SMALL AREA HOUSEHOLD PROJECTIONS

- A Parameterised Approach -

REVISION 1999

by

Frank Ip and Don McRae

BC STATS
Population Section
Ministry of Finance and Corporate Relations
Province of British Columbia

March 1999

CONTENTS

I. Background

As is the case with population projections, projections of the number of private households by small area are of value for planning purposes. Public and semi-public sector organizations require accurate small area household statistics on which to base planning decisions for resource allocation and development of regional infrastructure, as well as for analysing and formulating public policies. Private sector organizations require future information on the households to carry out market research and plan operating strategies for household-related consumption of goods and services.

The first set of sub-provincial private household projections for British Columbia was prepared in 1986, and projected the number of households by seventy-eight regions ranging in size from 170 to 190,000 for the years 1982 to 2006. The first projection was based on a model developed from data obtained from the 1976 and 1981 Censuses. In this update, pooled 1991 and 1996 Census data are used as the basis for projecting household counts to the year 2026. Many updates have been done since the projection model was originally developed in 1986. For descriptions and evaluations of various past updates, refer to different revisions of the unpublished reports “Small Area Household Projections” by BC STATS listed in the Reference section of this paper.

II. Summary of the Method

The traditional and most widely used method in household projections is the headship rate method. With this method, the trend in headship rates is projected forward and applied to the projected population to obtain a projection of households. One of the major advantages of this method is that it can project households by types and by different characteristics of the household heads. However, this method requires detailed age-sex specific and time series data, which are not always available by small area.

When BC STATS initially researched projecting households by small area, the main objective was to develop a projection model that was reliable, relatively inexpensive, and able to use the small area population projections that are produced and updated routinely. After extensive analysis and careful consideration of different household projection methods, a parameterisation approach was selected for model development.

The primary component of the British Columbia small area household projection methodology is a projection of the average number of persons per private household by small area. Once the number of persons per household for some future point in time has been predicted, the number of private households is derived by applying this ratio to an independently derived population projection.

Over time there has been a general decline in the average number of persons per household at the national, provincial and regional levels. It can be hypothesised that household formation and household size are influenced by a number of demographic and socio-economic factors such as rents, interest rates, house prices, per capita income, degree of urbanisation, female participation rates, marriage rates, divorce rates, the number of children, etc.

It was theorised that if the variation in the number of persons per household over time and across regions could be explained by the variation in predictable demographic variables, then projections of persons per household by small area could be derived. Cross-sectional multiple linear regression analysis was the fundamental approach adopted for the parameterisation of the relationship between persons per household, and a number of demographic variables thought to influence household formation and household size.

In the original model developed in 1986, pooled data from the 1976 and 1981 Censuses were used. Using average number of persons per household by seventy-eight sub-provincial areas as the dependent variable, three demographically driven independent variables were determined as predictors of household size[1]. These were:

1. *The child population aged 0 to 19 by small area.* Child population was theorised to have a positive correlation with the average number of persons per household. Generally, children under 19 do not form households of their own. Hence, in areas with a relatively high child population, the number of persons per household would also tend to be high.

2. *The married population by small area.* Marital status of the population will affect household formation. The married population accounts for a major proportion of household formation, and thus can influence average household size. The direction in which the size of the married population affects the average household size cannot be intuitively predicted. It depends largely on the life-style of different marital status groups. In this model, it was found that the average household size by small area was inversely related to the size of married population.

   Although no study has yet been conducted by BC STATS to further analyse this phenomenon, one possible explanation is that marriages sometimes result in the dissolution of larger households and the creation of smaller households, rather than the reverse. For example, a family of four may result in two families of three and two each, after the marriage of one child.

3. *The divorced and separated populations by small area.* Divorce and separation affect the type and way a household is formed. Again, the relationship between the size of the divorced and separated population and the average household size is a function of the life-style of the divorced/separated population. For example, people in this divorced/ separated category sometimes form lone-parent families or single-person households, which are generally associated with a smaller household size.

[1] See, "British Columbia Small Area Household Projections -Methodology", F. Ip and D. McRae, unpublished report, Planning and Statistics Division, Ministry of Finance and Corporate Relations, Government of British Columbia, April, 1986.

Given the limited data available at the sub-provincial level, socio-economic factors also thought to influence household size and household formation (i.e., interest rates, house prices, etc.) were not tested. As a result, the derived methodology was largely demographically driven.

The natural log was applied to all dependent and independent variables in order to minimise the disruptive effects due to extreme cases in the data on the generation of the parameters. For example, in the earlier model developed in 1986, it was found that in one area, the child population was almost 300 times greater than that for another region. The application of a log function attempts to normalise the range of the data, and thereby helps to remove some outlier effects during the model building phase. This method was proven effective by comparing the regression results from the logged and the unlogged equations. The results indicated that the logged equation increased the explanatory power of the model by almost 60 per cent.

The last step involved in the projection of households by small area was the projection of the independent variables themselves. Age/gender specific population projections by small area are already prepared by BC STATS using a Component/Cohort-Survival approach.[2] Hence, future counts of the total and child populations by small area were taken from the output of this population projection model.

Projections of the married, separated and divorced populations by small area were not readily available. Hence, some assumptions were made regarding the future growth or decline in these populations in each of the sub-provincial areas. Specifically, the trend over the last four censuses was extended to forecast future marriage and divorce/separation rates. For each sub-provincial area, two separate regression equations, one for marriage rates and the other for divorce/separation rates were created through time-series regression techniques using information obtained from the four censuses. The procedure is described in more detail in Appendix A of this paper.

In the updates subsequent to 1986, the structure of the model and regression techniques remained the same; however, the output performance of the model has been improved by incorporating additional information made available by the 1986, 1991 and 1996 Censuses.

[2] See "Population Extrapolation for Organizational Planning with Less Error (PEOPLE)," Population Section, Central Statistics Branch, Ministry of Finance and Corporate Relations, Government of British Columbia, June, 1993.

III. The 1998 Update (The 1996 Model)

A new regression equation was formulated using pooled 1991 and 1996 Census data as input and the cross-sectional regression model developed in 1986. In this update, household counts were projected for 89 sub-provincial areas instead of the 78 sub-provincial areas used in the previous version, because the Ministry of Health has adopted a new geographic scheme in which the province is divided into 89 analytical regions. Also, the projection period has been expanded to 2026.

Figure 1
The Regression Equation of the 1998 Update

ln **PPH = 0.65 + 0.38** *ln* **CPOP - 0.14** *ln* **MPOP - 0.22** *ln* **SDPOP**

Where	PPH	= Average number of persons per household
	CPOP	= Child (19 and under) Population
	MPOP	= Married Population
	SDPOP	= Separated and Divorced Population

Regression Statistics:

Normalised Parameters:		CPOP=3.8,	MPOP=-1.6,	SDPOP=-2.5
T-Value:	CONSTANT=24.5,	COP=27.9,	MPOP=-8.0,	SDPOP=-19.4
Std. Error:	CONSTANT=0.026,	CPOP=0.014,	MPOP=0.018,	SDPOP=0.012
F-Value:	445.8			
R-Square:	88.6			
R Bar-Square:	88.5			

Different combinations of pooled data from the four census periods were tested for use in the regression equation. The best overall results were obtained by pooling only the 1991 and 1996 Census data. Therefore, only 1991 and 1996 Census data were used for building the new regression equation for this update. Data from all four censuses were used to obtain inter-censal household estimates between the years 1986 and 1996.

The 1996 Census data used in this update were adjusted for net census undercount. A technical summary of the adjustment process is included in Appendix B of this paper.

V. Evaluation of the 1998 Updated Coefficients

A number of evaluations of the performance of the model were conducted subsequent to the first projection in 1986 and the various updates since that time. The evaluations examined the error characteristics of the projected/ backcasted results to actual sub-provincial household counts from different censuses. For a detailed description of model evaluation done prior to this update, refer to the "1994 Revision" version of this paper.

In this update, the backcasted results obtained from the new regression coefficients were tested against the actual census household counts of 1981, 1986, 1991 and 1996.

Table 4
Comparison of the 1996 Model (updated 1998)

	1981 Census	1986 Census	1991 Census	1996 Census
Average Absolute % Error	2.1	3.2	3.0	3.2
Std. Deviation of Error	1.6	4.1	2.3	3.1
Error at the Provincial Level	0.2	0.2	0.5	0.1
Min. % Absolute Error	0.0	0.06	0.06	0.01
Max. % Absolute Error	6.5	25	9.1	9.5
Count of Errors < 5%	73	65	62	74
Count of Errors Between 5 & 10%	5	11	16	15
Count of Errors Between 10 & 15%	0	0	0	0
Count of Errors > 15%	0	2	0	0
Count of Positive Errors	38	36	47	42
Count of Negative Errors	40	42	31	47

V. Conclusion

A number of advantages and disadvantages are associated with the British Columbia Small Area Household Projection model.

Advantages

- The historical data inputs required to formulate the household projection model by small area are minimal when compared to the traditional headship rate approach.
- The model itself produces an error structure that is consistent over time and within the range of acceptability.

Disadvantages

- A comparison of the "pure" and "overall" model error for 1991 and 1996 (Tables 1, 2 & 3) indicates that the use of actual data for the independent variables lowered the average absolute percent error from approximately 6.0 per cent to 2.5 per cent. Based on this result, it is anticipated that model accuracy would be significantly improved if more accurate and refined methodologies were available for projecting the independent variables used by the model.
- Currently, the model does not lend itself to projections of households by type, age of maintainer, families, family structure, etc., as do projections based on the maintainer rate method.

APPENDIX A

Married, Separated and Divorced Population Projections

I. Definitions

II. Method

I. Definitions

Marriage Rate, $M_{i,t}$ = Married Population $_{i,t}$ / Total 20^{+} Population $_{i,t}$

Separation plus Divorce Rate, $SD_{i,t}$ = (Separated + Divorced population) $_{i,t}$ / Married Population $_{i,t}$

Where **i = small area 1, 2, . . . , n.**
t = future years 1997, 1998, . . ., 2026

II. Method

The method for projecting married and divorced/separated population was changed in the 1991 Model. In the earlier Models, average annual compounded growth rates were calculated between the earlier censuses. These rates were then used for projecting the future married and divorced/separated populations for each small area.

The 1991 and 1996 Model employed a time-series regression method to project future marriage rates for the population aged 20 and over, and divorce/separation rates for the married population. Using the rates obtained from the 1981, 1986, 1991 and 1996 Censuses, for each small area a power regression equation was established in the form:

$$\ln(y) = a + b \times \ln(t)$$

Where **y** = marriage or divorce/separation rates at future year **t** (**t** takes on the values of 1997, 1998, . . . , 2026)

a and **b** are the regression coefficients obtained from regression analysis when **t** = 1981, 1986, 1991 and 1996, and **y** = marriage or divorce/separation rates for the respective years.

After the marriage and divorce/separation rates were obtained from the regression equation, married and divorced/separated populations were calculated according to the above definition of married population and divorced/separated population.

APPENDIX B

I. Adjustment for Net Census Undercount

II. Estimating Household Counts for 1976-1996

I. Adjustment for Net Census Undercount

Households: Adjustment Methodology

All input variables were adjusted for net census undercount. In the 1996 model, undercount was corrected by using the 1996 Census population undercount adjustment as a basis, and pro-rating the input variables accordingly by sub-provincial area.

II. Estimating Household Counts for 1976-1996

The household estimates between 1976 and 1990 have not changed with this update and the methodology is described below. The estimation of households between 1991 and 1996 was carried out using a similar method, except that the 1986, 1991 and 1996 census data were used to create the time-series regression equation.

The 1976-1991 household estimation method:

Three more regression equations were developed for 1976, 1981 and 1986 using only the 1976, 1981, 1986 adjusted data respectively. Then backcasting was performed for each of the four census years in order to calculate the error of closure and add-factoring. Estimates of independent variables for inter-censal years became necessary.

Married population, divorced/separated population and household population for inter-censal years were obtained using an interpolation technique based on four census points:

Two sets of interpolation results from using:

a) Three points from 1976, 1981 and 1986

b) Three points from 1981, 1986 and 1991

The combined results of 1976 from (a) and 1981 to 1991 from (b) formed the 1976 to 1991 estimates for the variables.

After obtaining the independent variables for 1977 to 1990, the 1976 coefficients were applied to obtain initial household counts for 1977 to 1980, 1981 coefficients for 1982 to 1995, and so on.

Initial household estimates were then adjusted by performing an error of closure together with add-factoring using the error obtained earlier from backcasting each of the four regression models from 1976, 1981, 1986 and 1991 (the 1991 model used the same pooled 1986/91 data as in the Projection).

In other words, 1977 to 1980 data were adjusted by an error of closure from 1981 backcasting followed by an add-factoring from 1976 backcasting.

References

1. Canada. Statistics Canada. **Interim Household and Family Projections for Canada, Provinces and Territories to 2006.** Ottawa. 1987.

2. F. Ip and D. McRae. "**British Columbia Small Area Household Projections - Methodology**", unpublished report. Planning and Statistics Division, Ministry of Finance and Corporate Relations, Government of B.C. 1986.

3. F. Ip. "**Small Area Household and Family Projection Update.**" unpublished report. BC STATS, Government of B.C. 1988.

4. U.S. Department of Economic and Social Affairs. **Methods of Projecting Households and Families**. New York. 1973.

5. F. Ip and D. McRae. "**Small Area Household Projections - A Parameterized Approach -**", unpublished report. Planning and Statistics Division, Ministry of Finance and Corporate Relations, Government of B.C. 1994.