UNIVERSITY OF WATERLOO
Faculty of Economics

A Comparison of Real and Simulated Immigrants:
Data from the Census Compared to the LifePaths Microsimulation Model

Statistics Canada
Ottawa, Ontario

Prepared by:

Jon MacKay
91067850
Masters of Economics
March, 2008

Table of Contents

Abstract....3

Introduction....4

Overview of the LifePaths Simulation Model and Modgen 5

Analysis: Comparison of Survey Data with LifePaths Generated Data....9

Calculating the Employment-to-Population Ratio....11

Simulating Employment Events in LifePaths....15

Comparing Results: Employment-to-Population Ratios in the Census and LifePaths....17

Using Standardization to Compare Employment-to-Population Ratios....22

Conclusions and Recommendations....25

Appendix A: Method Used to Remove Students from Census, LFS and LifePaths Tables....27

Appendix B: Complete Listing of Modgen File Immigrant.mpp....29

References....35

Illustration Index

Illustration 1: Simulated persons and families.....5

Illustration 2: Graphical representation of overlapping birth cohorts.....7

Illustration 3: Life cycle of a model - MPP file to Windows executable.....7

Illustration 4: Screen grab of typical model user interface.....8

Illustration 5: E/P ratios of males with less than a university degree, born in Canada (BIC) and immigrants. 17

Illustration 6: E/P ratios of females with less than a university degree, born in Canada (BIC) and immigrants.....18

Illustration 7: E/P ratios of those males with a university degree or greater, born in Canada (BIC) and immigrants.....20

Illustration 8: E/P ratios of females with a university degree or greater, born in Canada (BIC) and immigrants. 21

Index of Tables

Table 1: Population E/P ratios of males and females for Canadian-born, immigrants and standardized immigrants.....23

Table 2: Population E/P ratios of males and females for Canadian-born and immigrants who have lived in Canada for less than 5 years, and immigrants who have lived in Canada for more than 5 years.....23

A Comparison of Real and Simulated Immigrants:
Data from the Census Compared to Results Generated in the LifePaths Microsimulation Model

Abstract

This report is concerned with a limited evaluation of the data generated by Statistics Canada's longitudinal microsimulation model called LifePaths. In particular, this report is concerned with the treatment of the immigrant population within the LifePaths model as compared to what is known about the immigrant population in Canada based on the Census and Labour Force Survey (LFS). Currently, there are very few behavioral equations of the LifePaths model that specifically take into account the circumstances particular to immigrants. The LifePaths model generally does an excellent job of simulating the employment-to-population ratios of the Canadian-born population. However, a comparison of employment-to-population ratios for actual and simulated immigrants shows large differences between the LifePaths model and the Census. Based on this observed difference, this report outlines the use of a simple standardization technique to indicate whether existing variables in LifePaths are sufficient to accurately simulate the employment-to-population ratios of immigrants within Canada. While the current variables appear to be adequate for immigrant males, the standardization technique used indicates that more information is necessary to adequately model employment-to-population ratios of female immigrants. This report concludes by briefly examining a new variable, the number of years a person has been in Canada, which may be useful for improving LifePath simulations. Employment-to-population ratios of immigrant cohorts who have been in Canada less than five years, versus five years or more, are notably lower. It is suggested that taking a variable into account that measures length of time in Canada since immigration, improves the LifePaths model.

Introduction

Microsimulation models simulate representative samples of individual actors in order to draw conclusions about larger populations[1]. LifePaths, for example, simulates the economic life of individual Canadians in order to draw broader conclusions about the Canadian population as a whole. The developers of LifePaths have estimated behavioral equations based on the evaluation of micro-data drawn from sources such as the Labour Force Survey (LFS).

LifePaths is one of the microsimulation models produced by Statistics Canada to assist other government departments in developing and analyzing policy. In order to remain relevant, the model is reviewed often to ensure accuracy by comparing the results of the LifePaths simulated population, to what is known about the actual population from other sources, such as the Census and the LFS.

The LifePaths model relies in large part upon the LFS as a source for micro-data about Canadian individuals and families. Until 2006, the LFS did not contain questions identifying immigrants. This additional information in the LFS means that LifePaths developers can look forward to another source of longitudinal data about immigrants. This report uses a simple measure, employment-to-population ratios (E/P ratios), in order to compare the simulated population to the actual Canadian population which has been broken down into two groups: immigrants and Canadian-born persons. After comparing the real population to the simulated results, this report will explain why the set of variables currently used in LifePaths appears to be insufficient to accurately simulate the immigrant population. The results of a direct statistical standardization indicate that additional variables would be useful to model female immigrants. This report examines an additional variable for immigrants, measuring the number of years in Canada since their immigration, which is suggested as a useful starting point for further evaluation and possible expansion of the variables used in LifePaths.

1 Microsimulation Models homepage http://dissemination.statcan.ca/english/spsd/

Overview of the LifePaths Simulation Model and Modgen

LifePaths[2] is a case-based longitudinal microsimulation model. This means that each economic person (or "life path") is individually simulated from birth through to death. An individual's lifetime is represented as a series of events, such as: birth, level of schooling, employment status, marriage, child birth and death. In LifePaths, the probability of a person having basic characteristics such as gender and province of birth are determined using known demographic data. More complex characteristics of an individual's life, such as marriage and employment transitions, are determined using hazard functions which are estimated using longitudinal data[3] from a relevant survey, such as the Labour Force Survey (LFS). The LifePaths model aggregates each of these simulated individuals in order to generate the population of interest. A robust reporting utility allows model developers to design report tables

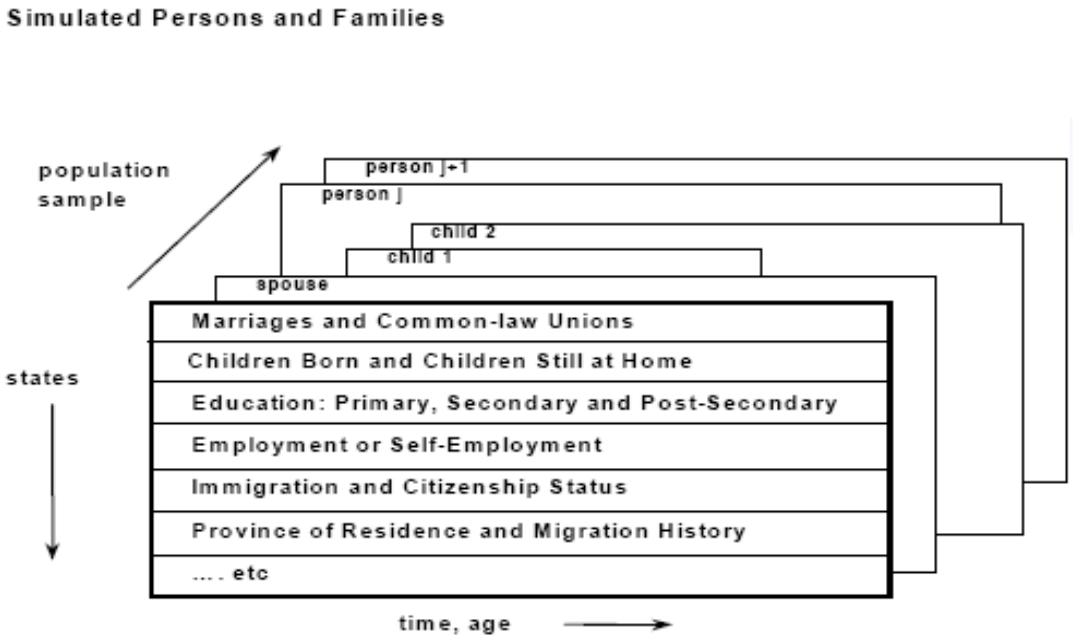

Illustration 1: Simulated persons and families – Illustrated example of case-based model and some relevant variables.
Source: The LifePaths Microsimulation Model - An Overview, p5

2 The homepage of LifePaths is http://dissemination.statcan.ca/english/spsd/LifePaths.htm
3 It is also possible to write non-empirical models using Modgen. See XEcon – Experimental Economy Theoretical Growth Model http://www.statcan.ca/english/spsd/XEcon.htm

which capture particular variables of interest.

A case-based model is designed to simulate an actor (such as a person) that has experiences over time. Various experiences are modeled using events. These events can be set to occur by the developer at particular times, such as each birthday, or only when a set of conditions is fulfilled. For example, there may be a given probability, or hazard, that someone experiences a divorce. A prerequisite for the divorce event is that the person actor is currently married. Waiting times for events often also have a stochastic element. The Monte Carlo variation of the model ensures that the simulated population generally has the same diversity observed in the actual population (LifePaths Overview, p5).

In some software that is used for creating simulation models, events take place only after discrete units of time – a month, or a year, for example. However, under the discrete time constraint the model builder often cannot know when one event has occurred relative to other events. For example, it is possible for a simulated individual to be married, have children and be divorced within a discrete modeling period. The problem is that the ordering of events is often of interest to researchers. Using discrete time, to continue with the example, the researcher cannot know if the children arrived before the marriage, during the marriage, or after the divorce. Modgen avoids this confusion by simulating continuous time, so that events can take place at any point in time, which allows researchers to know event ordering with certainty.

An additional strength of LifePaths involves the way in which cohorts of people born in the same year are managed. LifePaths has an overlapping cohort model, so that the model can generate the entire population range from newborns to the elderly. Historical data is directly comparable to LifePaths data from 1971 onwards (Overview, p6).

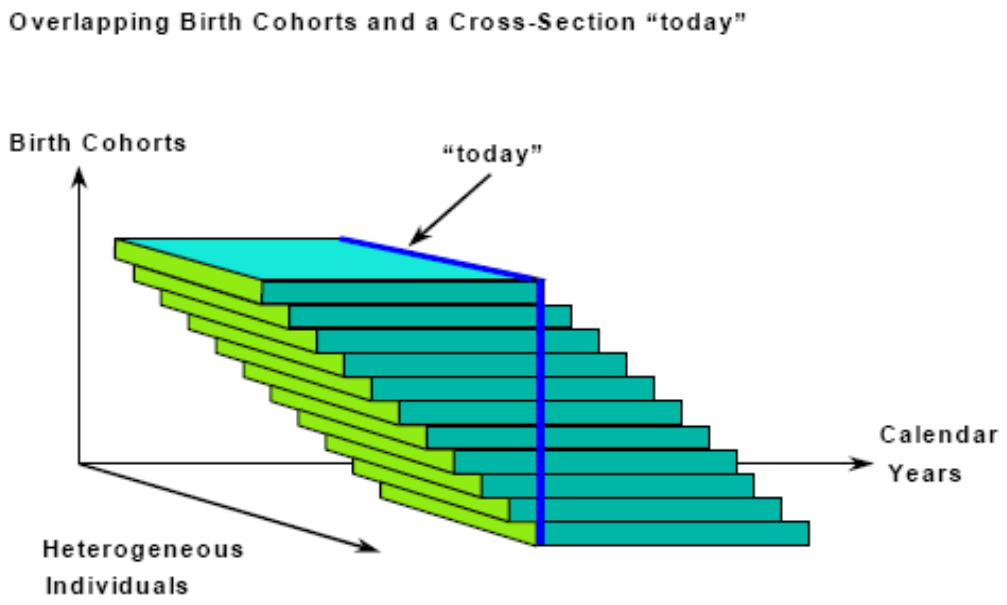

Illustration 2: Graphical representation of overlapping birth cohorts.
Source: The LifePaths Microsimulation Model - An Overview, p6

The LifePaths model is written using Modgen[4] which is short for model generator. Modgen refers to both a modeling language, which extends C++, and a tool to validate and pre-process model class files. The Modgen suite of tools is designed to allow users to easily create robust microsimulation models. Modgen class files (also referred to as modules) end with the extension ".mpp", and are preprocessed by a Windows executable called modgen.exe in order to generate ".cpp" files which can be compiled by Microsoft Visual Studio into a stand-alone Windows executable program. The Modgen suite is closely integrated into Microsoft Visual Studio which allows model developers to write and

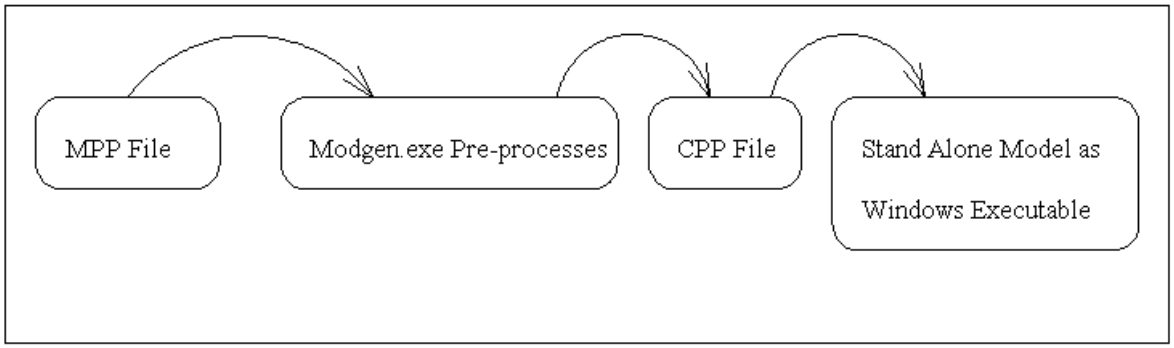

Illustration 3: Life cycle of a model - MPP file to Windows executable

4 The homepage of Modgen is http://dissemination.statcan.ca/english/spsd/Modgen.htm

compile their own models using a single integrated development environment (IDE).

Illustration 4 shows a screen capture of the graphical user interface (GUI) of a typical model generated using the Modgen suite. Model developers can allow users the freedom to alter key variables from this interface in order to test an economic hypothesis. The advantage of this approach is that a completed model can be sent to other colleagues who can easily alter the model for economic research through its GUI. These colleagues do not need to know how to program, or even have the Modgen tools installed. The model executable is a completely independent Windows application.

Simulated table output illustrating the distribution of weeks of paid work

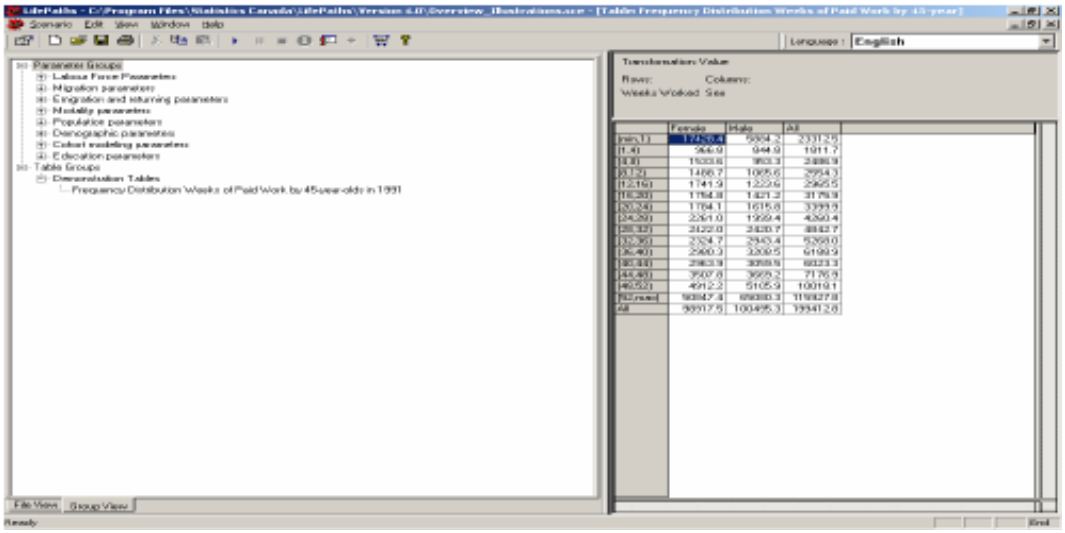

Illustration 4: Screen grab of typical model user interface. The right pane holds model generated data in a table that is defined by the model developer.
Source: The LifePaths Microsimulation Model - An Overview, p2

Analysis: Comparison of Survey Data with LifePaths Generated Data

LifePaths has been an active project at Statistics Canada for over a decade. Over that period, the model has become more robust as new modules have been added and updated in order to better simulate the actual Canadian population. LifePath developers have used the Labour Force Survey, as well as other data sources, as a source of longitudinal data for the model (Rowe and Huan, 2004).

As mentioned, this report focuses on career and work experiences of an individual. A number of variables are of interest in comparing the immigrant experience as captured in the Census versus the results from the LifePaths microsimulation model. Some of the important variables that are currently used by LifePaths in its career and work module (CareerWork), and also available in the surveys mentioned, include:

- Age of individual
- Gender
- Marital status
- Presence of children
- Educational attainment
- Province of residence
- Duration of time spent working

This brief study is limited to relatively basic economic variables that compose an individual's experience in LifePaths. For purposes of this report more complex variables in LifePaths, such as earnings, are ignored because the functions that determine an individual's earnings are largely driven by more basic underlying variables such as education and gender.

This report is mainly concerned with the career and work aspects of the LifePaths simulation model as it pertains to immigrants and the portion of the population born in Canada. In order to

examine this, it was decided to compare the simulated and actual employment-to-population ratios. Results taken from a large LifePaths simulation have been compared against results from both the Census and the LFS. The most recent Census data available was from 1996 and 2001, while the LFS only recently included questions identifying immigrants from 2006 onwards. While comparisons were made to each of these sources, this report will only show detailed results from the 2001 Census.

The Census, while an in-depth look at the Canadian population, is not a longitudinal survey. This makes comparison directly with the LifePaths model somewhat challenging. In order to validly compare the different surveys with LifePaths, it is necessary to ensure that the data is conceptually similar. An additional challenge is that the simulated population in the LifePaths model does not have the same level of detail that exists in the Census. For example, LifePaths currently does not include variables capturing details regarding an immigrant's country of origin, years since immigration, credentials, or language ability. For the purposes of the LifePaths model, immigrants are simply people that are “from away”.

Calculating the Employment-to-Population Ratio

One simple measure of employment for each subset of the Canadian population is the employment-to-population ratio. Showing the employment-to-population ratio (E/P ratio) across ages is one approach to graphically examining differences between the LifePaths generated populations and the actual population as captured by the Census.

To capture information such as the E/P ratio, LifePaths has a complex data output facility which allows users to create tables of results. Users specify table definitions within Modgen, and the tables are generated capturing the desired results. In order to capture the necessary data for this comparison of E/P ratios the following table definition was used in LifePaths:

```
table Person IMM_new02  // 2) 2001 Population
[ dominant && resident && year==2001
      && curtate_age >= 15
      && institutional_status == COMMUNITY
]
{
      isBIC // Is person Born In Canada?
      *
      split( children_at_home, CHILDREN_GROUPS) // Number of children
      *
      split( curtate_age, CAT_AGE ) // Age of individual, split by single years
      *
      sex // Boolean, 1 == Yes, 0 == No
      *
      split( weeks_worked_in_2000, quarter_year ) // Weeks worked in 2000
      *
      my_edu_level // Level of Education 1==less than HS, 2==HS|Trades, 3==BA+
      *
      {
            duration(employed, TRUE)/duration(),  //EN decimals=4 epr
            duration(employed, TRUE),              //EN decimals=4 dur_emp
            duration()                             //EN decimals=4 dur
      }
};
```

Text 1: LifePaths table declaration to generate employment-to-population ratios.
Source file: Immigrant.mpp, Source code by author.

The simple table definition above captures all of the necessary information (and more) for this portion of the project. The first line in the table is the name of the table followed by a filter statement in

square brackets ([]). The filter statement defines the characteristics of the individuals that are to be added to the table that is being defined. In this case, it is specified that the table can only contain "dominant" actors in the model that are: residents in Canada, in the year 2001, whose age must be greater than or equal to 15, and who must be in the general community (i.e., not in a long term care facility, for example).

While the other variables are self-explanatory, the `dominant` variable needs some explanation. Because LifePaths is a case-based model, it has dominant actors and non-dominant actors. Dominant actors are those individuals that are the focus of a given case. Non-dominant actors are those that are generated in order to fill-out the experiences of the dominant individual, such as spouses and children. The only purpose of non-dominant actors is to supplement the simulated life of the dominant actor. This characteristic of the model means that LifePaths is an open model (Overview, p5). Closed models, by contrast, usually have to simulate the entire population, and only then can the time be incremented. The closed model means that it cannot have any more accuracy than the data source from which it is derived. The open model nature of LifePaths means that if the Monte Carlo variation is too high for a certain subset of the population, the number of persons being simulated needs only to be increased in order to reduce variation.

The variables listed within the curly braces and separated by asterisks (*) are variables drawn from the existing LifePaths implementation. In order of appearance, these are:

- A boolean variable, to determine if person born in Canada (BIC), or an immigrant
- The number of (non-dominant) children of the (dominant) individual
- The age of the dominant individual
- The gender of the dominant individual
- The number of weeks worked in the year 2000
- The level of education attained by the dominant individual

The **duration()** call is the length of time an individual spends in the appropriate cell of the table. Thus, for example, if a person has a secondary school education but graduates from university half-way through the year in question, the duration call will add 0.5 to the total of secondary school, and 0.5 to university graduate.

The function **duration(employed, TRUE)** captures the amount of time an individual exists with the boolean variable **employed** in the state **TRUE**. So, the ratio of :

$$\frac{\text{duration(employed, TRUE)}}{\text{duration()}} \equiv \text{Employment-to-Population Ratio} \tag{1}$$

A variable from the Census which captures weeks worked in the past year has been used to construct a variable that would be comparable to the above ratio in the LifePaths table. The result is the proportion of weeks worked (PWRK). This variable is constructed using the weeks worked variable given in the Census divided by 52. This yields proportion of weeks worked.

$$PWRK = \frac{\text{Weeks Worked}}{52} \quad \text{where} \quad \sum_{i=1}^{n} (PWRK_i \cdot weight_i) \equiv \text{Total Duration Employed} \tag{2}$$

where *weight* is the weighting factor applied in the Census for a particular observation. Thus the employment-to-population ratio which is comparable to that developed in LifePaths would be:

$$\frac{\sum_{i=1}^{n} (PWRK_i \cdot weight_i)}{\sum_{i=1}^{n} weight_i} \equiv \frac{\text{Duration Employed}}{\text{Total population}} \equiv \text{Employment-to-Population Ratio} \tag{3}$$

This is equivalent to the probability that an individual might have been working at a point.

Similarly, the Labour Force Survey data can also be used to create a comparable employment-to-population ratio. Since the LFS is a monthly survey, there will be seasonal effects, and the annual

employment-to-population ratio should be examined. People are considered employed if the labour force status variable reveals that they are (a) currently employed, or (b) employed, but not currently at work. In order to remove the seasonal effects, the LFS data is collapsed and summed into months and then averaged over the year. This will remove seasonal effects from the data and the end result is comparable to the constructed LifePaths. This report will only focus on the Census data, however, because the sample size is much larger.

Simulating Employment Events in LifePaths

This report is concerned with the simulation of employment of immigrants, and the underlying variables in LifePaths that affect their employment. In order to determine whether the existing variables in LifePaths are sufficient to explain levels of employment, a graphical comparison of employment-to-population ratios has been used. In the cases that follow, the Census data will be used because it is a larger sample and so the resulting curves of the E/P ratio are smoother.

There are, however, other outstanding issues that require some mention regarding the simulation of life events in LifePaths. Employment events depend strongly upon other factors such as age and education level. The boolean `employed` used in the LifePaths table to determine the employment-to-population ratio can be altered in any of five different modules: `CareerWork.mpp`, `Disability.mpp`, `MaternityLeave.mpp`, `SavingsRPP.mpp` and `StudentWork.mpp`.

Employment of students is simulated in the `StudentWork.mpp` module. This module handles events such as student part-time employment, and first job upon graduation. This is important to note because there is thought to be significant under-reporting of employment by students in the Census. When asked if they are employed, students will often report that they are not despite holding a part-time job of some sort. As a result, the simulated employment-to-population ratio differs from that observed in the Census when students are included[5].

In order to ensure that the difference observed between LifePaths and the Census for young people is due to this discrepancy, students have been removed from both sets of data and the employment-to-population ratios have been compared. The results seem to indicate that students are the main cause of the difference between LifePaths and the Census (see Appendix). However, this report is

5 Personal communication with Geoff Rowe, senior analyst in Social Economics Analysis and Modeling Division (SEAMD), Statistics Canada, 2008.

broadly concerned with employment of adults, so rather than focusing on the StudentWork module in LifePaths, this report will focus on CareerWork, which covers the ages from the first job until the onset of retirement. The issue of student work is raised merely to illustrate the complexities of the simulation process, as well as some difficulties that exist in verification against authoritative data such as the Census and LFS.

Retirement also leads to different problems of comparison because of the complexities of people's decisions regarding when to leave work. Additionally, an aging population has to deal with increased hazard of illness and death which also complicates the analysis. Issues of retirement are handled in other LifePaths modules in addition to CareerWork. Rather than attempting to examine both the youth employment and patterns of employment closer to retirement, this report will simply ignore these age groups for the purposes of analysis and focus on the ages in between – roughly the early thirties to mid-fifties age range.

Comparing Results: Employment-to-Population Ratios in the Census and LifePaths

The graphs that follow detail the employment-to-population ratios (labeled EPR in the graphs that follow) of immigrants and those born in Canada (BIC). The population is further subdivided by level of education. The first educational category is whether one has attained an undergraduate university degree or greater. The other category includes those with an educational attainment less than an undergraduate degree, including college diplomas, high school, or less.

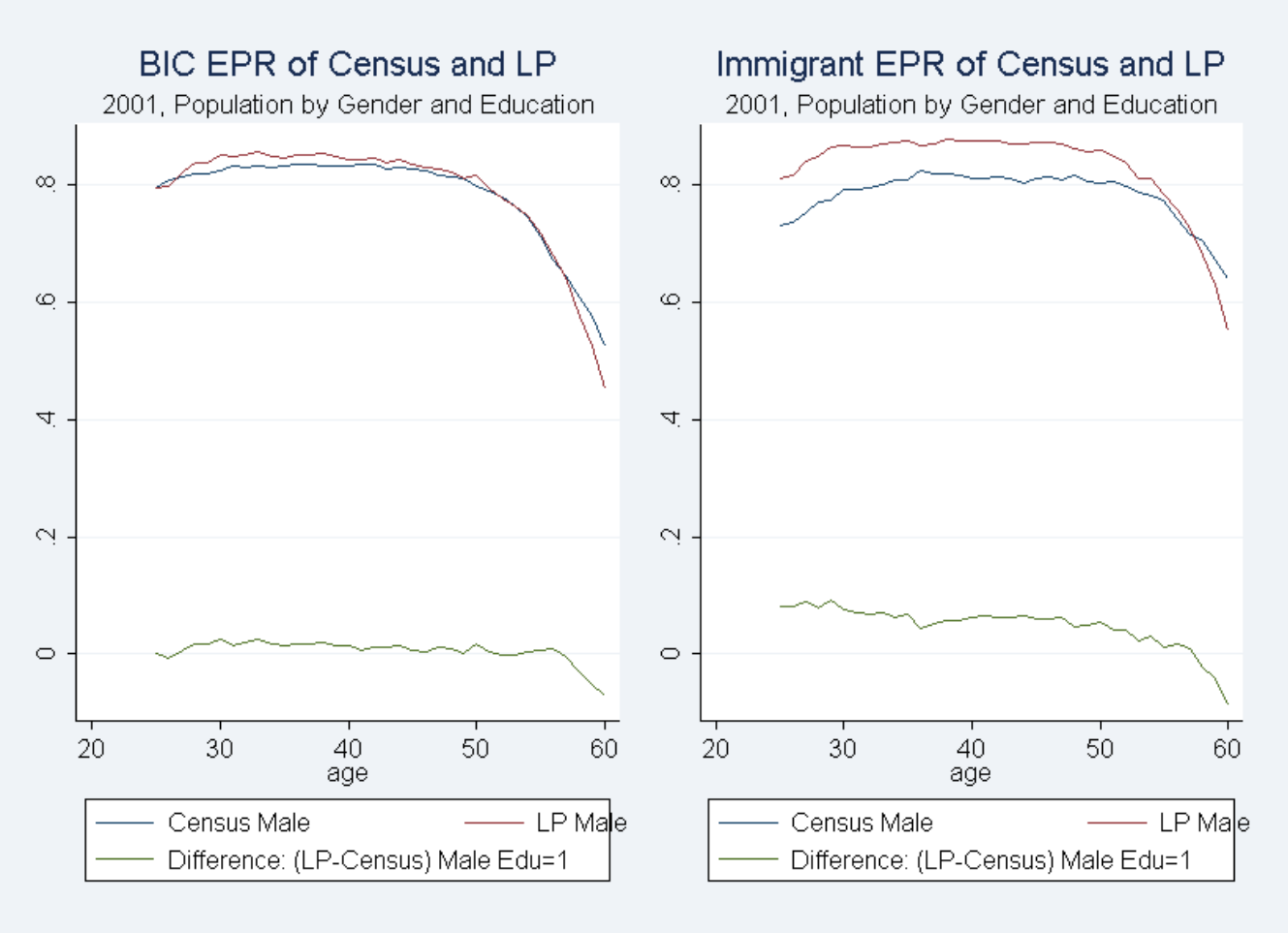

Illustration 5: E/P ratios of males with less than a university degree, born in Canada (BIC) and immigrants.

The employment-to-population ratio (EPR) shown above (Illustration 5) clearly show that LifePaths (LP) seems to be doing a good job simulating the employment situation of males born in

Canada (BIC) who have less than a university degree. The blue line represents the employment-to-population ratio of males at ages between 25 and 60 years. The red line is the employment-to-population ratio simulated by the LifePaths model. The green line is the difference between the actual and simulated employment-to-population ratios. For the portion of the population that was born in Canada, it is clear that the simulated model is very close to the actual ratio until the late 50s age range. Again, this upper bound is beyond the scope of this report because it involves other modules that represent complex issues such as disability, retirement and death.

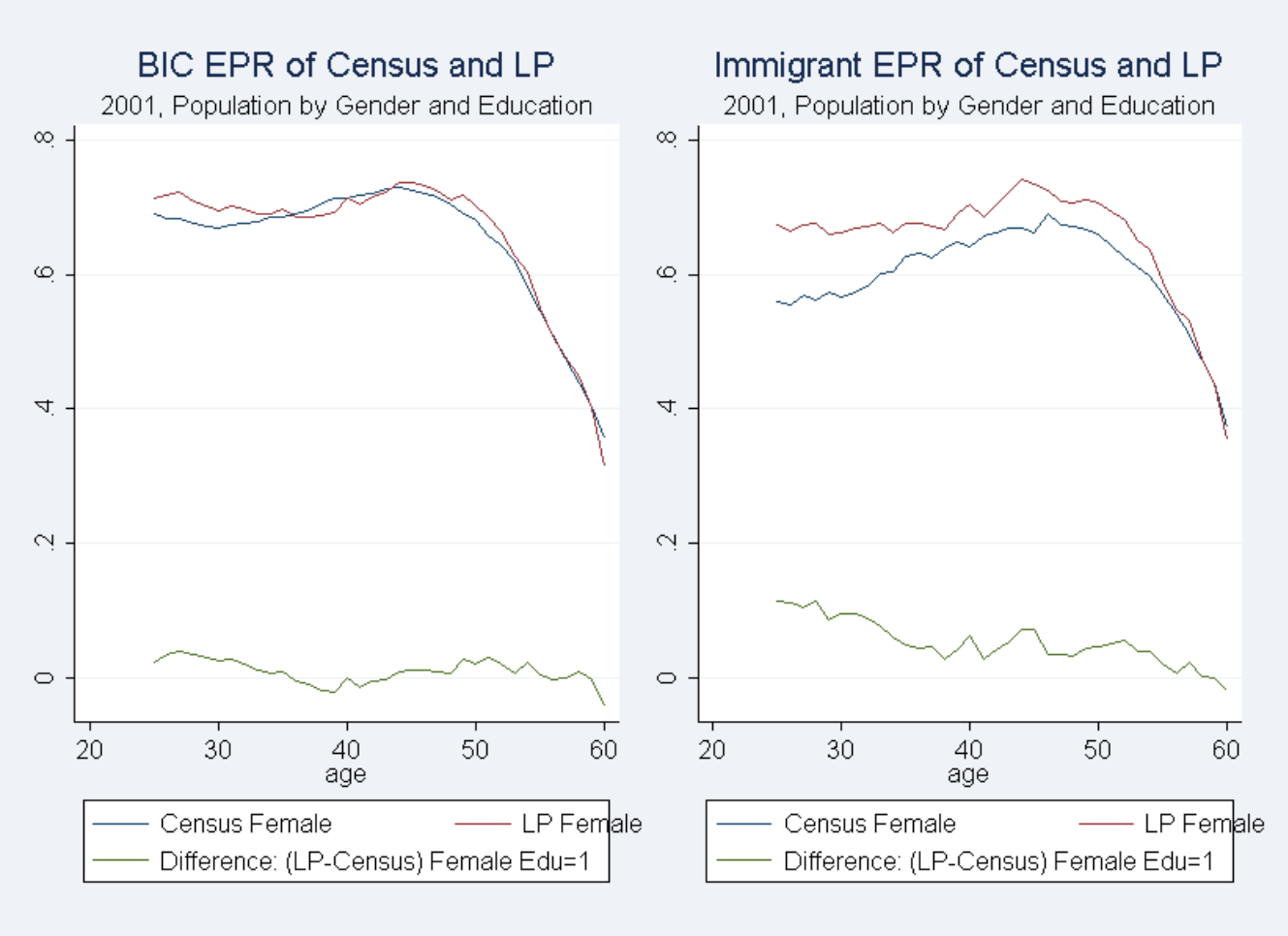

Illustration 6: E/P ratios of females with less than a university degree, born in Canada (BIC) and immigrants.

There is a larger difference for immigrants between LifePaths and the Census for males with less than a university degree. The average difference between the actual and simulated immigrant males

is about 8%, although it is slightly higher in the twenties and declines over time. This compares to about a 2% difference between the actual and simulated males that were born in Canada.

The next two graphs (Illustration 6) display the employment-to-population ratios for female immigrants and those born in Canada, who have less than a university degree. Females have a more complex employment situation because of pregnancy and maternity leave. In this situation, LifePaths handles changes to the employed boolean in the `MaternityLeave.mpp` module. This module simulates when the woman in question will take maternity leave, and for how long. Additionally, the simulated woman will also "decide" whether to return to the labour force.

With respect to the portion of the population born in Canada, there is less of a difference between the observed employment-to-population ratio of females from the Census and the employment-to-population ratio of simulated females from LifePaths, which show an average difference of about 2%. Immigrants in the Census compared to those in simulated in LifePaths have a difference in employment-to-population ratios of approximately 8%, on average. In both cases, however, the difference is more pronounced at ages less than 40 years.

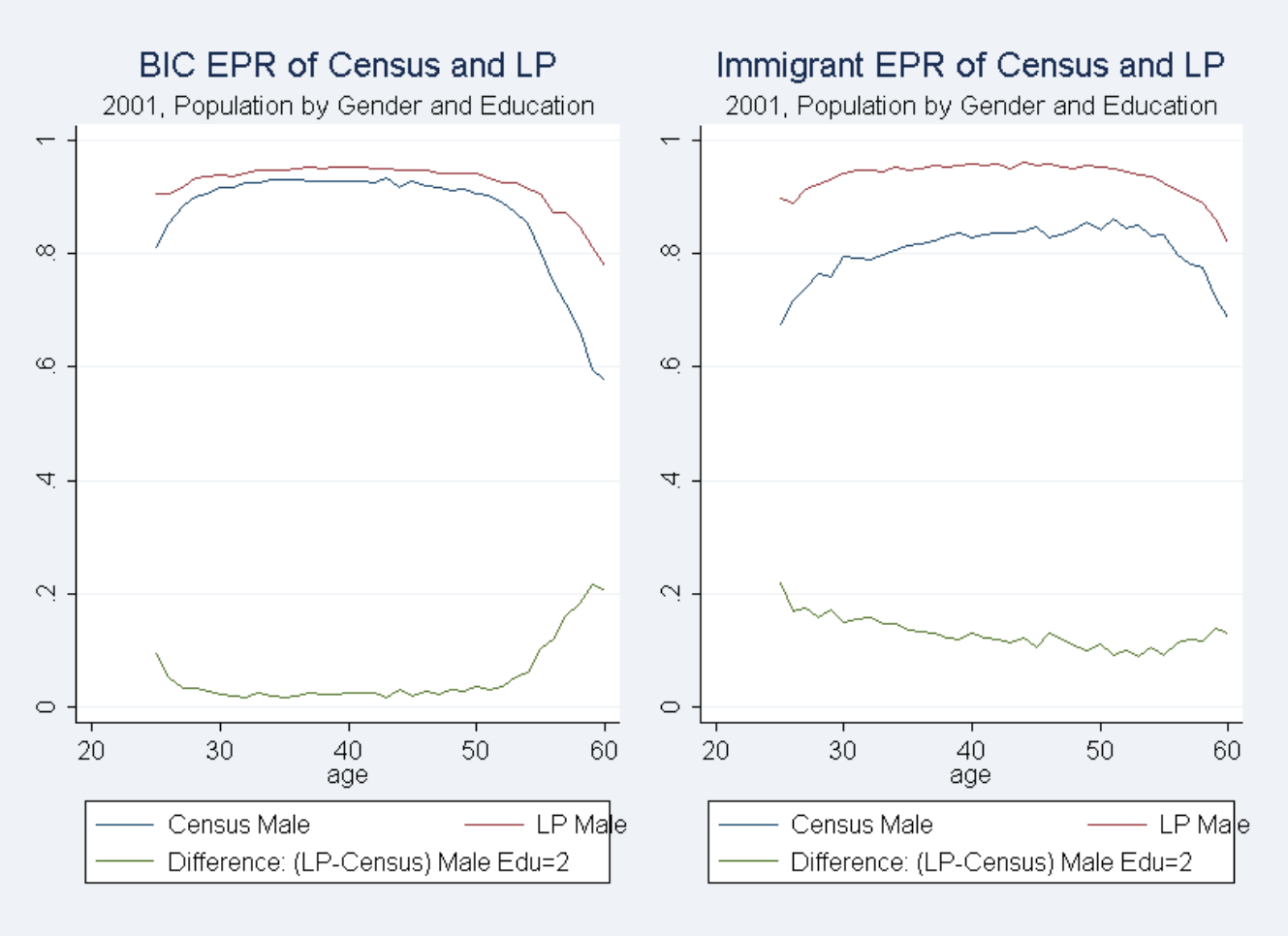

Illustration 7: E/P ratios of those males with a university degree or greater, born in Canada (BIC) and immigrants.

For males with a university degree or greater there is a visible separation between the employment-to-population ratio derived from the Census and that derived from a LifePaths simulation. For males, with a university degree or greater and born in Canada, there is an average difference between the simulated employment-to-population ratio and the actual of about 3%. For immigrants with the same characteristics the difference is, on average, about 12% although it is higher in the twenties and gradually declines over time.

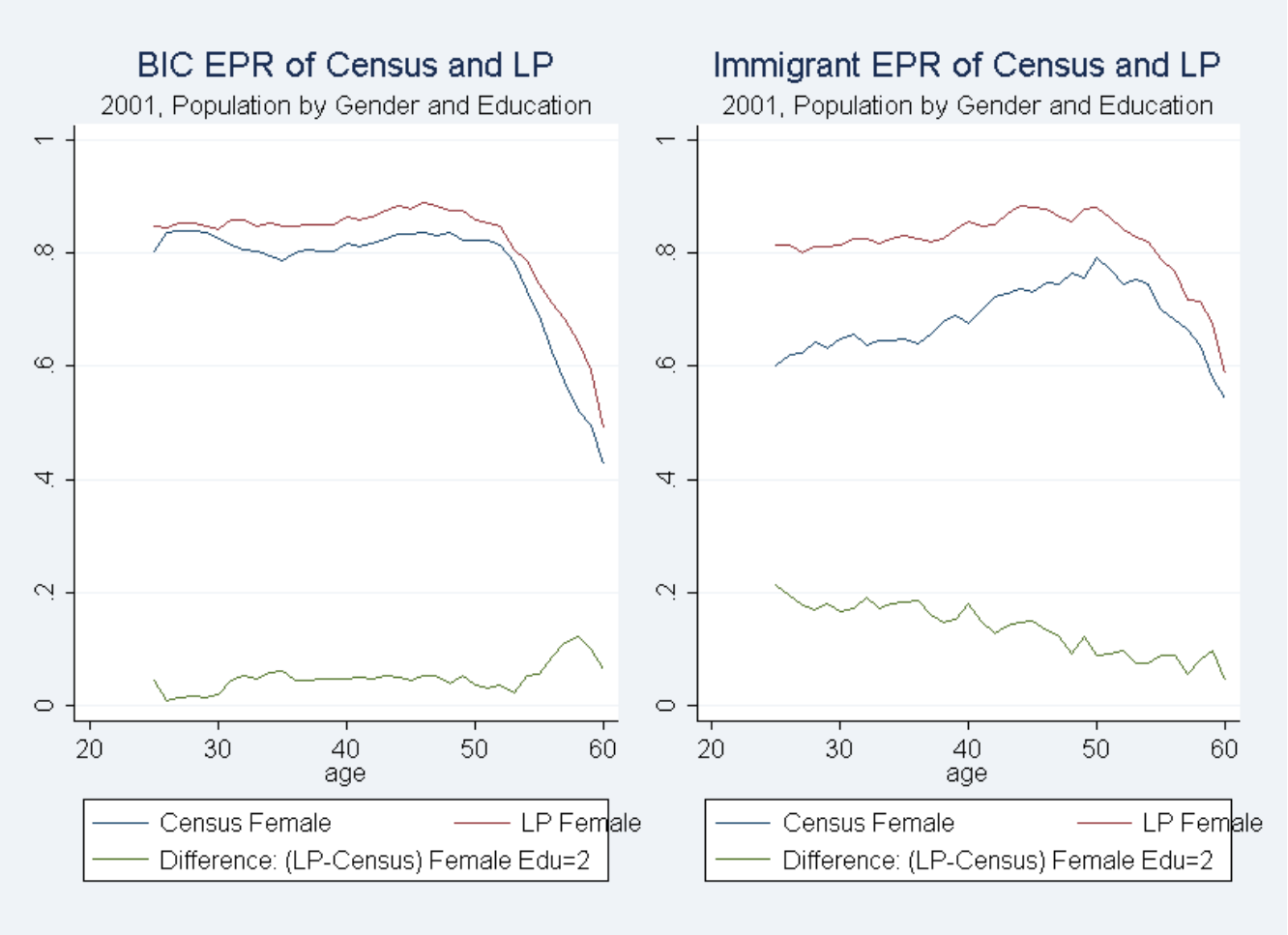

Illustration 8: E/P ratios of females with a university degree or greater, born in Canada (BIC) and immigrants.

Females who were born in Canada and earned a university degree or greater (Illustration 8), show an average difference of about 4% between the Census and LifePaths simulated employment-to-population ratios. Females who are immigrants show a much larger average difference of approximately 14%. The difference is closer to 20% in the mid-twenties and decreases over time.

Using Standardization to Compare Employment-to-Population Ratios

LifePaths has been designed to model the Canadian individuals and families, but very little work has been done focusing specifically on immigrants. The section above shows that the model fails to closely match the employment-to-population ratios of immigrant males and females in different age categories. The question remains whether the LifePaths model requires more informative variables to add to its behavioral equations, or whether the current variables may be sufficient if used more appropriately.

In order to give a rough indication of this question, direct standardization is used across age categories, and relevant variable categories, such as: educational level, marital status, the presence of children and provincial location. Educational level is divided into three categories for those with less than a high school diploma, those with high school but less than a university degree, and those with at least an undergraduate university degree and above. The marital status variable is true or false, as is the presence of children. Provincial location is either east (up to Ontario) or west; LifePaths contains only provincial level information with no distinction between urban and rural.

The basic idea behind direct standardization is to compare the employment-to-population ratios by removing the effects of the other variables in question. In this case, direct standardization is used to indicate whether the existing variables in the LifePaths model are sufficient to yield an accurate employment-to-population ratio.

The approach is relatively simple. Using the Census data from 2001, the employment-to-population ratios are calculated for the population with ages between 30 and 55, for both immigrants and the portion of the population born in Canada. Using direct standardization, a new E/P ratio for the entire population between 30 and 55 years of age is determined by holding the E/P ratio of immigrants constant under the variables of interest, but by using the population distribution of the portion of the Canadian-born population. The approach of direct standardization is meant to remove confounding factors, and to allow for direct comparison of the E/P ratios of the respective populations.

	Male E/P Ratio	Female E/P Ratio
Canadian-born	0.83	0.71
Immigrant	0.81	0.65
Standardized Immigrant	0.80	0.66

Table 1: Population E/P ratios of males and females for Canadian-born, immigrants and standardized immigrants. Data from 2001 Census.

Table 1 (above) shows that the variables in question seem to be roughly adequate for males. For females the variables used in LifePaths seem inadequate to explain the difference in employment-to-population ratios between Canadian-born women and immigrants. This overall result, while rough, indicates that more information is needed by LifePaths if it is to accurately simulate the E/P ratios of female immigrants.

One possible variable to take account of in the model is the number of years of experience that an immigrant has in Canada (Table 2, below). It seems reasonable that the longer immigrants are in Canada the more they will come to know the work requirements, and remedy any skill deficiencies that they may have. This means that immigrants will likely initially be less employable until they gain language, cultural, or directly job-related skills. The Census captures the individual's year of immigration, so it is simple to recalculate the employment-to-population ratios for those with more than five years experience of being in Canada versus those who have been in Canada longer.

	Male E/P Ratio	Female E/P Ratio
Canadian-born	.83	.71
Immigrants > 5 years	.84	.68
Immigrants < 5 years	.70	.49

Table 2: Population E/P ratios of males and females for Canadian-born and immigrants who have lived in Canada for less than 5 years, and immigrants who have lived in Canada for more than 5 years. Data from 2001 Census.

The results of Table 2 (above) are striking. It is clear that employment-to-population ratios for immigrants who have been in Canada for less than five years are much lower than those who have been in the country for more than five years. By comparison, the employment-to-population ratios of immigrants who have been in Canada for more than five years are much closer to the portion of the population born in Canada.

Conclusions and Recommendations

This report has given a broad overview of microsimulation and its importance in policy formation. In particular, this report has focused on examining the relative accuracy of Statistics Canada's LifePaths microsimulation model in generating employment-to-population ratios for immigrants as compared to the portion of the population born in Canada.

The importance of having comparable measures in the Census and the output of the LifePaths simulation model has been stressed, and the formulations used in this case have been demonstrated. Additionally, there is a brief discussion of the tables facility in LifePaths. The collection of LifePaths modules has been discussed with respect to the employment variable in order to illustrate the importance of understanding the complex relationships between modules. After alluding to the complexities of student work status as well as the changes which take place closer to retirement, this report has focused on the more stable working ages between thirty and fifty-five.

The employment-to-population ratios differs between those born in Canada and immigrants, when comparing the results from LifePaths and the 2001 Census. Additional differences seem to increase by educational attainment. In all cases, however, the immigrant employment-to-population ratios are overestimated by LifePaths.

By using direct statistical standardization, it appears that the employment-to-population ratio of immigrant males could be adequately modeled using the variables which currently exist in LifePaths. Immigrant females, however, did not show much difference after the direct standardization. This leads to the conclusion that additional explanatory variables are needed to adequately model this group's employment-to-population ratios.

By using an additional variable from the 2001 Census, which asked respondents the year in

which they immigrated, it is possible to derive the number of years that each respondent has lived in Canada. This variable, quantifying years lived in Canada, is thought to be a reasonable indicator of the adjustment of immigrants to the conditions of the labour market. Given the opportunity, most immigrants are prepared to work as soon as they arrive. While a lack of language may be a barrier to employment, often it is Canadian attitudes toward foreign credentials and work experience. The longer people live in Canada, the more likely they are to be able to participate in the labour force.

As a simple indication, the employment-to-population ratio was calculated for the immigrants by group, depending on whether they had been in Canada for longer than five years or not. The results indicate that immigrants who have been in the country for longer than five years have an employment-to-population ratio closer to the Canadian-born population, whereas both male and female immigrants who have been in Canada for less than five years have a lower employment-to-population ratio. While the results of this measure do not conclusively indicate that this new variable would be the best addition to LifePaths, it is an indication of a good place to start.

Appendix A: Method Used to Remove Students from Census, LFS and LifePaths Tables

In order to remove students from the LFS and Census in such a manner as to remain comparable to LifePaths, different approaches were used. In each case, students were removed from the survey in a different way, and because the Census takes place once every five years, and the LFS is a monthly on-going survey, it was necessary to create two different branches of the original table used to collect employment-to-population ratio data.

Census

In the Census, there is a boolean variable (ATTENDR) that describes whether the individual has been in school during the previous eight months. Because tables in Modgen are continuously updated, it is not possible to simply exclude those people that have recently been in school. Instead, a similar concept can be used in the table, although not exactly the same as that used in the Census, by including only those in the table that have not been in school during the previous eight months.

In code that follows, the boolean variable census_attendr8months is true only if the simulated individual has been in the state ES_OUT (a boolean created elsewhere in the model which is true if the person in question is out of school) for eight months.

```
// duration_trigger returns a boolean depending on whether the person has been
// out of school at least 8 months (8/12)
logical census_attendr8months = duration_trigger( es_state, ES_OUT, 0.66666666 );

...

table Person EPR_cen_rates03 // 3) Census 2001 Employment to Population Ratio
[ dominant && resident && year==2001 && curtate_age >= 15 && institutional_status
== COMMUNITY && census_attendr8months]
{
      isBIC // isBIC
      *
      split( curtate_age, CAT_AGE ) // cat_age
      *
      sex // ismale
      *
      my_edu_level // edu
      *
      {
```

```
            duration(isFtWorker, TRUE),          //EN decimals=4 ft_dur
            duration(isPtWorker, TRUE),          //EN decimals=4 pt_dur
            duration(isUnemployed, TRUE),        //EN decimals=4 unemp_dur

            duration(employed, TRUE)/duration(), //EN decimals=4 lp_total
            duration(employed, TRUE),            //EN decimals=4 dur_emp
            duration()                           //EN decimals=4 dur
      }
};
```

LFS

Creating a comparable table in LifePaths for the LFS is much more straightforward. The LFS is a monthly survey, and so individuals who are currently in school are simply excluded from the set when the employment-to-population ratios are calculated. In Modgen, a comparable idea is easy to implement, and a boolean variable is created based on whether a person is currently in school.

```
// #2 - LFS knows when a person is currently a student. This boolean will
// be true when a person is currently a student in LifePaths.
logical lfs_isNotStudent = (es_state == ES_OUT);

...

table Person EPR_lfs_rates05 // 5) LFS 2006 Employment to Population Ratio
[ dominant && resident && year==2006
     && curtate_age >= 15
     && institutional_status == COMMUNITY
     && lfs_isNotStudent
]
{
      isBIC // isBIC
      *
      split( curtate_age, CAT_AGE ) // cat_age
      *
      sex // ismale
      *
      my_edu_level // edu
      *
      {
            duration(isFtWorker, TRUE),          //EN decimals=4 ft_dur
            duration(isPtWorker, TRUE),          //EN decimals=4 pt_dur
            duration(isUnemployed, TRUE),        //EN decimals=4 unemp_dur

            duration(employed, TRUE)/duration(), //EN decimals=4 lp_total
            duration(employed, TRUE),            //EN decimals=4 dur_emp
            duration()                           //EN decimals=4 dur
      }
};
```

Appendix B: Complete Listing of Modgen File Immigrant.mpp

The entire Immigrant module is included here because tables in Modgen are not often well understood, and either are some of the more complex function implemented in the language.

```
/*
	Create user table to generate a listing of changes to immigrant portion of the population in LifePaths.
	These results will then be compared to the actual data collected in 1996-7 in the LFS to see how close
	the model is to reality.

	Immigrant is defined in different ways throughout the LP code.
*/

partition AGE_GROUPS { // Age group
	15, 25, 35, 45, 55, 65
};

partition CHILDREN_GROUPS { // Age group
	1, 2, 3, 4, 5
};

partition EARNINGS_GROUP {
	0, 10000, 20000, 30000, 40000, 50000, 60000, 80000, 100000, 120000
};

partition quarter_year {   // wksworked
	13, 26, 39
};

partition CAT_AGE { 20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,
                                    51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66} ;//EN age group

classification HAS_SSG { // Is secondary school graduate
	SSG_TRUE, // 1
	SSG_FALSE // 0
};

classification HAS_UNIV {
	UNIV_TRUE, // 1
	UNIV_FALSE // 0
};

classification WORK_STATUS {
	STUDENT,
	WORKFORCE
};

classification WORKER_TYPE {
	WT_NE, // 0
	WT_PE, // 1
	WT_SE, // 2
	WT_PE_SE // 3
};

classification PROV_RESID { // Province
	P_QUEBEC,      // 24
	P_ONTARIO,     // 35
	P_ALBERTA,     // 48
	P_BC,          // 59
	P_OTHER                // 100
};

classification EDUC_LEVEL { // Education
	EDU_PRI, // 1
	EDU_HS,  // 2
	EDU_BA,  // 3
	EDU_MA   // 4
};

classification WRK_STATUS { // Work Status
	WRK_FT, // 0
	WRK_PT, // 1
	WRK_UNEMP // 2
};

actor Person {

	logical isBIC = (immig_cat==NON_IMMIG);
```

```
	// Ages. Use my_ca in the table filter for the disaggregated view
	int my_ca = curtate_age;

	// Education level
	EDUC_LEVEL my_edu_level =     (ed_level_earnings == EARN_PHD)?            EDU_MA :
	                              (ed_level_earnings == EARN_MA)?             EDU_MA :
	                              (ed_level_earnings == EARN_BA)?             EDU_BA :
	                              (ed_level_earnings == EARN_NON_U)?          EDU_HS :
	                              (ed_level_earnings == EARN_SPS)?            EDU_HS :
	                              (ed_level_earnings == EARN_SEC_GRAD)?       EDU_HS :
	                              (ed_level_earnings == EARN_PRIMARY)?        EDU_PRI : EDU_PRI;

	// -----------------------------------------
	// EPR - Create two sets of tables (sigh...) for Census and LFS because of
	// their differing concepts of education
	// -----------------------------------------
	// #1 - Match the census (somewhat) ATTENDR variable which records whether a person
	//              has attended school in the last 8 months
	// duration_trigger returns a boolean depending on whether the person has been out of school
	// at least 8 months (8/12)
	logical census_attendr8months = duration_trigger( es_state, ES_OUT, 0.66666666 );

	// #2 - LFS knows when a person is currently a student. This boolean will
	// be true when a person is currently a student in LifePaths.
	logical lfs_isNotStudent = (es_state == ES_OUT);
	// -----------------------------------------

	// Weeks worked. From: CareerWorkTables.mpp
	// year_before_2000 exists in CareerWorkTables
	logical year_before_1995 = ( year < 1995);     //EN Year prior to 1995 flag
	logical year_before_1996 = ( year < 1996);     //EN Year prior to 1996 flag
	double  weeks_worked_in_1995 = //EN Weeks worked in 1995
	   52.0 * ( value_at_exits(year_before_1996, TRUE, workdur)
	               - value_at_exits(year_before_1995, TRUE, workdur) );

	// Categorize by province
	PROV_RESID my_prov_res =      ( prov_of_res == QUE )        ? P_QUEBEC :
	                              ( prov_of_res == ONT )        ? P_ONTARIO :
	                              ( prov_of_res == ALTA ) ? P_ALBERTA :
	                              ( prov_of_res == BC ) ? P_BC : P_OTHER;

	// Determine worker type based on wages: PE SE PE+SE NE
	WORKER_TYPE worker_type1995 = ( ( abs(wages_1995) != 0 ) && ( abs(se_earnings_1995) != 0) ) ? WT_PE_SE :
	                              ( ( abs(wages_1995) != 0 ) && ( abs(se_earnings_1995) == 0) ) ? WT_PE :
	                              ( ( abs(wages_1995) == 0 ) && ( abs(se_earnings_1995) != 0) ) ? WT_SE :
	                              WT_NE;

	WORKER_TYPE worker_type2000 = ( ( abs(wages_2000) != 0 ) && ( abs(se_earnings_2000) != 0) ) ? WT_PE_SE :
	                              ( ( abs(wages_2000) != 0 ) && ( abs(se_earnings_2000) == 0) ) ? WT_PE :
	                              ( ( abs(wages_2000) == 0 ) && ( abs(se_earnings_2000) != 0) ) ? WT_SE :
	                              WT_NE;

	WORKER_TYPE worker_type2005 = ( ( abs(wages_2005) != 0 ) && ( abs(se_earnings_2005) != 0) ) ? WT_PE_SE :
	                              ( ( abs(wages_2005) != 0 ) && ( abs(se_earnings_2005) == 0) ) ? WT_PE :
	                              ( ( abs(wages_2005) == 0 ) && ( abs(se_earnings_2005) != 0) ) ? WT_SE :
	                              WT_NE;


	// Combined SE and PE income
	double se_earnings_1995 = value_at_exits( year, 1995, se_earnings )
	                                    - value_at_exits( year, 1994, se_earnings );
	double wages_1995 = value_at_exits( year, 1995, wages ) - value_at_exits( year, 1994, wages );
	double pe_se_wages_1995 = (se_earnings_1995 + wages_1995);
	double se_earnings_2000 = value_at_exits( year, 2000, se_earnings )
	                                    - value_at_exits( year, 1999, se_earnings );
	double wages_2000 = value_at_exits( year, 2000, wages ) - value_at_exits( year, 1999, wages );
	double pe_se_wages_2000 = (se_earnings_2000 + wages_2000);

	double se_earnings_2005 = value_at_exits( year, 2005, se_earnings ) - value_at_exits( year, 2004, se_earnings
);

	double wages_2005 = value_at_exits( year, 2005, wages ) - value_at_exits( year, 2004, wages );
	double pe_se_wages_2005 = (se_earnings_2005 + wages_2005);

	// Determine work status: full-time (30hrs/week+), part-time (<30hrs/week), unemployed
	logical isPtWorker = (employed && !working_FT);
	logical isFtWorker = (employed && working_FT);
	logical isUnemployed = (!employed);
```

```
        WRK_STATUS wrk_status = ( isUnemployed ) ? WRK_UNEMP : ( isFtWorker ) ? WRK_FT : WRK_PT;

        // Weeks worked
        logical year_before_2005 = ( year < 2005);
        logical year_before_2004 = ( year < 2004);
        double  weeks_worked_in_2005 = //EN Weeks worked in 2005
                                52.0 * ( value_at_exits(year_before_2005, TRUE, workdur)
                                   - value_at_exits(year_before_2004, TRUE, workdur) );

};

// ---------------------------------------------
// Table groups
// ---------------------------------------------

table_group IMM_INFO1 // Immigrant Employment Categories
{
        IMM_new01,
        IMM_new02,
        IMM_new03
};

table_group EPR_CENSUS_TABLES // CENSUS - EPR Employment to Population Ratios
{
        EPR_cen_rates01,
        EPR_cen_rates03,
        EPR_cen_rates05
};

table_group EPR_LFS_TABLES // LFS - EPR Employment to Population Ratios
{
        EPR_lfs_rates01,
        EPR_lfs_rates03,
        EPR_lfs_rates05
};



// ---------------------------------------
// Census Employment to Population Ratios
// ---------------------------------------
table Person EPR_cen_rates01 // 1) Census 1996 Employment to Population Ratio
[ dominant && resident && year==1996 && curtate_age >= 15 && institutional_status == COMMUNITY &&
census_attendr8months ]
{
        isBIC // isBIC
        *
        split( children_at_home, CHILDREN_GROUPS) // children
        *
        split( curtate_age, CAT_AGE ) // cat_age
        *
        sex // ismale
        *
        my_edu_level // edu
        *
        {
                duration(isFtWorker, TRUE),                  //EN decimals=4 ft_dur
                duration(isPtWorker, TRUE),                  //EN decimals=4 pt_dur
                duration(isUnemployed, TRUE),         //EN decimals=4 unemp_dur

                duration(employed, TRUE)/duration(), //EN decimals=4 lp_total
                duration(employed, TRUE),                    //EN decimals=4 dur_emp
                duration()                                                  //EN decimals=4 dur
        }
};


table Person EPR_cen_rates03 // 3) Census 2001 Employment to Population Ratio
[ dominant && resident && year==2001 && curtate_age >= 15 && institutional_status == COMMUNITY &&
census_attendr8months]
{
        isBIC // isBIC
        *
        split( curtate_age, CAT_AGE ) // cat_age
        *
        sex // ismale
        *
        my_edu_level // edu
        *
        {
```

```
                duration(isFtWorker, TRUE),                  //EN decimals=4 ft_dur
                duration(isPtWorker, TRUE),                  //EN decimals=4 pt_dur
                duration(isUnemployed, TRUE),         //EN decimals=4 unemp_dur

                duration(employed, TRUE)/duration(), //EN decimals=4 lp_total
                duration(employed, TRUE),                    //EN decimals=4 dur_emp
                duration()                                                  //EN decimals=4 dur
        }
};



table Person EPR_cen_rates05 // 5) Census 2006 Employment to Population Ratio
[ dominant && resident && year==2006 && curtate_age >= 15 && institutional_status == COMMUNITY &&
census_attendr8months]
{
        isBIC // isBIC
        *
        split( curtate_age, CAT_AGE ) // cat_age
        *
        sex // ismale
        *
        my_edu_level // edu
        *
        {
                duration(isFtWorker, TRUE),                  //EN decimals=4 ft_dur
                duration(isPtWorker, TRUE),                  //EN decimals=4 pt_dur
                duration(isUnemployed, TRUE),         //EN decimals=4 unemp_dur

                duration(employed, TRUE)/duration(), //EN decimals=4 lp_total
                duration(employed, TRUE),                    //EN decimals=4 dur_emp
                duration()                                                  //EN decimals=4 dur
        }
};

// ----------------------------------------
// LFS Employment to population ratios
// ----------------------------------------
table Person EPR_lfs_rates01 // 1) LFS 1996 Employment to Population Ratio
[ dominant && resident && year==1996 && curtate_age >= 15 && institutional_status == COMMUNITY && lfs_isNotStudent ]
{
        isBIC // isBIC
        *
        split( curtate_age, CAT_AGE ) // cat_age
        *
        sex // ismale
        *
        my_edu_level // edu
        *
        {
                duration(isFtWorker, TRUE),                  //EN decimals=4 ft_dur
                duration(isPtWorker, TRUE),                  //EN decimals=4 pt_dur
                duration(isUnemployed, TRUE),         //EN decimals=4 unemp_dur

                duration(employed, TRUE)/duration(), //EN decimals=4 lp_total
                duration(employed, TRUE),                    //EN decimals=4 dur_emp
                duration()                                                  //EN decimals=4 dur
        }
};

table Person EPR_lfs_rates03 // 3) LFS 2001 Employment to Population Ratio
[ dominant && resident && year==2001 && curtate_age >= 15 && institutional_status == COMMUNITY && lfs_isNotStudent]
{
        isBIC // isBIC
        *
        split( curtate_age, CAT_AGE ) // cat_age
        *
        sex // ismale
        *
        my_edu_level // edu
        *
        {
                duration(isFtWorker, TRUE),                  //EN decimals=4 ft_dur
                duration(isPtWorker, TRUE),                  //EN decimals=4 pt_dur
                duration(isUnemployed, TRUE),         //EN decimals=4 unemp_dur

                duration(employed, TRUE)/duration(), //EN decimals=4 lp_total
                duration(employed, TRUE),                    //EN decimals=4 dur_emp
                duration()                                                  //EN decimals=4 dur
```

```
	}
};

table Person EPR_lfs_rates05 // 5) LFS 2006 Employment to Population Ratio
[ dominant && resident && year==2006 && curtate_age >= 15 && institutional_status == COMMUNITY && lfs_isNotStudent]
{
	isBIC // isBIC
	*
	split( curtate_age, CAT_AGE ) // cat_age
	*
	sex // ismale
	*
	my_edu_level // edu
	*
	{
		duration(isFtWorker, TRUE),		//EN decimals=4 ft_dur
		duration(isPtWorker, TRUE),		//EN decimals=4 pt_dur
		duration(isUnemployed, TRUE),	//EN decimals=4 unemp_dur

		duration(employed, TRUE)/duration(), //EN decimals=4 lp_total
		duration(employed, TRUE),		//EN decimals=4 dur_emp
		duration()					//EN decimals=4 dur
	}
};


// ------------------------------------------
// Tables for collecting data
// ------------------------------------------

table Person IMM_new01 // 1) 1996 Population
[ dominant && resident && year==1996 && curtate_age >= 15 && institutional_status == COMMUNITY ]
{
	isBIC // isBIC
	*
	split( children_at_home, CHILDREN_GROUPS) // children
	*
	split( curtate_age, CAT_AGE ) // cat_age
	*
	sex // ismale
	*
	wrk_status // wrkstatus
	*
	my_edu_level // edu
	*
	split( weeks_worked_in_1995, quarter_year ) // wksworked
	*
	{
		duration(), // decimals=4 lp_total
		duration(isFtWorker, TRUE),		//EN decimals=4 ft_dur
		duration(isPtWorker, TRUE),		//EN decimals=4 pt_dur
		duration(isUnemployed, TRUE),	//EN decimals=4 unemp_dur
		duration(employed, TRUE)/duration(), //EN decimals=4 epr
		duration(employed, TRUE),		//EN decimals=4 dur_emp
		duration()					//EN decimals=4 dur
	}
};

table Person IMM_new02 // 2) 2001 Population
[ dominant && resident && year==2001 && curtate_age >= 15 && institutional_status == COMMUNITY ]
{
	isBIC // isBIC
	*
	split( children_at_home, CHILDREN_GROUPS) // children
	*
	split( curtate_age, CAT_AGE ) // cat_age
	*
	sex // ismale
	*
	wrk_status // wrkstatus
	*
	my_edu_level // edu
	*
	split( weeks_worked_in_2000, quarter_year ) // wksworked
	*
	{
		duration(), // decimals=4 lp_total
		duration(isFtWorker, TRUE),		//EN decimals=4 ft_dur
		duration(isPtWorker, TRUE),		//EN decimals=4 pt_dur
		duration(isUnemployed, TRUE),	//EN decimals=4 unemp_dur
```

```
                duration(employed, TRUE)/duration(), //EN decimals=4 epr
                duration(employed, TRUE),                       //EN decimals=4 dur_emp
                duration()                                                      //EN decimals=4 dur
        }
};

table Person IMM_new03 // 3) 2006 Population
[ dominant && resident && year==2006 && curtate_age >= 15 && institutional_status == COMMUNITY ]
{
        isBIC // isBIC
        *
        split( children_at_home, CHILDREN_GROUPS) // children
        *
        split( curtate_age, CAT_AGE ) // cat_age
        *
        sex // ismale
        *
        wrk_status // wrkstatus
        *
        my_edu_level // edu
        *
        split( weeks_worked_in_2005, quarter_year ) // wksworked
        *
        {
                duration(), // decimals=4 lp_total
                duration(isFtWorker, TRUE),                     //EN decimals=4 ft_dur
                duration(isPtWorker, TRUE),                     //EN decimals=4 pt_dur
                duration(isUnemployed, TRUE),          //EN decimals=4 unemp_dur
                duration(employed, TRUE)/duration(), //EN decimals=4 epr
                duration(employed, TRUE),                       //EN decimals=4 dur_emp
                duration()                                                      //EN decimals=4 dur
        }
};
```

References

Allison, Paul. Event History Analysis - Regression for Longitudinal Data. Beverly Hills, CA: Sage, 1984.

Box-Steffensmeier, Janet, and Bradford Jones. Event History Modeling - A Guide for Social Scientists. New York: Cambridge University Press, 2004.

Rowe, Geoff, and Huan Nguyen. "Longitudinal Analysis of Labour Force Survey Data." Survey Methodology 30.1 (2004): 105-14. <http://www.statcan.ca/english/freepub/11-522-XIE/2002001/session9/rowe.pdf>

Canada. Statistics Canada."Census of Canada" (web site) Statistics Canada. Last Modified April 2008. Accessed 10 April 2008. <http://www12.statcan.ca/english/census/index.cfm>.

---. "Experimental Economy Theoretical Growth Model (XEcon)" (web site) Statistics Canada. Last Modified August 2002. Accessed 10 April 2008.<http://www.statcan.ca/english/spsd/XEcon.htm>.

---. "Labour Force Survey (LFS)" (web site) Statistics Canada. Last Modified April 2008. Accessed 10 April 2008. <http://www.statcan.ca/>.

---. "LifePaths" (web site) Statistics Canada. Last Modified 28 August 2007. Accessed 1 March 2008. <http://www.statcan.ca/english/spsd/LifePaths.htm>.

---. The LifePaths Microsimulation Model: An Overview. Accessed 1 March 2008. Statistics Canada Web Site <http://www.statcan.ca/english/spsd/LifePathsOverview_E.pdf>.

---. "Microsimulation Models." (web site) Statistics Canada. Last Modified 29 August 2002. Accessed 1 March 2008. <http://dissemination.statcan.ca/english/spsd/index.htm>.