

INDIRECT ESTIMATES OF AGE-SPECIFIC
INTERREGIONAL MIGRATION FLOWS IN
CANADA

K. Bruce Newbold
School of Geography and Geology
McMaster University, Hamilton ON

INCOMPLETE DRAFT

September, 2004

Acknowledgments: This paper was prepared for presentation at the 4th Colorado Conference on Migration, convened at the Aspen Lodge, Estes Park, Colorado, on September 24-26, 2004, and supported by the National Institute of Child Health and Development and by the National Institute of Aging.

INDIRECT ESTIMATES OF AGE-SPECIFIC INTERREGIONAL MIGRATION FLOWS IN CANADA

Introduction

Like the U.S. population, the Canadian one is highly mobile, a characteristic that is attributed to a variety of factors including peripatetic traditions inherited from immigrant forebears, an apparent lack of rootedness to place, and the relatively open nature of land and housing markets. Historically, internal movement was related to the opening of new frontiers and land (i.e., westward expansion), the sponsored settlement of interior regions, or the discovery of natural resources such as gold and the ensuing gold rushes. More recently, internal migration in Canada has reflected employment opportunities, particularly in Ontario and Alberta, the amenity inspired attraction of British Columbia on the west coast, the lack of employment opportunities in the Atlantic region, and the role of language as a modifier of migration flows between Quebec and the Rest of Canada. Still, the existing literature (i.e., Long and Boertlein 1976; Newbold 1997; Newbold and Liaw 1990) suggests a relative similarity between Canada and the United States with respect to internal migration trends and the reasons for internal migration, even after controlling for the number of options in the choice set. In fact, the literature goes so far as to suggest that the two countries share a similar point in the mobility transition (Zelinsky 1971).

Building upon earlier work (Rogers *et al.* 2003; Rogers and Jordan 2004), the following paper utilizes and evaluates three methodologies to infer age-specific, origin-destination specific migration flows. The intent is to provide a national case study, focusing upon the Canadian example using data drawn from the 1991 and 1996 Censuses.

Data and Methods

Migration data derived from the 1991 and 1996 Public Use Microdata Files (PUMF) produced by Statistics Canada, a sample derived from long-form questionnaires that represent 3 percent of the population. Migration is analyzed at the regional scale defined by a five-by-five matrix, with the regions defined by the Atlantic provinces, Quebec, Ontario, the Prairies, and British Columbia. The analysis excludes the foreign-born and those who did not report a place of residence. In addition, those resident in one of the three Canadian Territories or Prince Edward Island were excluded due to aggregation issues and sparse flows in-and out-of these regions. The methods drawn upon to indirectly estimate migration flows are based upon those described in Rogers *et al.* (2003) (see also Rogers and Jordan 2004).

Results (Preliminary)

Method 1: Estimates Based on Infant Migration Propensities

Rogers and Jordan (2004) utilized regression analysis to develop age-specific migration probabilities from infant migration data. Infant migration refers to those who are 0-4 years old at the time of the census and living outside their region of birth. Given observed regularities in the age profiles of migration, levels of migration in this youngest age group must be related to migration levels in other age groups.

Figure 1 plots the aggregate conditional survivorship proportion ($S_{ij}(+)$) against the corresponding first age-group-specific component of that aggregate proportion, $S_{ij}(-5)$. As in the

U.S. case, there is a close correspondence between the conditional survivorships proportions for infant migration and the survivorships for all other age groups, with the predicted, linear line offering a good summary of the relationship between these two variables. Since the initial analysis (not shown) revealed that the intercept term was statistically insignificant, the constant was set to zero for the simple and multiple cases. For those flows (i.e., $S_{ij}(-5)$) that were missing, the value was suppressed from the regression analysis. Unlike the U.S. experience, retirement peaks were not noted with the Canadian data, so no attempt to disaggregate flows based upon retirement migration experiences.

When the conditional survivorship proportions for infants are used to predict survivorships for other age categories using simple regression (Table 1), there is considerable variation in the goodness-of-fit. Overall, the adjusted R^2 exceeds 0.93 for the 1986-91 and 1991-96 intervals, and for both census periods combined. Corresponding t-statistics exceed 16.0 in all three cases. Age specific regression analyses are less robust, with declining model fit associated with increasing age. Slightly better results are noted when ${}_iK_j(+)\%$ is added to the regression (Table 2), with a modest increase in the adjusted R^2 . For example, the adjusted R^2 increases from 0.94 to 0.97 (simple and multiple regression models, respectively), holding the 1986-91 period constant. Similar results are noted for the age-specific models, with declining model fit with increasing age. Concurrently, the coefficient for ${}_iK_j(+)\%$ was not significant in several of the oldest age groups.

Table 3 presents the results of logistic regression analysis, therefore ensuring non-negative estimates for $S_{ij}(x)$ that range between 0 and 1. In general, the R^2 values are lower than observed earlier (i.e., 0.83 for 1986-91, all ages). As before, the adjusted R^2 values are greatest for the younger age groups: 0.72 for $x = 0$, and poorest for the older age groups: 0.52 for $x = 80$.

With the exception of the 65-69 year old model, the ${}_iK_j(+)$ % is insignificant at the $p < 0.05$ level. In several cases as well, the estimated coefficient for $S_{ij}(-5)$ was also insignificant, including the 55-59 and 65-79 age groups. Figure 3 illustrates the observed and predicted age patterns for flows from the Atlantic region to Ontario, and Ontario to Atlantic.

Method 2: Estimates Based on Age-Specific Net Migration Data

The second methodology, as described in Rogers *et al.* (2003) utilizes net migration flows. Methodologically, this method presented new challenges, including the treatment of the net migration measure ($N_i(x)$) in a multi-regional context. If, for example, origin-destination specific measures of net migration are used, the result is a computationally extensive set of regression analyses to determine the total predicted $S_{ij}(x)$, from which the other predicted age-specific values can be computed. In addition, the region that is used as the focus for the regression equation influences the outcome and leads to a different result if the other region is used instead. Alternative estimation routines, particularly enabling estimation within a multi-regional context, are required.

As an alternative and interim step, a series of bi-regional migration systems was considered. For example, migration between the Atlantic region and the 'Rest of Canada' (ROC), along the corresponding 'Rest of Canada' to the Atlantic, was evaluated. Similar migration flows for the remaining four regions were considered as well. While clearly not an ideal solution to this problem, in that it does not provide origin-specific flows *per se*, the results are preliminary. Moreover, migration flows out of the region of interest could be disaggregated to the destination regions based upon previous shares of inter-regional flows (not shown).

Table 4 reports the outcome of the current analysis using the bi-regional set of systems. Of the five regions included in the analysis, the best fit was obtained for the Atlantic region (see Figure 3). The application of the same model to other regions either resulted in poor model fit (as measured by the adjusted- R^2 value: 0.18 for the Prairies and 0.25 for Quebec) or poor association with the observed survivorships (Table 5, Figure 4). Model fit was improved considerably in some cases when the focus region was reversed, reflecting the problem with the current estimation methodology noted above. For instance, model fit (measured by the adjusted- R^2 value) improved from 0.61 to 0.83, with the estimated coefficients reflecting the observed values much better. Still, this was not universal, with model fit declining when estimated for Quebec (adjusted- $R^2 = 0.23$) and BC (0.78), and improving somewhat for the Prairies (0.32).

Method 3: Estimates Based on Two Successive Distributions of Population Stocks

The final evaluated method illustrated using Canadian data utilizes birthplace specific information to estimate migration flows. Illustrating the predicted primary migration flow between Ontario and British Columbia for 1991-96, Figure 5 is illustrative of the inferred results using this method. Without adjusting for missing or zero flows, the predicted survivorships frequently include negative values. Additionally, the age-profiles are frequently disjointed (as opposed to the smooth, typical age schedules of migration).

More than likely, the results obtained reflect a number of data-driven issues and problems. Specifically, a number of observations are missing, reflecting both the suppression of data by Statistics Canada for confidentiality reasons, as well as missing (non-sampled) flows, a problem that is particularly apparent in this method. Clearly, attempts to implement this

methodology at smaller spatial scales (states, provinces, metropolitan areas, etc.) would lead to increasingly sparse matrices and potentially large estimator biases.

Several solutions were tested to deal with these problems, including the substitution of arbitrarily small values (0.25 for the off-diagonal flows, and 1.0 for the diagonal flows) for the missing flows. However, this is a rather arbitrary solution, and one that does not have a substantive base for its implementation. More robust solutions include ‘borrowing’ the missing data from somewhere else, such as previous census periods, or estimating the missing values through some other means. Although borrowing data from elsewhere is a convenient and appropriate method, earlier census periods did not have the information needed to complete the flows, and typically suffering from the same data limitations noted previously. Essentially, when the data was disaggregated by (i) region of birth, (ii) age, (iii) region of residence at the start of the census interval, and (iv) region of residence at the end of the census interval, the flows do not exist for some regions. Nor are these just missing flows for a particular age group, but missing flows across all ages for specific origins (that is, for example, there are no Prairies to other region flows), the result of data suppression by Statistics Canada.

Alternatively, the conditional survivorships may be used. Average conditional survivorships, for example, could be applied. However, the average value tends to be much larger than those typically observed in older age groups, meaning an adjustment factor is required. The age-specific conditional survivorships generated in Method 1 (based upon the origin-destination flows) may be borrowed and used as substitutes for missing flows. Reasonably complete with a limited number of zero flows, their substitution increases the total conditional survivorship, such that the sum is greater than 1.0, requiring the re-scaling of individual

conditional survivorships such that they sum to 1.0 before proceeding with the model application.

Unfortunately, the estimation of migration flows was little improved following the substitution of the age-specific survivorships. While substantially decreasing the number of missing or zero cells, some flows retained a zero value, reflecting the lack of data even when region of birth was not controlled for (Method 1). Additionally, some of the estimated $S_{ij}(x)$ retained negative signs, even after adjustment of the input data. Most likely, the poor results reflect relatively sparse inter-regional flows when controlling for birthplace. Such sparse flows may over- or under-represent actual observed flows between regions.

Conclusions

Given observed regularities in the age-migration schedule, it is possible to infer origin-destination specific migration flows, with three alternative methods tested in this paper using data from the 1991 and 1996 Canadian Censuses. While the results presented here are preliminary, the application of the three methodologies raises two significant issues. First, missing data, zero flows, and sparse flows were problematic in the application of the models. Although minimal in the context of Methods 1 and 2, the issue was significant in Method 3, where the data is disaggregated by birthplace, age, and region of residence at the beginning and end of the Census period. Similarly, sparse inter-regional flows likely result in biased estimates, creating the potential to over- or underestimate the importance of particular regional linkages. Most likely, the presence of sparse flows was responsible for the results generated in Method 3. More serious problems would exist as the scale of resolution is increased. Although a number of

alternatives to missing flows were proposed and tested in the current analysis, the adjustments did not appear to impact upon the validity of the results. Additional work is required to identify appropriate estimation tools and routines in these situations.

Second, the estimation results derived from Method 2 are dependent upon the focus region. Even when the five-region system was reduced to a two-region system, the outcome of the regression analysis depended on whether the 'Rest of Canada' or the specific region was chosen, with no clear rule to determine the setup of the regression equations.

References

Long L, and Boertlein CG. 1976. *The geographical mobility of Americans: An international Comparison*. Current Population Reports Series P-23, no. 64. Washington DC: U.S. Department of Commerce, Bureau of the Census.

Newbold KB. 1997. Primary, return and onward migration in the U.S. and Canada: Is there a difference? *Papers in Regional Science* 76(2): 175-198.

Newbold KB, and Liaw KL. 1990. Characterization of primary, return and onward interprovincial migration in Canada: Overall and age-specific patterns. *Canadian Journal of Regional Science* 13: 17-34.

Rogers A, Raymer J, and Jordan L. 2003. Inferring migration flows from birthplace-specific population stocks. Population Program, Institute of Behavioral Science, University of Colorado, WP-03-02.

Rogers A, and Jordan L. 2004. Estimating migration flows from birthplace-specific population stocks of infants. *Geographical Analysis* 36(1): 38-53.

Zelinsky W. 1971. The hypotheses of the mobility transition. *Geographical Review* 41:219-249.

Table 1. Total and Age-Specific Simple Regression Models, Canada: 1986-91 and 1991-96

(Constant=0)

	Sl($S_{ij}(-5)$)	t	Adj- R^2	N
Year (all ages):				
1986-91	16.01	16.6	0.94	18
1991-96	14.89	17.8	0.95	18
1986-96	15.50	24.2	0.94	36
Age (pooled):				
0	1.10	18.4	0.90	36
5	0.90	20.4	0.92	36
10	0.83	19.6	0.91	36
15	1.70	18.9	0.91	36
20	2.13	23.2	0.94	36
25	1.55	23.1	0.94	36
30	1.17	18.3	0.90	36
35	0.91	17.0	0.89	36
40	0.71	17.1	0.89	36
45	0.58	16.2	0.88	36
50	0.55	11.8	0.79	36
55	0.49	9.9	0.73	36
60	0.47	12.4	0.81	36
65	0.39	11.0	0.77	36
70	0.33	10.8	0.76	36
75	0.30	11.0	0.77	36
80	0.41	9.0	0.69	36

Table 2. Total and Age-Specific Multiple Regression Models, Canada: 1986-91 and 1991-96 (Constant=0)

	Sl($S_{ij}(-5)$)	t	Sl(K_j)	t	Adj- R^2	N
Year (all ages):						
1986-91	5.70	2.1	2.27	3.9	0.97	18
1991-96	9.76	4.1	1.07	2.3	0.96	18
1986-96	8.42	4.6	1.52	4.0	0.96	36
Age (pooled):						
0	0.57	3.1	0.11	3.0	0.92	36
5	0.40	3.2	0.11	4.1	0.95	36
10	0.46	3.4	0.08	3.0	0.93	36
15	1.00	3.5	0.15	2.5	0.92	36
20	1.16	4.3	0.21	3.8	0.95	36
25	0.86	4.3	0.15	3.6	0.95	36
30	0.65	3.2	0.11	2.6	0.92	36
35	0.40	2.4	0.11	3.3	0.91	36
40	0.41	3.0	0.06	2.3	0.90	36
45	0.26	2.3	0.07	3.0	0.90	36
50	0.15	1.0	0.09	2.8	0.83	36
55	0.08	0.5	0.09	2.7	0.77	36
60	0.15	1.2	0.07	2.8	0.84	36
65	0.18	1.5	0.05	1.9	0.79	36
70	0.11	1.1	0.05	2.3	0.79	36
75	0.22	2.3	0.02	0.9	0.77	36
80	0.37	2.3	0.01	0.3	0.68	36

Table 3. Total and Age-Specific Log-Odds Regression Models, Canada: 1986-91 and 1991-96 (Constant=0)

	Int.	<i>t</i>	SI($S_{ij}(-5)$)	<i>t</i>	SI((K_j))	<i>t</i>	Adj- R^2	N
Year (all ages):								
1986-91	-3.35	-19.3	92.37	2.4	8.17	1.1	0.83	17
1991-96	-3.35	-18.8	88.89	2.5	8.14	1.3	0.79	17
1986-96	-3.35	-28.3	91.55	3.7	8.02	1.7	0.82	35
Age (pooled):								
0	-5.95	-45.5	66.4	2.5	8.49	1.7	0.72	35
5	-6.29	-48.9	71.0	2.7	9.00	1.8	0.76	35
10	-6.54	-44.2	85.9	2.8	7.92	1.4	0.73	35
15	-5.70	-44.4	79.0	3.0	7.96	1.6	0.77	35
20	-5.22	-46.2	70.3	3.0	7.40	1.7	0.77	35
25	-5.54	-46.2	83.8	3.4	4.68	1.0	0.75	35
30	-5.92	-40.0	88.4	2.9	4.67	0.8	0.68	35
35	-6.25	-39.1	85.6	2.6	6.13	1.0	0.66	35
40	-6.53	-40.7	98.4	3.0	3.53	0.6	0.66	35
45	-6.65	-38.0	76.2	2.1	7.00	1.0	0.60	35
50	-6.83	-43.7	74.2	2.3	7.97	1.3	0.67	35
55	-7.01	-35.6	74.2	1.9	8.03	1.1	0.57	34
60	-6.92	-38.2	73.8	2.0	7.85	1.1	0.61	33
65	-7.38	-33.5	29.6	0.7	18.62	2.2	0.56	34
70	-7.17	-37.4	36.8	1.0	13.34	1.9	0.56	33
75	-6.85	-38.7	56.6	1.7	4.87	0.8	0.51	30
80	-6.78	-37.8	72.0	2.1	3.39	0.5	0.52	32

Table 4. Multiple Regression Results: Model 2, Canada

	$K_j(x)/$		$r_{ij}(x)$		Adj- R^2	N
	$K_i(x)*r_{ij}(x)$					
	(X1)	t	(X2)	t		
Atlantic – ROC	0.043	8.2	0.595	13.3	0.87	17
Quebec – ROC	0.019	1.5	0.096	2.7	0.25	17
Ontario – ROC	0.032	1.4	0.165	3.6	0.61	17
Prairies - ROC	0.156	2.7	0.023	1.7	0.18	17
BC - ROC	0.002	0.1	0.679	3.0	0.80	17

Table 5. Observed and Estimated S_{ij} : Model 2

	Observed	Estimated
Atlantic – ROC	0.5707	0.5952
ROC – Atlantic	0.0416	0.0428
Quebec – ROC	0.1871	0.0969
ROC – Quebec	0.0539	0.0194
Ontario – ROC	0.3798	0.1654
ROC – Ontario	0.1438	0.0314
Prairies – ROC	0.5600	0.0232
ROC – Prairies	0.1088	0.1555
BC – ROC	0.5099	0.6793
ROC – BC	0.1459	0.0019

Figure 2. Method 1: Observed and predicted age profiles for selected migration streams, 1991-96

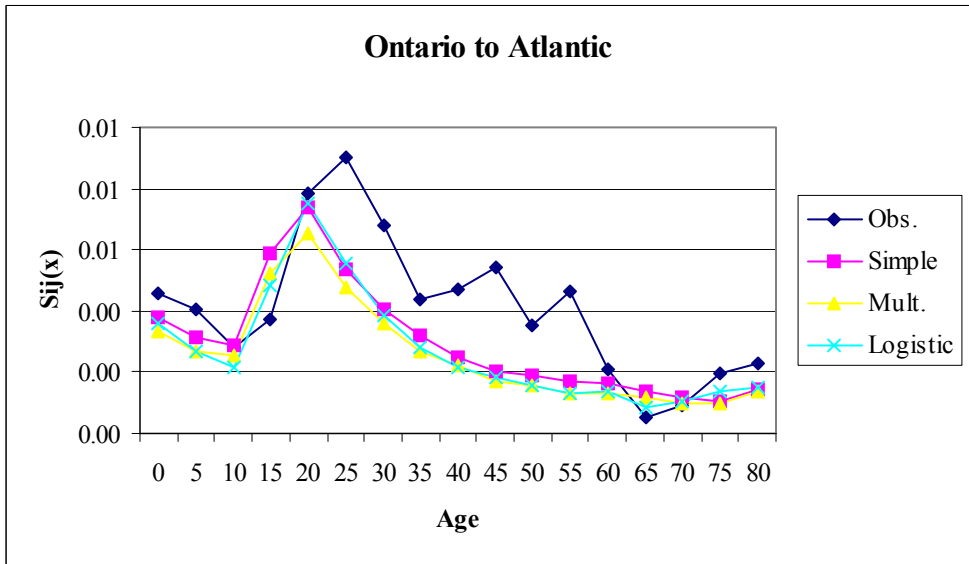
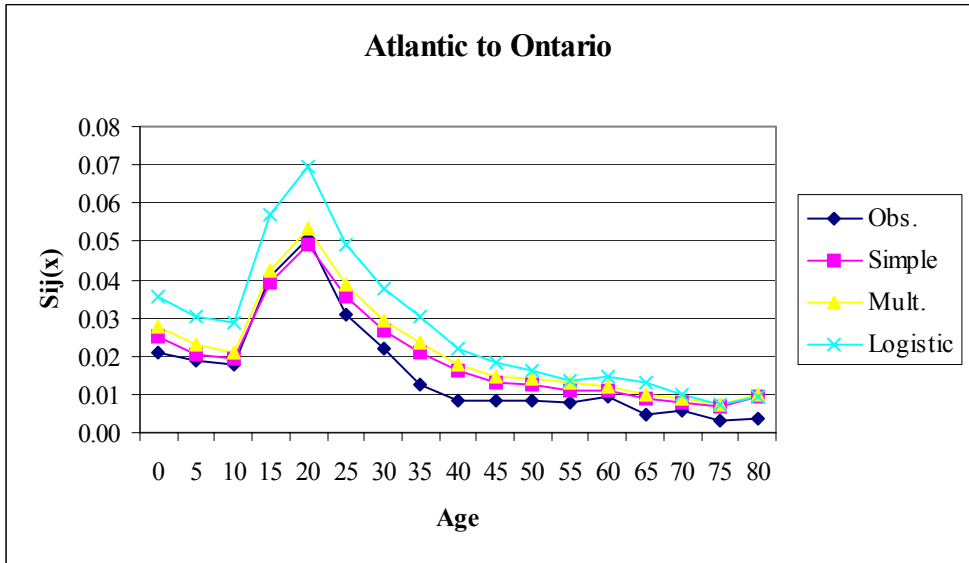


Figure 3. Method 2: Observed and predicted migration age profiles, Atlantic Region and ROC, 1991-96: Derived by Method 2.

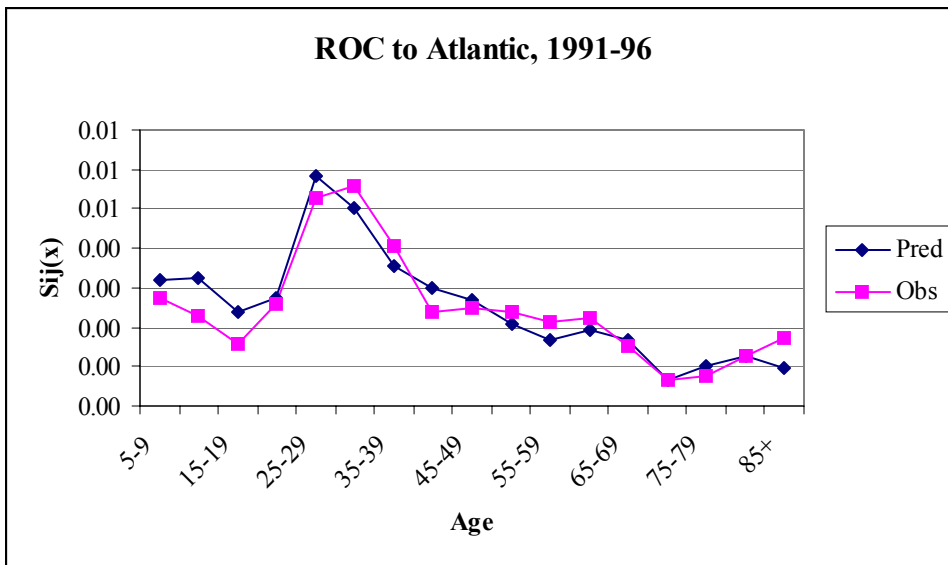
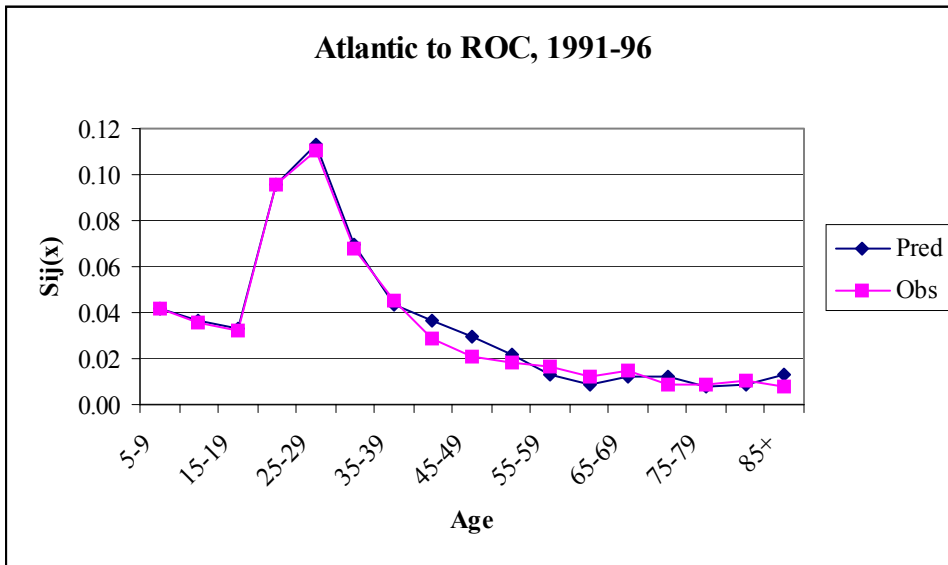


Figure 4. Method 2: Observed and predicted migration age profiles, Prairie Region and ROC, 1991-96:

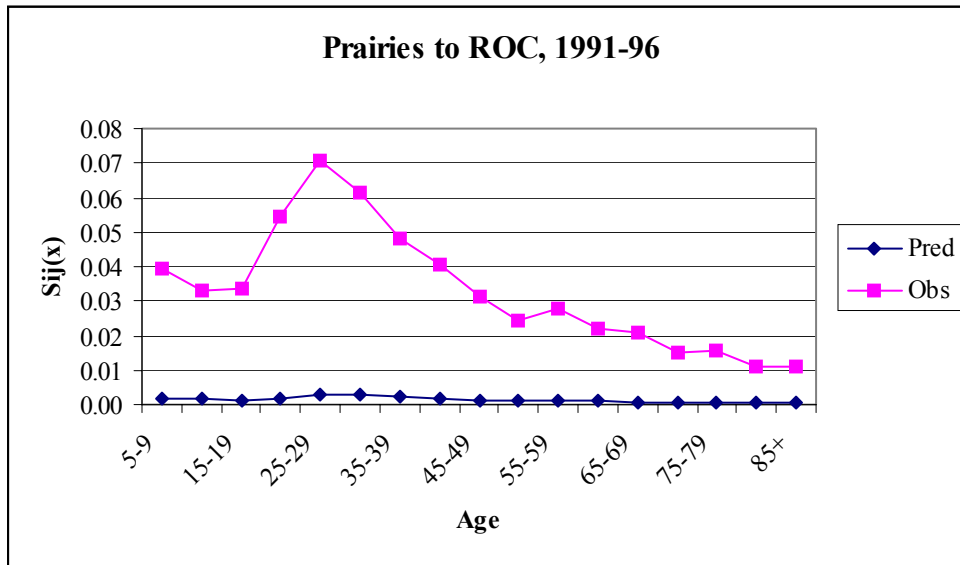


Figure 5. Method 3. Observed and predicted migration age profiles for Primary migration between Ontario and BC, 1991-96.

