Analyzing Interprovincial Migration Determinants in Canada: A Random-Effects Gravity Model Approach

Peter Woyzbun

November 2014

1 Introduction

Canadian interprovincial migration has experienced a steady increase over the past few years, with 320,000 people having moved across provinces between the years 2012-2013.[1] This represents a movement of slightly less than 1% of the total population and is the highest level in more than 25 years. Figure 1 provides a visualization of this increase in flow from 2000 to 2013. As first intuition would suggest, the biggest beneficiaries of migrant inflows are provinces that offer the greener economic pastures of commodity resource booms – Alberta and Saskatchewan. However, interprovincial migration flows have an impact that reaches beyond the provincial level.

A flow of migrants also represents a reallocation of labour, which impacts aggregate unemployment. As such, understanding what factors drive interprovincial migration is a key element in understanding how the aggregate economy adjusts to shocks. This paper looks to identify and examine such factors.

Our approach is to use a random-effects panel data gravity model and data for the period 2000 to 2013. The parameters are estimated using the generalized least squares method. The idea is to identify the aggregate determinants of migration flows. These determinants can be broadly classified as "push", "pull", or "frictional" factors. Push factors encourage outflows of migration, pull factors encourage inflows, and frictional factors discourage both.

We find that unemployment, the number of people aged 15-29, population size, and the CPI are all push factors. Wage levels, GDP per capita, and population size are pull factors, and CPI, unemployment, distance, and the price of new homes are frictional factors.

[1] When we say "interprovincial" or "provinces" we are referring to the territories as well.

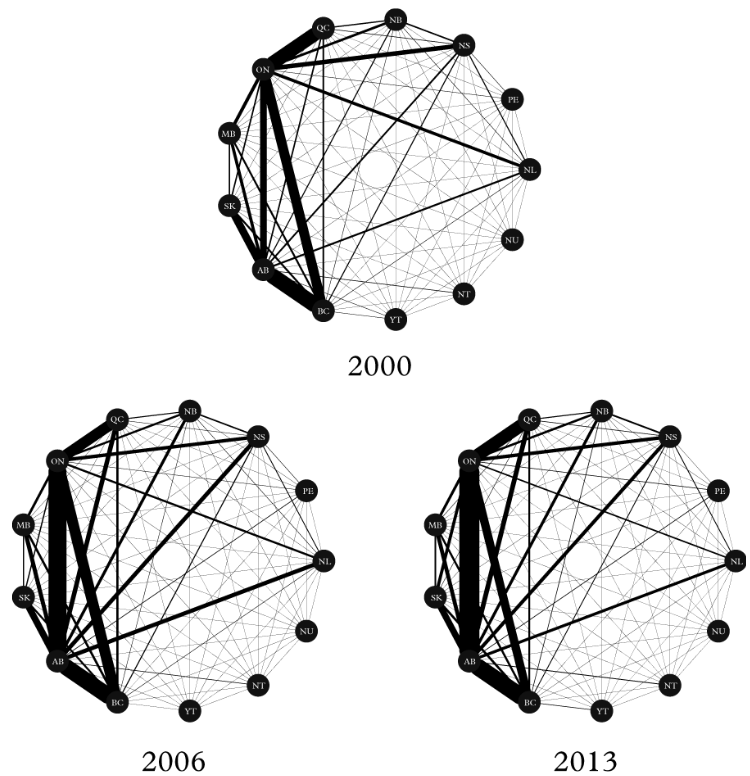

FIGURE 1.
Weighted graphs of interprovincial migration flows for 2000, 2006, and 2009. The thickness of the lines connecting provinces indicates the size of the migration flows. The graphs were generated in R using the igraph package and then edited in Adobe Illustrator for styling purposes.

The structure of the paper is as follows. We begin with some background and a review of the relevant literature. We then outline our model and variable selection. Next, we describe our data, panel design, and model specification. Finally, we outline our results, provide a brief discuss, and conclude the paper.

2 Background

In the literature, the analyzing of interprovincial migration generally involves one of two broad approaches. The first approach is to look at the determinants of migration on the level of individuals. That is, identifying the factors that increase or decrease an individual's probability of relocating. The second approach, and the approach taken in this paper, is to look at aggregate determinants. This approach ignores the characteristics of individuals and instead takes into account differences in conditions between provinces.

One example of the first approach is the work of Finnie (2004), who used panel logit models to determine the probability of an individual moving from one province to another between two given years. He found that moving was positively related with a province's unemployment rate, and identified a host of other influencing factors. Another example is an earlier study by Robinson and Tomes (1982). They used probit models and determined that individual migration depended, in part, on potential wage gains, language, and education levels.

An example of the second approach is the work of Amirault, Munnik and Miller at the Bank of Canada (2012). They used a gravity model and determined that employment rates, household incomes, and language are important in explaining migration, and that migration was greater within provinces than between provinces. Coloumbe, using the second approach as well, also found that unemployment was an important factor, but also labour productivity, and the rural/urban differential structure of the provinces (2006).

Our approach is of the second category, but we use a model based partly on the work of Tranos, Gheasi and Nijkamp (2013). They were interested in the determinants of country-to-country migration, rather than intranational, but their approach is well suited to our needs. They used a gravity model with random effects. We outline our own variation of the gravity model and our variable selection in the next section.

3 The Gravity Model

The basis of our approach is the generalized gravity model of immigration. The model takes into account the size (per capita GDP) of the source and destination province, and the distance between them. Migration between province i and j at time t, M_{ijt}, is defined:

$$M_{ijt} = \alpha \cdot \frac{(GDPpc_{it})^{\beta_1} \cdot (GDPpc_{jt})^{\beta_2}}{(dist_{ij})^{\beta_3}}$$

where $GDPpc_{it}$ is per capita GDP in province i at time t, $GDPpc_{jt}$ is per capita GDP in province j at time t, $dist_{ij}$ is the distance between the largest cities of province i and j, and α is a proportionality constant. The beta terms represent the sensitivity of M_{ijt} to changes in the aforementioned variables. By performing the relevant log transformations, and letting $\beta_0 = \log \alpha$, the model can be expressed in linear form:

$$\log(M_{ijt}) = \beta_0 + \beta_1 \log(GDPpc_{it}) + \beta_2 \log(GDPpc_{jt}) + \beta_3 \log(dist_{ij}).$$

The general idea behind the model is that migration flows between provinces should be proportional to their respective GDP levels and inversely proportional to the distance between them. The distance acts as a "friction" to migration. To account for other factors influencing migration flows, we extend the general gravity model and include a host of other economic variables. The final model is:

$$\begin{aligned}\log(M_{ijt}) = \beta_0 &+ \beta_1 \log(GDPpc_{it}) + \beta_2 \log(GDPpc_{jt}) + \beta_3 \log(dist_{ij}) \\ &+ \beta_4 \log(pop_{it}) + \beta_5 \log(pop_{jt}) + \beta_6 \log(grads_{it}) + \beta_7 \log(grads_{jt}) \\ &+ \beta_8 popshare_{it} + \beta_9 border_{ij} + \beta_{10} \log(UE_{it}) + \beta_{11} \log(UE_{jt}) \\ &+ \beta_{12} \log(CPI_{it}) + \beta_{13} \log(CPI_{jt}) + \beta_{14} newhouse_{it} \\ &+ \beta_{15} newhouse_{jt} + \beta_{16} wage_{jt}\end{aligned}$$

where pop_{it} and pop_{jt} are the populations in i and j respectively. Population variables are common additions to migration gravity models. It is generally understood that the higher a population in a source region, the more likely people are to migrate. Further, that they are drawn by destinations with higher populations and thus larger labour markets (Lewer & den Berg, 2008).

The variables $grads_{it}$ and $grads_{jt}$ represent the number of post-secondary graduates in the source and destination provinces. These components are included to test the common association between graduates and higher degrees of mobility found in other research (Kim, 2009). The variable $popshare_{it}$ represents the share of the population in the source province between the ages of 15 and 29. This group is also associated with a higher degree of mobility due to their having a longer period of time to realize the benefits of migration, and less ties to jobs and family (Amirault et al., 2012).

The $border_{ij}$ variable is a dummy variable that indicates whether or not i and j share a border. It is intended to determine whether or not a shared border increases migration flows.

The variables UE_{it} and UE_{jt} indicate the unemployment levels in the source and destination provinces. The importance of unemployment in determining migration flows has implications on a national level, as movement from regions of high unemployment to regions with labour shortages decreases aggregate unemployment (Bernard, Finnie, & St-Jean, 2008).

We capture the effect of price levels, by way of the consumer price index, in the source and destination provinces by including CPI_{it} and CPI_{jt}, respectively. Our intention is to determine whether migration is sensitive to price differentials between provinces. A higher CPI in a destination region, as compared to a source region, should be a barrier to migration, while a lower CPI should be a pull factor, as it implies a *ceteris paribus* increase in the standard of living.

The two variables $newhouse_{it}$ and $newhouse_{it}$ are new housing price indexes. They represent differences in the cost of purchasing a new home for a given year. We've included them to determine whether the cost of purchasing a new home influences migration decisions. Lastly, the variable $wage_{jt}$ is the average weekly wage in the destination province. It is included to determine whether or not wages are a pull factor for migration.

In the next section we outline and provide the sources of our data, and then describe the panel data structure we used in the estimation of our model.

4 Data

All of our data, aside from the geographical measures, are from various tables in Statistics Canada's CANSIM database.[2] The data is yearly and covers the period 2000–2013. The only time series that required altering was GDP, which we converted to a per capita measure using provincial population levels.[3] We've outlined each variable with its associated CANSIM table in Table 1 below.

The 2000-2013 observation window was chosen because we're interested in the recent determinants of migration flows, rather than long-run historical trends. The data is complete from a temporal perspective, but there are instances of missing values for the territories.

The CANSIM database does not have figures for the Yukon, Northwest Territories, or Nunavut, in the tables for: the number of post-secondary graduates

[2]The distances between largest cities was derived using optimal routes calculated by Google Earth. The dummy *border* variable is, of course, trivial.

[3]The actual assembling of the data into the necessary panel format was a more involved process. We outline this in the Technical Apendix.

Variable	CANSIM Table	Variable	CANSIM Table
migration_ijt	051-0019	popshare_it	051-0019
GDPpc_it	384-0038	grads_it	477-0030
GDPpc_jt	384-0038	grads_jt	477-0030
UE_it	282-0087	CPI_it	326-0021
UE_jt	282-0087	CPI_jt	326-0021
pop_it	051-0001	newhouse_it	327-0046
pop_jt	051-0001	newhouse_jt	327-0046
wage_jt	281-0049		

TABLE 1.
Variables and associated CANSIM tables.

(*Table 477-0030*), unemployment rates (*Table 282-0087*), and the price index for new homes (*Table 327-0046*). There are also no CPI figures for Nunavut. However, given the small contribution to total migration by the territories, and the panel structure of our analysis – which we outline in the next section – these missing figures are not a grave concern.

4.1 Panel Design

In general, a panel has the form X_{kt}, $k = 1, ..., N$, $t = 1, ..., T$, where k is the individual dimension, and t represents the time. In our design, the individual dimension is the flow from province i to province j, and the time dimension is the given year. This is illustrated in Table 2 below.

panel_id	year	migration	distance	$\cdots$
ONAB	2000	10358	3470	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

TABLE 2.
A sample panel from the data set.

The example panel is the flow between Ontario and Alberta in 2000. The full set of explanatory variables is not shown. In the next section, we outline the estimating of the parameters for our panel data model.

5 Model Specification

Our parameters are estimated in Stata using a generalized least squares (GLS) random-effects model. We chose a GLS approach due to the presence of autocorrelation between observations in our data set. In particular, we found that there is first-order correlation, so that an observation at time t is correlated with the observation at $t + 1$. We identified this AR(1) disturbance using Wooldridge's Test.

Wooldridge's test provides an F-statistic by which one can accept or reject the null hypothesis that there is no first-order correlation at a given confidence level (Drukker, 2003). The test is implemented in Stata using the xtserial function.[4] The results of the test are provided below:

$$F(1, 89) = 58.023, \quad Pr(F > 0.0000)$$

It is clear that the null hypothesis is strongly rejected and there is a serial correlation. As such, we use an autoregressive model to estimate our parameters. We chose a random-effects model, as opposed to a fixed-effects model, so that we could include time invariant variables. This follows the approach of Tranos, Gheasi, and Nijkamp (2013). To implement the model, we used the xtregar function in Stata.[5] We cover the results of implementing the model in the next section.

6 Empirical Results

The results of our model estimation were quite positive. We'll begin by examining the overall performance of the model. The results are provided in Table 3. The R^2 overall value of 0.7433 suggests that our model explains a respectable amount of what determines migration flows.

R^2 within	0.2113	Wald chi-squared	589.09
R^2 between	0.7509	Prob >chi-squared	0.0000
R^2 overall	0.7433		

TABLE 3.
Model statistics

[4]The xtserial function was developed by David Drukker. For more information see: http://www.stata.com/support/faqs/statistics/ panel-level-heteroskedasticity-and-autocorrelation/

[5]For more information on the xtregar function, see: http://www.stata.com/ manuals13/xtxtregar.pdf

We'll now discuss our findings for each of the explanatory variables in turn. The results of our coefficient estimates and their associated test statistics are outlined in Table 4.

The coefficient for the first variable, $GDPpc_{it}$, is not statistically significant, while the coefficient for $GDPpc_{jt}$ is significant and positive. This suggests that GDP per capita is a pull, rather than push, factor. Migration flows tend toward provinces with higher per capita GDP levels, but the GDP of the source province does not have a significant impact on its outflows.

Neither of the coefficients for the next two variables, $grads_{it}$ and $grads_{jt}$, which represent the numbers of post-secondary graduates in the source and destination province, are statistically significant. The coefficient for $popshare_{it}$ – the share of the population in the source province between the ages of 15-29 – however, is significant and positive. It is likely that, since the majority of graduates fall within the 15-29 age range, $popshare_{it}$ captures their mobility.

The coefficient of the distance variable is statistically significant and negative. This fits to the assumptions of the gravity model – that distance is a friction to migration. Not sharing a border however, is not, as is evident in the dummy variable's lack of significance. This is not surprising given the size of Canadian provinces and territories.

Both CPI_{it} and CPI_{jt} have significant coefficients but they are of opposite signs. A higher CPI in the source province is associated with increased outflows, while a higher CPI in the destination province is associated with decreased inflows. This makes sense intuitively, as one would expect a higher cost of living to make a province less appealing to settle in. On the other hand, a higher cost of living makes a province more appealing to leave. Further, the CPI in a destination province has a greater impact than the CPI in a source province. This makes sense given that leaving a province has a cost associated with it, while *not* migrating to a province does not.

The coefficients of both UE_{it} and UE_{jt} are also significant, and display a relationship to migration flows similar to that of the CPI. Higher unemployment will increase the outflow of migrants in a given province, while also acting as a friction decreasing the inflow of migrants. As before, this result is not surprising. A lack of employment makes a province less appealing for settlement to both those living there, and those not.

The population variables are both significant and positive, indicating that provinces with larger populations have both larger inflows and outflows of migrants. This fits with the conventional wisdom and is not a surprising result.

The price of new homes appears to only be significant as a friction to migration inflows. This seems reasonable, as we would expect higher prices on new homes to act more as a deterrent for settlement than as a reason to leave.

The last variable, $wage_{jt}$, is statistically significant and positive, indicating that higher average wages act as an enticement and increase migration inflows.

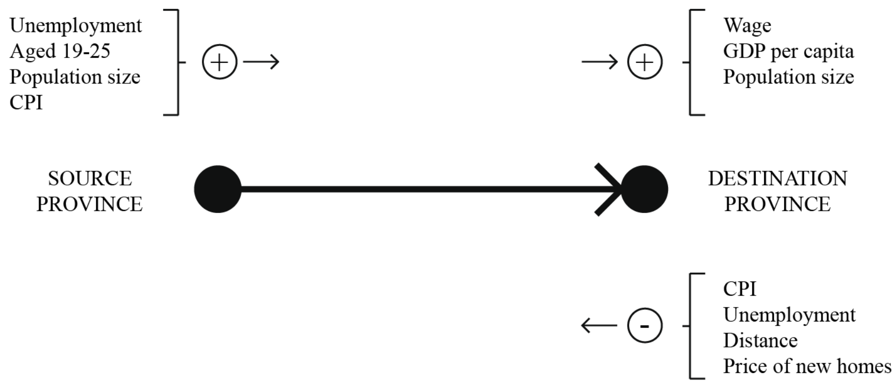

FIGURE 2.
Factors impacting interprovincial migration flows, as determined by our model. The positive determinants above the source province indicate push factors. The positive determinants above the destination province indicate pull factors. The negative determinants beneath the destination province indicate frictions decreasing the inflow of migrants.

We've summarized our findings with the diagram in Figure 2. It groups the statistically significant determinants based on their being push, pull, or frictional factors. In the next section we discuss our results and their implications.

7 Discussion

The model does well in explaining the aggregate determinants of interprovincial migration flows. This indicates that migration determinants persist over the 13 year time frame of our data set. However, the model might be augmented with additional variables to expand its explanatory ability. One example could be to include variables for regional climates and tax rate differentials. Another possible avenue for further analysis would be to apply the model to earlier periods. This could determine how the determinants of interprovincial migration flows change over time.

Further, variables might be included to capture the changes in commodity prices and the impact this has on migration flows. The idea being to determine the importance of provincial commodity booms in shaping migration flows.

8 Conclusion

By bridging methods commonly used in migration studies, network theory, time series models, and econometric analysis of panel data, we have provided a better understanding of the determinants of interprovincial migration.

We find that macroeconomic factors such as GDP per capita, inflation, and unemployment, are highly statistically significant in determining both migration inflows and outflows across provinces. Further, while education has traditionally been an important source of mobility – particularly in an age where barriers to human capital flows have been reduced – our model found little statistical significance for such a predictor variable. Our estimates also confirm that distances between provinces play a frictional role in interprovincial migration.

The value of our study is our gravity model panel data approach, as well as our use of Statistics Canada's CANSIM tables. Further, we provide an up-to-date analysis and our data encompasses the contemporary economic period.

Variables	Coef.	Std.Err.	z	P >z	95% Confidence Interval	
ln_gdppc_it	0.08157	0.195099	0.42	0.676	-0.300817	0.463957
ln_gdppc_jt	0.7798857	0.1939977	4.02	0.000***	0.3996572	1.160114
ln_grads_it	-0.1032991	0.0664166	-1.56	0.120	-0.2334732	0.026875
ln_grads_jt	-0.0385215	0.0655744	-0.59	0.557	-0.1670449	0.090002
popshare_it	8.744503	1.948799	4.49	0.000***	4.924927	12.56408
ln_dist_ijt	-0.6852085	0.1148231	-5.97	0.000***	-0.9102576	-0.4601593
border_ijt	-0.0449847	0.2464861	-0.18	0.855	-0.5280887	0.4381193
ln_cpi_it	2.754924	0.677417	4.07	0.000***	1.427211	4.082637
ln_cpi_jt	-4.903474	0.663943	-7.39	0.000***	-6.204779	-3.60217
ln_ue_it	0.1094327	0.0551827	1.98	0.047**	0.0012766	0.2175888
ln_ue_jt	-0.3654882	0.0557407	-6.56	0.000***	-0.4747379	-0.2562385
ln_pop_it	0.8385077	0.0900201	9.31	0.000***	0.6620715	1.014944
ln_pop_jt	0.6099672	0.0885521	6.89	0.000***	0.4364082	0.7835262
newhouse_it	0.0000158	0.0008196	0.02	0.985	-0.0015906	0.0016223
newhouse_jt	-0.0052393	0.000969	-5.41	0.000***	-0.0071385	-0.0033402
wage_jt	0.0020542	0.000309	6.65	0.000***	0.0014486	0.0026599
_cons	-8.546033	2.933318	-2.91	0.004***	-14.29523	-2.796835
	Note: * p <0.1, ** p <0.05, *** p <0.01					

TABLE 4.
Model coefficient estimates, standard errors, test statistics, and confidence intervals

References

Amirault, D., de Munnik, D., & Miller, S. (2012). *What drags and drives mobility: Explaining canada's aggregate migration patterns.*

Bernard, A., Finnie, R., & St-Jean, B. (2008). Interprovincial mobility and earnings. *Perspectives on Labour and Income*, *9*(10). doi: 10.1073/pnas.36.1.48

Coulombe, S. (2006). Internal migration, asymmetric shocks, and interprovincial economic adjustments in canada. *International Regional Science Review*, *29*(2), 199-223. doi: 10.1073/pnas.36.1.48

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *The Stata Journal*, *3*(2), 918-938. Retrieved from http://ageconsearch.umn.edu/bitstream/116069/2/sjart$_st0039.pdf$ doi: 10.1073/pnas.36.1.48

Finnie, R. (2004). Who moves? a logit model analysis of interprovincial migration in canada. *Applied Economics*, *36*, 1759–1779. doi: 10.1073/pnas.36.1.48

Kim, T. (2009). Transnational academic mobility, internationalization and interculturality in higher education. *Intercultural Education*, *20*(5), 395-405. doi: 10.1073/pnas.36.1.48

Lewer, J. J., & den Berg, H. V. (2008). A gravity model of immigration. *Economics Letters*, *99*(1), 164-167. Retrieved from http://ageconsearch.umn.edu/bitstream/116069/2/sjart$_st0039.pdf$ doi: 10.1073/pnas.36.1.48

Robinson, C., & Tomes, N. (1982). Self-selection and interprovincial migration in canada. *The Canadian Journal of Economics*, *15*(23), 474-502. doi: 10.1073/pnas.36.1.48

Tranos, E., Gheasi, M., & Nijkamp, P. (2013). *International migration: A global complex network.*

Technical Appendix: Data Cleaning

In order to organize the data into a panel structure we had to process and restructure the CANSIM tables. This involved first using Excel to remove unnecessary headers and add in columns for missing data where appropriate. We then used algorithms in R to perform the necessary restructuring of the data.

The first and most important restructuring was for the interprovincial migration data. The CANSIM table was organized so that a given year of migration flows was one row, and each flow was labeled by two separate cells in the same column indicating the source and destination of the flow. Using R we formated the data so that the destination and source were one cell in the same row as the migration flow. This way, each row represented the outflow from one province to another for a given year. This is illustrated in Figure 3.

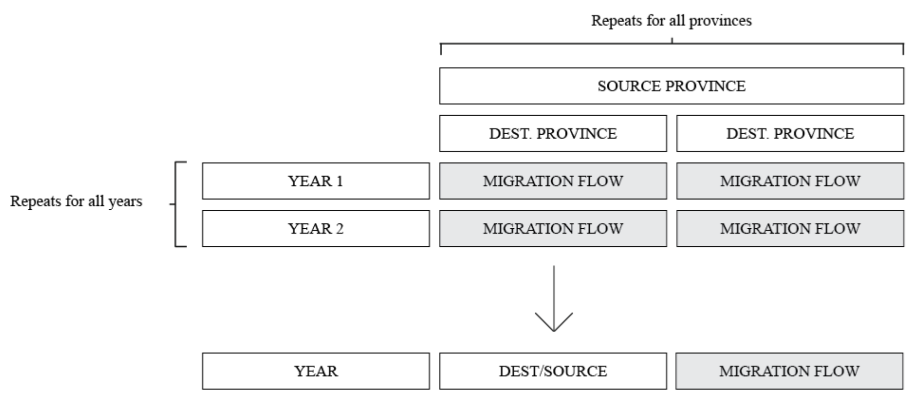

FIGURE 3.
Transformation of interprovincial migration data.

The result was 2184 rows of data in a panel format that we could then add the explanatory variables to. [6] To do so, we wrote algorithms in R that mapped each CANSIM table data entry to the appropriate year, source province, and destination province.

[6] This is the number of permutations for 13 provinces taken 2 at a time, and then multiplied by the 14 years in the data set.