basis that would have to be reallocated in order to achieve zero misallocation.

1 INTRODUCTION

(I) PURPOSE

The purpose of this paper is to outline the methodology currently used to estimate the population for British Columbia small areas, in particular, municipalities, regional districts and local health areas.

(II) BACKGROUND

In 1978 the Central Statistics Bureau (the precursor to BC Stats) began work on a program of current population estimates for sub-provincial areas. A number of synthetic approaches to population estimation were proposed and tested [Central Statistics Bureau 1980]. By 1979, the Regression Approach – specifically the Difference-Correlation Method (DCM) – was chosen as the primary method of small area population estimation. An alternative method known as Proportional Allocation (PA), was also used to estimate the population for certain classes of areas.

The accuracy of the initial methodology was assessed relative to the 1976 Census. With the advent of the 1981, 1986, 1991 and 1996 Censuses of Canada additional testing of the methodology was carried out, resulting in further refinements to the initial model and input data sources [BC Stats 1993, Central Statistics Bureau 1989, McRae and Gieringer 1987, McRae 1982, Central Statistics Bureau 1982].

2 ESTIMATING THE TOTAL POPULATION BY SMALL AREA - METHODOLOGY -

(I) THE DIFFERENCE-CORRELATION METHOD

Population estimates for most small areas are derived by first estimating the change in each area's proportion of the provincial population since the last census.[3] This estimate, along with the proportion each area represented in the last census, enables a post-censal estimate of an area's population proportion to be made. This estimated proportion is then applied to the Statistics Canada estimate of the British Columbia provincial population resulting in an estimate of the small area.[4]

3 The Difference-Correlation Method is used to estimate the population for 111 out of 179 municipal regions. These 111 regions represent 97 percent of the provincial population. For Local Health Areas, the Difference-Correlation Method is used to estimate all but 9 out of 98 regions, representing over 99 percent of the population.

4 For details of the methodology underlying the estimate of the total British Columbia population see: Statistics Canada, (Demography Division), *Population Estimation Methods, Canada*, Catalogue 91-528E, Supply and Services Canada, March 1987, p 9-16.

The method by which the change in each area's share of the provincial population since the last census is estimated is known as Difference-Correlation. Briefly, this method uses information derived from a set of indicators symptomatic of regional population. Past changes in these indicators are examined and correlated with known changes in regional populations. Once a relationship is derived, changes in population shares are estimated by applying this relationship to the observed changes in the population indicator shares.

The population indicators used are:

1. The number of residential electrical (hydro) connections as of July 1 of each year.
2. The number of Old Age Security (OAS) recipients aged 65 and over as of July 1 of each year.

The relationship between the change in population share and the change in indicator share is derived via the use of least squares linear regression. This involves regressing the change in indicator share between two or more censuses on the change in population shares.

The general form of the Difference-Correlation regression equation is:

$$\Delta \hat{R}_i = \sum_{j=1}^{N} \beta_j \bullet \Delta X_{ij} + u_i \qquad [1]$$

Where:

β_j is the regression coefficient for symptomatic variable j.

N is the number of symptomatic variables, $j = 1,.., N$.

u_i is the error term which has mean (m) = 0, and variance = s_u^2

ΔX_{ij} is the difference in the area's proportion of symptomatic variable j between the previous census year and the estimating year. For example, the difference in a municipality's share of total provincial residential electrical accounts between the last census and July 1 of the estimating year.

$$\Delta X_{ij} = \frac{X_{ij}^{t}}{X_{Tj}^{t}} - \frac{X_{ij}^{t_0}}{X_{Tj}^{t_0}}$$

$X_{ij}^{t}, X_{ij}^{t_0}$ is symptomatic variable j for area i in the base (t_0) and estimating (t) years.

$X_{Tj}^{t}, X_{Tj}^{t_0}$ is symptomatic variable j for British Columbia in the base (t_0) and estimating (t) years.

$\Delta \hat{R}_i$ is the difference between the previous census year and the estimating year in the ratio of population for area *i* to the total provincial population.

$$\Delta \hat{R}_i = \frac{P_i^t}{P_T^t} - \frac{P_i^{t_0}}{P_T^{t_0}}$$

$P_i^t, P_i^{t_0}$ is the population of area *i* for the base (t_0) and estimating (t) years.

$P_T^t, P_T^{t_0}$ is the British Columbia population for the base (t_0) and estimating (t) years.

Having estimated $\Delta \hat{R}_i$ for all areas *i* using equation [1], the population estimate is calculated by solving the equation:

$$P_i^t = \left[\frac{P_i^{t_0}}{P_T^{t_0}} + \Delta \hat{R}_i \right] \bullet P_T^t \qquad [2]$$

(II) DEVELOPING THE REGRESSION EQUATIONS

(a) Data Collection

The accuracy of the small area population estimates derived from a regression-based model is heavily dependent on the ability of the population indicator data selected to accurately reflect population change. As a result, the selection as well as the care with which the indicator data is compiled by small area are important parameters in the overall population estimation methodology. As noted earlier, the population indicators used by the current model are the number of residential electrical accounts (**H**) and the number of people 65 and over who are receiving Old Age Security (**O**). At one time the number of Family Allowance recipients aged 1-14 were also used, but since the federal government discontinued the Family Allowance program, this data source is no longer available. Efforts are being made to examine other indicator data to be incorporated into the Generalized Estimation System (GES) model to replace the Family Allowance data. Currently, school enrolment data is being tested.

Information on the number and age of persons registered in the Old Age Security program is available from Human Resource Department Canada. This data is aggregated to specific geographic areas on the basis of the postal code of the cheque mailing address. All persons in Canada aged 65 or over who are Canadian citizens or legal residents are eligible for OAS if they have resided in Canada for the 10 years immediately prior to application. Hence, the number of OAS recipients closely approximates the population 65 years and over.

B.C. Hydro, West Kootenay Power and Light and other organizations make available to BC Stats information on the number of residential electrical accounts by municipality as of July 1 of each year.[5] An account is associated with an electrical meter and generally refers to an occupied private dwelling unit. Since 1991, information on the average daily consumption of electricity has been used as a proxy for occupancy. Residential electrical accounts for a house or duplex, mobile home and row house with an average consumption during the sixty days prior to July 1 of less than 3 kilowatt hours per day are assumed to be vacant and removed from the data set. In addition, for data supplied by B.C. Hydro, where the premise code attached to the data indicates the meter is for a seasonal dwelling or a non-dwelling (e.g., a swimming pool or lamp post), the meter is removed from the data set.

Counts of the number of residential electrical meters by municipality are compiled by each of the organizations from which the data is obtained. Both the major electrical power suppliers are required to pay a portion of their residential billing to the municipalities in lieu of property taxes. As a result, the power companies have designed their accounting systems to correspond to customers within municipal boundaries, thereby removing the need to geo-code the municipal data based on postal codes. However, data for local health areas still need to be geo-coded.

In order to construct the regression equations, indicator data (i.e., hydro and OAS) were collected for three points in time: 1986, 1991 and 1996. The change in share observed for both these population indicators (i.e., 1986/91 and 1991/96) was related via least squares linear regression to the change in share recorded by the 1986, 1991 and 1996 Censuses.

(b) Correction for Heteroscedasticity

A fundamental assumption of Ordinary Least Squares Regression is that of constant variance of the residual term. In this particular application, tests indicated that the assumption of homoscedasticity (homogeneity of the variances) was violated.

[5] Other organizations that provide information on the number of residential electrical connections are: Princeton Power and Light, City of Kelowna, City of Penticton, District of Summerland, City of New Westminster, City of Grand Forks, City of Nelson.

The disturbance variable (or error term) u given in equation [1] did not have the same variance for all values of the explanatory variables. In this case, the usual least square estimation procedure is no longer optimal, since the estimators no longer have minimum variance. As a result, an alternative known as Weighted Least Squares Linear Regression was used.[6] Weights were derived and used to transform the variables such that the assumption of homoscedasticity was satisfied. To accomplish this, two groups of areas were established on the basis of the size of their estimated change in share of population ΔR_i since σ^2_{ui} increases as ΔR_i increases. The groups were distinguished according to whether ΔR_i was above or below the median value. A weight for each group (i.e., that with small and large ΔR_i) was then estimated such that:

$$S^2_s = S^2 \bullet w^2_s$$

and

$$S^2_1 = S^2 \bullet w^2_1$$

Where:

- S^2_s is the estimated variance of the error term (σ^2_u) for the “small” group,
- S^2_1 is the estimated variance of the error term (σ^2_u) for the “large” group,
- S^2 is the estimated variance of the error term (σ^2_u) for all areas,
- w^2_s, w^2_1 is the estimated value of the weights.

Equation [1] is thus transformed into:

$$\Delta \hat{R}_i \bullet w_k = \sum_{j=1}^{N} B_j \bullet \Delta X_{ij} \bullet w_k + v_i \qquad [3]$$

Where:

$$v_i = u_i \bullet w_k$$

and

$w_k = w_s$ for all k *1* “small” group (i.e., with small $\Delta \hat{R}_i$),

$w_k = w_1$ for all k *1* “large” group (i.e., with large $\Delta \hat{R}_i$).

The weights increase the variance of the disturbance term for areas with small ΔR_i and hence small S^2_u, and decrease the variance for areas with large ΔR_i. The result is a constant S^2_v where $S^2_v = w^2_k \bullet S^2_u = S^2_v$ for all *i*.

---

[6] For more detail see: Pindyck and Rubinfield, *Econometric Models and Economic Forecasts*, 2nd Edition, New York: McGraw-Hill, 1976, p. 140-152.

(c) Municipal and Unorganized Area Population Estimation Equation

The weighted regression equation used to produce the post-censal population estimates for British Columbia municipalities and unorganized areas within regional districts was developed from a combination of cross-sectional and time-series observations. That is, the weighted regression equation was based on the "cross-sectional" variation observed among the areas for the two time periods of 1986/91 and 1991/96. The derived equation used to estimate the municipal populations for the period 1997 to 2001 is given below.

$$\Delta \hat{R}_i = 0.803939 \bullet \Delta H_i + 0.115522 \bullet \Delta O_i \qquad [4]$$

T-Statistic	17.6	6.6
Standard Error	(0.045560)	(0.017390)

R^2=68.3%
F Value=301.1
Error degrees of freedom=280

Where:

$\Delta \hat{R}_i$ is the estimated change in the share of population between the previous census year and the estimating year for area *i*.

ΔH_i is the estimated change in the share of residential electrical meters between the previous census year and the estimating year for area *i*.

ΔO_i is the estimated change in the share of old age security recipients aged 65 and over between the previous census year and the estimating year for area *i*.

(d) Local Health Area Population Estimation Equation

Similar to municipalities, the weighted regression equation used to produce the post-censal population estimates for British Columbia local health areas was developed from a combination of cross-sectional and time-series observations. The derived equation used to estimate the health area populations for the period 1997 to 2001 is given below. Note that Old Age Security data was not significant when combined with the hydro data and was removed from the equation.

$$\Delta \hat{R}_i = 1.041657 \bullet \Delta H_i \qquad [5]$$

T-Statistic	22.7
Standard Error	(0.045866)

R^2=78.1%
F Value=515.8
Error degrees of freedom=145

Where:

$\Delta \hat{R}_i$ is the estimated change in the share of population between the previous census year and the estimating year for area *i*.

ΔH_i is the estimated change in the share of residential electrical meters between the previous census year and the estimating year for area *i*.

(e) Allowing for Missing or Incorrect Observations

In this particular application of the Difference-Correlation Method, there were some cases when one of the population indicators was either unavailable or suspected of being incorrectly compiled.[7] In such cases, municipal estimates were prepared using only one of the population indicators according to the following equations:

$$\Delta \hat{R}_i = 0.368953 \bullet \Delta O_i \qquad [6]$$

T-Statistic	13.0
Standard Error	(0.029120)

R^2=36.9%
F Value=170.1
Error degrees of freedom=291

$$\Delta \hat{R}_i = 0.971493 \bullet \Delta H_i \qquad [7]$$

T-Statistic	23.9
Standard Error	(0.040615)

R^2=67.1%
F Value=572.1
Error degrees of freedom=281

Where:

$\Delta \hat{R}_i$ is the estimated change in the share of population between the previous census year and the estimating year for area *i*.

ΔH_i is the estimated change in the share of residential electrical meters between the previous census year and the estimating year for area *i*.

---

7 Generally, there is only a small number of cases in which data is missing or rejected. For example, in the case of the 1996 municipal and unorganized area estimates approximately 9 out of 179 areas were affected. For local health areas, only 3 out of 98 areas were affected.

ΔO_i is the estimated change in the share of old age security recipients aged 65 and over between the previous census year and the estimating year for area *i*.

For local health areas, the equation[8] was as follows:

$$\Delta \hat{R}_i = 0.344001 \bullet \Delta O_i \quad [8]$$

T-Statistic	9.1
Standard Error	(0.037610)

R^2=36.0%
F Value=83.7
Error degrees of freedom=149

Where:

$\Delta \hat{R}_i$ is the estimated change in the share of population between the previous census year and the estimating year for area *i*.

ΔO_i is the estimated change in the share of old age security recipients aged 65 and over between the previous census year and the estimating year for area *i*.

(III) THE PROPORTIONAL ALLOCATION METHOD

As noted earlier, the current British Columbia small area population estimation model uses a combination of two estimating methodologies. The primary method is Difference-Correlation which is used to estimate the population for 97 percent or more of the province. Population estimates for the remaining 3 percent or less are prepared using a method known as Proportional Allocation.

8 Since the OAS data was excluded from the combined equation due to insignificance, the Hydro equation does not change.

A Proportional Allocation-based population estimate is made by first relating the symptomatic indicator share to the census population share for an area. If one assumes this relationship to be stable in future years, a current estimate of the population can be made using the following equation:

$$\frac{\left[\frac{X_i^{t_0}}{X_T^{t_0}}\right]}{\left[\frac{P_i^{t_0}}{P_T^{t_0}}\right]} = \frac{\left[\frac{X_i^{t}}{X_T^{t}}\right]}{\left[\frac{P_i^{t}}{P_T^{t}}\right]} \qquad [10]$$

Solving for P_i^t:

$$P_i^t = \frac{X_T^{t_0} \bullet P_i^{t_0} \bullet X_T^t \bullet P_T^t}{X_i^{t_0} \bullet P_T^{t_0} \bullet X_T^t} \qquad [11]$$

Where:

P_i^t	is the estimated population for area *i* at time *t*.
$X_T^{t_0}, X_T^t$	are the indicator variable totals in the base and estimating years.
$P_i^{t_0}$	is the census population for area *i*.
$X_i^{t_0}, X_i^t$	are the indicator variable counts for area *i* in the base and estimating years.
$P_T^{t_0}, P_T^t$	are the total populations in the base and estimating years.

The major difference between the Difference-Correlation Method with one indicator and that of Proportional Allocation is that the Difference-Correlation Method takes the fitted least squares value (in essence a type of average over time and/or across regions) of the past relationship between the difference (over time) in indicator shares and the difference in population shares, and applies this "average" relationship to the present. The magnitude of the slope parameter of the resulting regression line could be greater or less than one depending on the historical relationship between differences in indicator shares and differences in population shares.

The Proportional Allocation Method, on the other hand, takes the relationship between the ratio of population shares at a previous point in time for a given place and applies this relationship to the present. The theoretical "slope parameter" of the Proportional Allocation line is not allowed to float to its least squares "average" value, but rather is fixed at one.

The theoretical advantage of the DCM approach over PA is that average historical relationships are implicitly stable and therefore are useful in the post-censal period when re-estimation of the coefficients is impossible. However, a disadvantage of the least squares estimate of the slope parameter is the possibility of outliers in the data adversely affecting the slope of the regression line, or similarly, of averages taken over many areas and time periods being inappropriate for a given place and time.

Empirical evidence indicated that for smaller population centres (less than 4,000), the Proportional Allocation Method produced population estimates with a lower (absolute) percent error on average relative to census than the Difference-Correlation Method. As a result, the Generalized Estimation System (GES) uses both approaches to population estimation. The Difference-Correlation Method is used to estimate the population of the larger centres, and the Proportional Allocation Method is used for the smaller areas.

3 EVALUATION OF THE SMALL AREA POPULATION ESTIMATES FOR BRITISH COLUMBIA

(I) EVALUATION MEASURES

In order to approximate the error associated with the developed methodology, a comparison was made between the 1996 Census population adjusted for net census undercount and incompletely enumerated Indian Reserves, and the estimated population.[9] The accuracy of the 1996 population estimates was assessed through a number of summary measures. These were:

- average absolute percent error or difference (AAPE)
- median absolute percent error (MAPE)
- standard deviation about the average of the absolute percent errors
- frequency distributions of percent errors
- Index of Dissimilarity (ID)

Of the preceding summary measures, the last one (ID) warrants a further explanation:

[9] It is estimated that there was a net undercount (i.e., those not counted, minus those counted more than once, or counted when they should not be counted) of 145,579 persons in British Columbia in the 1996 Census. Of this total, it is estimated that 3,136 persons living on 19 Indian Reserves were incompletely enumerated during the 1996 Census. See Section 4 for a summary of Census adjustments.

When population estimates are used as a basis for the allocation of funds, the extent to which a given total population is misallocated among the component parts will directly affect the extent to which there exists misallocation of the population-based funding. Where misallocation occurs, one or more areas may obtain more than the intended entitlements at the expense of one or more other areas. It is therefore desirable to have some measure of the misallocation that could result. Such a measure is found in the Index of Dissimilarity (ID).[10]

The Index of Dissimilarity gives the percentage of all funds allocated on a per capita basis that would have to be reallocated in order to achieve zero misallocation. For example, an ID of 1.0 implies that a shift of one percent of the per capita grants budget is required in order to change the allocation of funds from that obtained using the estimated populations to that obtained using the census figures. Hence, ID measures the inequity associated with the set of population estimates when used for population-based grants.

Algebraically, the Index of Dissimilarity is represented as:

$$ID = \left[\frac{\sum_{i=1}^{n} \left| C_i - E_i \right|}{\sum_{i-1}^{n} C_i} \right] \times 100 \times \frac{1}{2} \qquad [12]$$

Where:

C_i is the census population for area *i*.
E_i is the estimated population for area *i* adjusted such that $\sum C_i = \sum E_i$.

(II) EVALUATION RESULTS

The various summary measures for the evaluation of the present municipal and regional district population estimation methodology are given in Table 1.[11] Tables 2 and 3 give a breakdown of the error by population size. Tables 4 and 5 give summary measures for the evaluation of the local health area population estimation methodology.

---

10 See: Shryock and Siegel, *The Methods and Materials of Demography*, 1971, p. 232-233; and *State of Washington Population Trends 1979*, Office of Financial Management Population, Enrollment and Economic Studies Division, August 1979, p. 62.

11 It is assumed that the population estimates at the provincial level are correct. Hence, the GES methodology is being evaluated independent of any errors made by Statistics Canada at the provincial level.

In addition, tests indicate that the error structure of the resulting population estimates is not significantly affected by the time period upon which the regression coefficients were developed. Hence, even though the GES model coefficients were developed using data from the same time period over which the evaluation was conducted (i.e., 1991/96), the results indicate an accurate representation of the true error structure.

Table 1

1996 Population Estimation Accuracy for British Columbia Municipalities and Regional Districts (1997/01 Model)

	Municipalities	Regional Districts
AAPE*	4.8%	2.2%
MAPE!	3.4%	2.0%
ID+	1.2%	0.9%
% Overestimated	63.2%	82.8%
No. of Observations:		
less than 5%	125	28
5%-10%	39	1
10%-15%	9	0
greater than 15%	9	0

* Average Absolute Percent Error
! Median Absolute Percent Error
\+ Index of Dissimilarity

Table 2

1996 Population Estimation Accuracy for British Columbia Municipalities by Size (1997/01 Model)

Population Size	AAPE*	MAPE!	Number of Observations
0 - 500	12.5%	9.8%	7
501 - 1,000	8.7%	5.9%	19
1,001 - 2,500	8.1%	6.2%	27
2,501 - 5,000	4.4%	3.6%	31
5,001 - 10,000	2.2%	1.4%	26
10,001 - 25,000	3.6%	3.5%	38
25,000 and over	1.9%	1.5%	34
Overall	4.8%	3.4%	182

* Average Absolute Percent Error
! Median Absolute Percent Error

Table 3

1996 Population Estimation Accuracy for British Columbia Regional Districts by Size[12] (1997/01 Model)

Population Size	AAPE*	MAPE!	Number of Observations
0 - 25,000	1.9%	1.5%	4
25,001 - 50,000	3.2%	3.7%	8
50,001 and over	1.8%	1.8%	17
Overall	2.2%	2.0%	29

* Average Absolute Percent Error
! Median Absolute Percent Error

Table 4

1996 Population Estimation Accuracy for British Columbia Local Health Areas (LHAs) (1997/01 Model)

AAPE*	4.5%
MAPE!	2.7%
ID+	1.4%
% Overestimated	66.3%
No. of Observations:	
less than 5%	72
5%-10%	18
10%-15%	3
greater than 15%	5

* Average Absolute Percent Error
! Median Absolute Percent Error
\+ Index of Dissimilarity

[12] Note that this evaluation does not include the Stikine region. This region is very small and its population is unstable and the indicators do not accurately reflect population in the area. As a result, the average absolute percent error for this region was 42 percent. Including the region in the evaluation distorts the true error associated with the model when estimating regional districts.

Table 5

1996 Population Estimation Accuracy for British Columbia Local Health Areas by Size (1997/01 Model)

Population Size	AAPE*	MAPE!	Number of Observations
0 - 10,000	7.4%	4.4%	29
10,001 - 30,000	3.9%	2.5%	32
30,001 and over	2.6%	1.9%	37
Overall	4.5%	2.7%	98

* Average Absolute Percent Error
! Median Absolute Percent Error

(III) CHANGE IN THE EXPECTED ERROR DURING THE FIVE-YEAR POST-CENSAL PERIOD

The expected error for each year of the five-year post-censal range cannot be accurately determined. However, the variance associated with the dependent variable (i.e. the change in share of the population, ΔR_i) suggests that the error is expected to increase with time.

Assuming that ΔR_i is a function of one independent (symptomatic) variable only, ΔX_i, then the variance of ΔR_i for a particular value of the symptomatic variable, ΔX_i, is given as:

$$\mathrm{Var}\left(\Delta \hat{R}_i \middle| \Delta X_i\right) = \sigma_u^2 \bullet \left[1 + \frac{1}{n} + \frac{(\Delta X_i)^2}{\sum (\Delta X_i)^2}\right] \qquad [13]$$

The right hand term usually includes the term $\Delta\overline{X_i}$ which in this case is 0 since $\sum \Delta X_i$ must equal 0.

Historical data clearly indicates that both ΔX_i and ΔR_i tend to increase with time. That is, in general there are larger changes in the share of the population over five years than over one year. Given this observation, there are two reasons why $\mathrm{Var}(\Delta R_i)$ may increase with time. First, it was shown that heteroscedasticity resulted from the fact that the variance of the error term, σ_u^2, increased as ΔR_i (or ΔX_i) increased. Hence, σ_u^2 increases with time (since ΔR_i generally increases with time), increasing $\mathrm{Var}(\Delta R_i)$ in the above equation. Second, the $(\Delta X_i)^2$ term increases with time, since ΔX_i tends to increase with time. For these reasons, $\mathrm{Var}(\Delta R_i)$ will likely be at a maximum at the end of the five year post-censal period.

Since the population is estimated by equation [2]:

$$P_i^t = \left[\frac{P_i^{t_0}}{P_T^{t_0}} + \Delta \hat{R}_i \right] \bullet P_T^t \qquad [2]$$

when $Var(\Delta R_i)$ is at a maximum, the error in P_i^t will also be at a maximum.

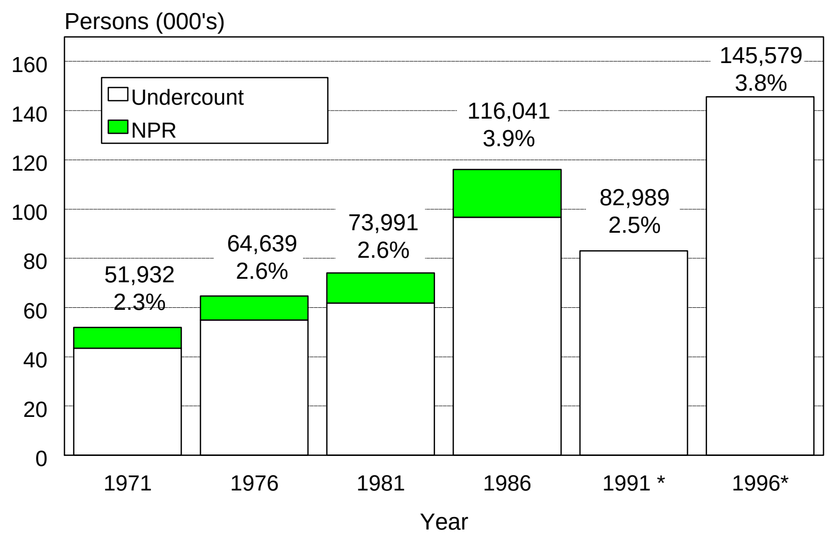
4 CENSUS ADJUSTMENTS

In January of 1993, Statistics Canada, after extensive consultation and research, announced that the 1991 Census base used for post-censal population estimates would be adjusted for net census undercount as measured by the Reverse Record Check and the Over-coverage study. In addition, in order to have a consistent time series of population data, similar adjustments would be made to the national and provincial population estimates back to 1971. The population adjustments prior to 1991 would also include estimates of non-permanent residents (i.e., student visa holders, work permit holders, refugee claimants and ministerial permit holders) in order to be consistent with the 1991 Census data.[13]

These adjustments have had a significant effect on the official population counts for British Columbia as seen in Figure 1.

Figure 1

Net Addition to the B.C. Population Due to Net Census Undercount and the Addition of Non Permanent Residents (NPR)



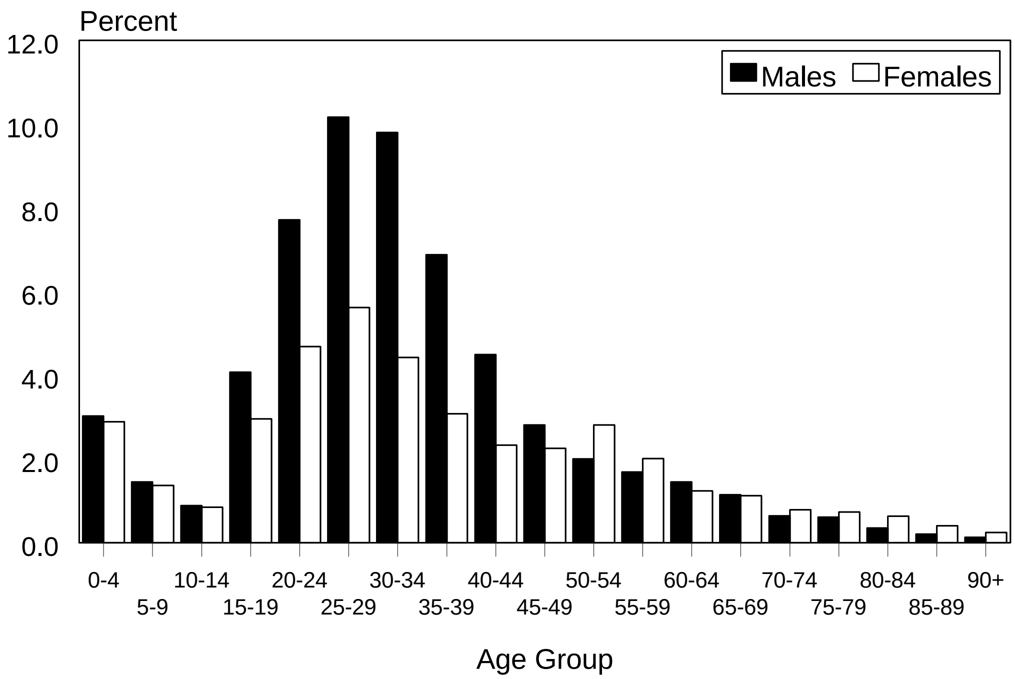
BC Stats *Net Census Undercount only

[13] For a description of the methods used to adjust the provincial population estimates for net census undercount and non-permanent residents see Tourigny, Bureau and Clark [1998] and Declos [1993].

On a sub-provincial basis, the effect of these changes is very similar to that observed at the provincial level. As can be seen from Figure 2, the net census undercount is not evenly distributed by age and gender. For this reason, Statistics Canada distributed the undercount based on not only share of population, but on the age and gender distribution of the area as well. To maintain consistency, the provincial net census undercount was distributed sub-provincially based on the population share by age and gender.

Figure 2

Distribution of Net Census Undercount 1996



BC Stats

The addition of non-permanent residents to the sub-provincial census counts for 1971, 1976, 1981 and 1986 was based on the regional distribution of non-permanent residents recorded by the 1991 Census. This distribution favoured the Greater Vancouver Regional District municipalities where almost 80 percent of the non-permanent residents reside.

The final adjustment was the addition to the 1986, 1991 and 1996 Census population counts for persons living on incompletely enumerated Indian Reserves. Statistics Canada estimated that 5,840 persons living on 63 Indian Reserves were not counted in the 1986 Census, 3,472 persons on 18 Indian Reserves were not counted in 1991, and 3,136 persons on 19 Indian Reserves were not counted in 1996. Estimates of the missed population for 1986, 1991 and 1996 were prepared for each of the missed Indian Reserves based on information from the 1981, 1986 and 1991 Censuses, and assumptions with respect to population growth rates for each Reserve.

REFERENCES

BC STATS (1993). "Generalized Estimation System (GES) – Local Health Area Population Estimation Methodology." BC Stats, Ministry of Government Services, Government of British Columbia.

BC STATS (1993). "Generalized Estimation System (GES) – Municipal and Regional District Population Estimation Methodology." BC Stats, Ministry of Government Services, Government of British Columbia.

CENTRAL STATISTICS BRANCH (1989). "Generalized Estimation System (GES) – Municipal and Regional District Population Estimation Methodology." Unpublished report, Central Statistics Branch, Ministry of Finance and Corporate Relations, Government of British Columbia.

CENTRAL STATISTICS BUREAU (1980). "British Columbia Municipal Population Estimation Methodology." Unpublished report, Central Statistics Bureau, Ministry of Industry and Small Business Development, Government of British Columbia.

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.

DECLOS, M. (September 15, 1993). "Revised Postcensal and Intercensal Estimates - Canada, Provinces and Territories, 1971-1993 - Total Population." Unpublished report, Statistics Canada, Demography Division, Ottawa.

MCRAE, D. G. (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.

MCRAE, D. G. (1985). "The Use of Hydro Accounts in the British Columbia Regression-based Population Estimation Model." *Survey Methodology*, December 1985: page 197-201.

MCRAE, D. G. (1986). "A Regression Approach to Small Area Population Estimation." *Small Area Statistics International Symposium ,85 (Contributed Papers)*. Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Canada, August 1986: page 80-104.

MCRAE, D. G., AND M. GIERINGER (1987). "1986 Evaluation of the British Columbia Municipal and Census Division Population Estimates." Paper presented to the Federal-Provincial Committee on Demography, December 10-11, 1987, Ottawa, Ontario, Canada.

STATISTICS CANADA (1987). *Population Estimation Methods, Canada (91-528E)*. Supply and Services Canada, March 1987, Ottawa, Canada.

TOURIGNY, J., BUREAU, M., AND C. CLARK (September 24, 1998). *Final Direct Estimates of 1996 Census Coverage Error*. Unpublished Report, Statistics Canada, Social Survey Methods Division, Ottawa.

TOURIGNY, J., BUREAU, M., AND C. CLARK (September 24, 1998). *Revised Direct Estimates of 1991 Census Coverage Error*. Unpublished Report, Statistics Canada, Social Survey Methods Division, Ottawa.

BIBLIOGRAPHY

CHOW, G. 1960. "Test of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica* 28: 591-605.

D'ALLESANDRO, F. AND J. TAYMAN. 1981. "Ridge Regression for Population Estimation: Some Insights and Clarifications." Paper presented at the Annual Meeting of the Population Association of America, Washington, D.C.

MANDELL, M. AND J. TAYMAN. 1982. "Measuring Temporal Stability in Regression Models of Population Estimation." *Demography*, February 1982: 135-146.

O'HARE, W. 1976. "Report on a Multiple Regression Method for Making Population Estimates." *Demography* 13: 369-380.

O'HARE, W. 1980. "A Note on the use of Regression Methods in Population Estimates." *Demography* 17: 341-343.

PINDYCK, R. S. AND D. L. RUBINFIELD. 1981. *Econometric Models and Economic Forecasts*. Second Edition. New York: McGraw-Hill.

RAO, C. R. 1973. *Linear Statistical Inference and its Application*. Second Edition. New York: John Wile and Sons.

SCHMITT, R. AND A. CROSETTI. 1954. "Accuracy of the Ratio-Correlation Method for Estimating Postcensal Population." *Land Economics*. 30: 279-281.

SHRYOCK, H. S. AND J. S. SIEGEL. 1971. *The Methods and Materials of Demography*. U.S. Department of Commerce.

SWANSON, D. 1980. "Improving Accuracy in Multiple Regression Estimates of Population Using Principals from Casual Modelling." *Demography*. 17: 413-427.