Airport Activity as a Predictor of Regional Economic Development

Brody Reid

500993204

April 6, 2020

Abstract

While it can be intuited that a city with an active airport and strong intercity connections might have more local business and job creation, it is a difficult task to quantitatively describe this relationship. Here, an analysis of this nature is attempted for Canadian cities: are airport-level factors, such as number of passengers and number of flights, enough to predict economic development in a region? We employ regressions in order to predict individual median income in a region and then to predict the number of employed individuals in a region. While some of the analyses produce unrealistic results, we see that for citizens of Toronto a 10% increase in the number of air passengers produces a 0.8% increase in the number of employed individuals. Similarly, for citizens of Montreal, this same 10% increase in the number of air passengers yields a 0.5% increase in the number of employed individuals.

Introduction

Naturally we may consider airports and economic development to be tightly linked, but quantatative analyses of such relationships are scarce. An example of one such relationship is given in Figure 1: It describes the number of passenger boardings per capita and regional population growth from 1990-2000 in major U.S. cities. Naively we may assume that this relationship is obvious—a city with a busy airport means more people are travelling to the region for business and thus you would expect to see business and economic indicators increase. However, as Green shows in his 2007 paper, these analyses are very sensitive to external factors and not so straight-forward.[1] Therefore, one must approach an analysis between air-travel-effects and economic-effects carefully.

In this paper, we build off the work of Baker et. al., Bilotkach, and Green in order to examine how Canadian air traffic impacts a region's economic development.[1–3] We seek to describe, analytically, how airport-level factors might affect regional economic activity. The airport-level factors considered are the

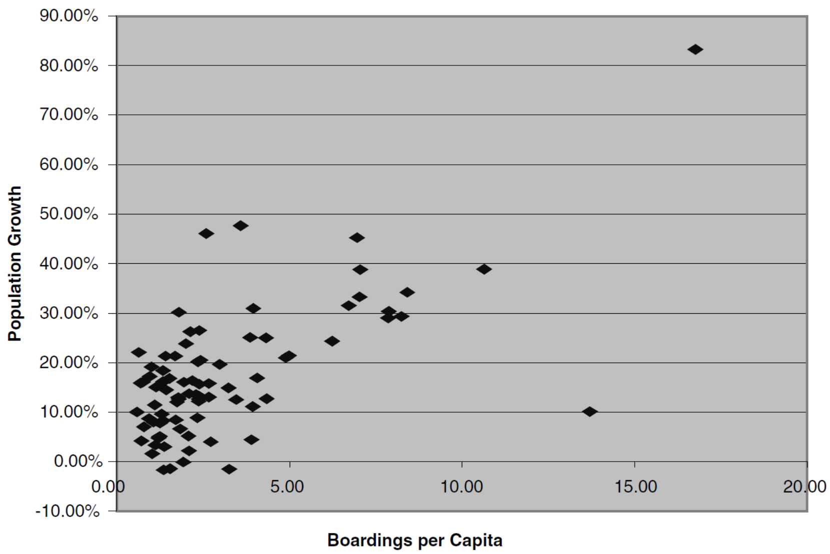

Figure 1: Boardings per capita in 1990 and population growth from 1990-2000 in major U.S. cities.[1]

number of flights to and from an airport, and the number of passengers arriving at and departing from an airport. The regional economic indicators are GDP, total number of employed individuals, and individual median income. To do this analytic description, we take the airport-level variables as predictors and total number of employed individuals as the response. The data spans a 16 year period starting in 2001—since our data overlaps during 2001-2016.

Literature Review

Academic studies on this topic are scarce; however, three appropriate papers on the topic were discovered and will serve as aids to this project. These papers quantitatively explore the relationship between an airport's activity and that region's economic development. This project will attempt to discover insights on this topic for Canadian cities.

A paper by V. Bilotkach exemplifies the complexity of adequately sourcing and preparing data for an analysis between airports and economic development.[3] Bilotkach uses regional U.S. financial data to create indicators of economic development—total employment, number of business establishments, and weekly minimum wage—and indicators of airport activity—number of flights departing an airport, seats provided, and passengers carried. The paper concluded that, indeed, an airport's activity boosts economic development: A 10% increase in the number of flights into a region's airport yields a 0.13% increase in employment, a 0.1% increase in the number of business establishments, and about

Table 1: Variables in the analysis.

Variable*	Label	Units
Individual median income	*INC*	2017 Canadian dollars
Number of passengers	*ACT*	Thousands of people
Number of flights	*MOV*	
Number of employed	*EMP*	Thousands of people
Regional GDP	*GDP*	2010 U.S. dollars (billions)

0.2% increase in average weekly wage. The paper also concluded that, keeping everything else constant, connecting an airport to a new destination creates 98 jobs and opens four new businesses.

The second paper, by Baker et. al, used Australian data to investigate whether the number of flights departing and arriving at an airport would affect the income of the region's population.[2] The paper showed that a linear combination of the two series was stationary and therefore had a long-run equilibrium. Then, using a causality test, the authors determined that there was a bi-directional relationship between the two variables.

The third paper, by R. K. Green, showed that in the U.S., passenger activity at a region's airport is a powerful predictor of economic growth.[1] Green found that a one standard deviation increase in passenger boardings per capita produces an 8.0% increase in employment growth. These regressions were run again with cargo volume and employment growth and it was found that there was no effect on employment growth.

Data and descriptions

Data

The analysis was conducted using regional Canadian airport data in order to predict individual median income in that region. The airport-level factors include aircraft departures and arrivals and passenger departures and arrivals. Initial analyses showed that if we only included these two variables, we were perhaps over-estimating their effect on the response variable. This introduced the possibility of having omitted variable bias. Omitted variable bias occurs when a confounding variable is left out of the regression equation: This leads to a strong relationship between the predictor variables and the response that, in reality, is illusory.[4] So, confounding variables were sought out in order to control the strength of the relationship between the airport-level predictors and median income. The total number of regionally employed individuals (*EMP*)

*All variables are numeric types.

Table 2: Included variables in each i-variable model and their corresponding errors for Toronto.

	ACT	MOV	EMP	GDP	RMSE	MAE
1			✓		627	603
2	✓	✓			453	435
3		✓	✓	✓	437	407
4	✓	✓	✓	✓	350	316

and regional gross domestic product (GDP) data were sourced in order to curb the problem of omitted variable bias. Descriptions of the variables are given in Table 1.

Airport-level data

To label our airport-level attributes, we refer to the aviation industry where aircraft departures and arrivals are called *aircraft movements* (MOV), and the number of passenger departures and arrivals is called *airport activity* (ACT). These two datasets were sourced from the government of Canada's open data portal and span the years 2001-2016.[5] These datasets required much cleaning, and necessitated an *R* package called `sqldf`. This package—which allows the user to query dataframes using *SQL* commands—was installed in order to filter for the desired cities and date range. For example, when cleaning MOV, the

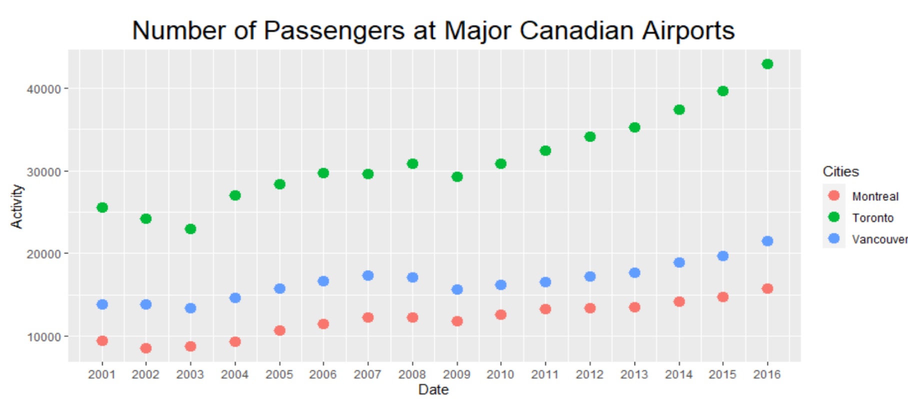

Figure 2: Note the steady drop in passengers during 2001-2003. This could be due to stricter security measures and citizens being hesitant of travelling after the September 11th attacks.

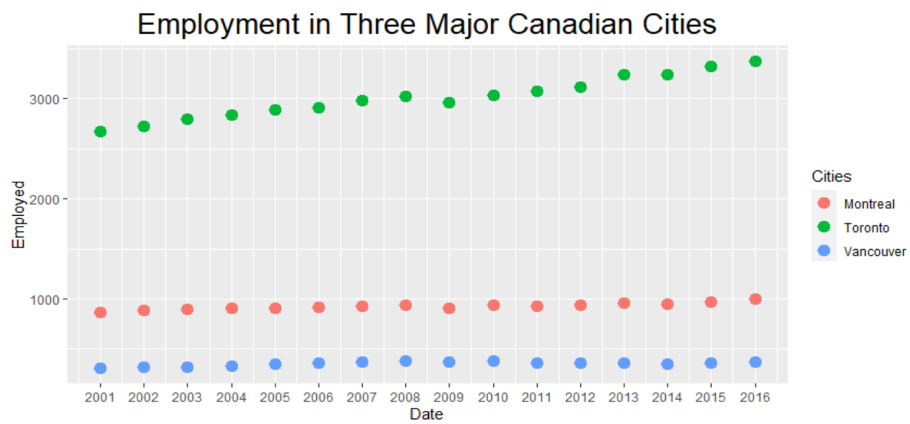

Figure 3: In Toronto, employment saw a linear rise between 2001 and 2016—except for 2008. This slight decrease in employment is likely due to the 2008 financial crisis. Vancouver and Montreal, however, remained relatively untouched by this crisis.

following *SQL* code was used in *R*:[**]

```
sqldf("
    SELECT Date, Airport, SUM(Movements) as Movements
    FROM movements
    WHERE Type = 'Total, itinerant and local movements'
    AND Peak = 'Number of movements'
    GROUP BY Date
    ")
```

In Figure 2 we see that the number of individuals travelling to and from Toronto, Vancouver, and Montreal grows relatively linearly—except around the time of the September 11th attacks in 2001 and around the time of the 2008 financial crisis. These disturbances can be seen in each city, but it is most prominent in Toronto's data.

Economic growth data

To determine economic development, the first analysis will use the median individual incomes of Canadians as the response and use total number of employed individuals and regional GDP as controls. In the second analysis we will use employed individuals as the response and use income and GDP as controls. These datasets were sourced from the government of Canada's open data portal and span the years 2001-2016.[5]

Figure 3 shows Toronto having a fast-growing, linear trend in the number of employed individuals, whereas Vancouver and Montreal have the same linear trend but the growth is very slow. A small decrease in the number of employed

[**]Similar code was used to clean all variables and can be found here. Samples of the cleaned data can be found in the appendix.

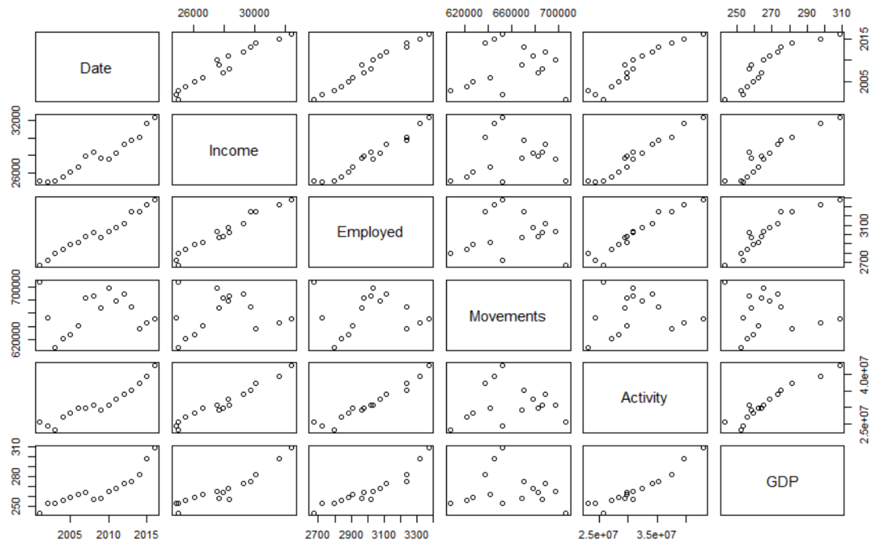

Figure 4: A matrix plot of the Toronto dataset from 2001-2016.

individuals can be observed between 2008 and 2009. This could be due to the 2008 financial crisis, where 400,000 Canadians lost their jobs.[6]

Method

The best fitting models were calculated for each number of variables (i.e., the best fit model with 1 variable, the best fit model with 2 variables, etc.). This was done using 10-fold cross validation in R using the `train()` function from the `leaps` package. For Toronto using income as the response, each i-variable model and their corresponding root mean square error (RMSE) and mean absolute error (MAE) are listed in Table 2. The four-variable model was desired since its RMSE and MAE were the lowest. So for each city, i, and intercept α, our regression equation for income is

$$INC_i = \alpha_i + \beta_{ACT_i} ACT_i + \beta_{MOV_i} MOV_i + \beta_{EMP_i} EMP_i + \beta_{GDP_i} GDP_i + \epsilon_i, \quad (1)$$

where ACT and MOV are the predictor variables, and EMP and GDP are the control variables. And for the regression for employment, our equation is

$$EMP_i = \alpha_i + \beta_{ACT_i} ACT_i + \beta_{MOV_i} MOV_i + \beta_{INC_i} INC_i + \beta_{GDP_i} GDP_i + \epsilon_i, \quad (2)$$

where ACT and MOV are the predictor variables, and INC and GDP are the control variables.

Table 3: Income regressions with equation (1).

	Toronto		Vancouver		Montreal	
	Coefficients	p-value	Coefficients	p-value	Coefficients	p-value
Intercept	-7040	0.34	2800	0.77	-14,900	0.082
	(7070)		(9140)		(7770)	
Activity	0.108	0.29	0.404	0.57	**0.866**	0.00028
	(0.0961)		(0.686)		(0.166)	
Movements	0.00889	0.059	-0.0238	0.26	0.00801	0.29
	(0.00422)		(0.0200)		(0.00724)	
Employed	**6.13**	0.0056	9.502	0.56	1.39	0.89
	(1.79)		(15.7)		(10.0)	
GDP	27.0	0.26	258	0.18	**196**	0.0017
	(22.7)		(181)		(47.6)	
RMSE	304		689		381	
AIC	240		267		248	
Adj R^2	0.96		0.94		0.98	

Bolded coefficients are significant at the 0.05 level.

Discussion

This analysis required that each city had its own regression run: Table 3 shows the results from the income regression using equation (1), and Table 4 shows the results from the employment regression using equation (2).

In the model with income as the response, we find that a 10% increase in passenger activity yields a 1.1% increase in Toronto's individual median income, a 2.6% increase in Vancouver, and a 3.6% in Montreal. These values seem somewhat realistic, even if a 2-3% increase in income is quite high. None of the reviewed papers reported values for this type of prediction so there is no way of comparing these values.

In the Montreal regression for employment, all else equal, we see that a 10% increase in passenger activity results in a 0.5% increase in employed individuals. In Toronto, the same 10% increase in passenger activity provides a 0.8% increase in number of employed individuals. In Vancouver, however, since the coefficient for activity is negative, we get that a 10% increase in passenger activity leads to an 11% *decrease* in employed individuals. This is interesting since in that regression, passenger activity was the only variable that was significant, *even*

Table 4: Employment regressions with equation (2).

	Toronto		Vancouver		Montreal	
	Coefficients	p-value	Coefficients	p-value	Coefficients	p-value
Intercept	1580	0.051	-131	0.45	**528**	0.034
	(723)		(169)		(218)	
Activity	0.00861	0.47	**-0.0242**	0.050	0.00416	0.66
	(0.0161)		(0.010)		(0.00921)	
Movements	-0.000889	0.12	0.000635	0.10	0.0000438	0.85
	(0.000521)		(0.000354)		(0.000229)	
Income	**0.0843**	0.0056	0.00341	0.56	0.00126	0.89
	(0.0246)		(0.00562)		(0.00906)	
GDP	-2.26	0.43	6.27	0.078	2.13	0.35
	(2.75)		(3.22)		(2.19)	
RMSE	36		13		11	
AIC	172		140		135	
Adj R^2	0.96		0.52		0.83	

Bolded coefficients are significant at the 0.05 level.

though it was negatively correlated. And the two variables on their own have a correlation value of 0.62. This problem could usually be attributed to overfitting, however, our R^2 value is only 0.52, so this would not be the case.

One analysis found similar results: Brueckner found that, on average, a 10% increase in passenger activity results in a 1% increase in employment in the region.[7] And Green found that, on average, a 10% increase in passenger activity results in a 2.8% increase in employment.[1] So the values we received for this 10% activity increase in Toronto and Montreal are quite realistic—at 0.8% and 0.5%, we can be confident in these results. However, for Vancouver, we need to review the data and method since an 11% decrease seems unrealistic.

In the Vancouver regression for income, we get an interestingly high R^2 value of 0.94; however, *none of the predictor or control variables are considered significant.* This problem might be due to overfitting, since, for example, income and passenger activity alone have a correlation of 0.95. So adding more variables to this model will only compromise its ability to predict accurately.

Despite gaining a few good insights, I believe this analysis needs a more robust model. The data for most of the regressions were too perfectly correlated to begin with: this made it difficult to extract real-world conclusions from the

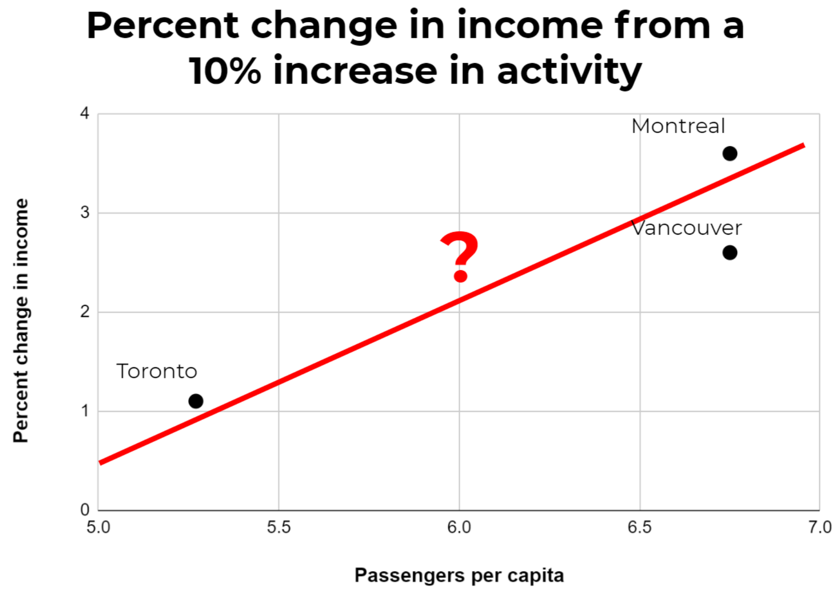

Figure 5: Possible future research.

analysis. For future research, more external factors should be considered which could control for seasonality of passenger activity and flight movements. One of which could be weather data: either a snowfall measure or extreme wind measure which would account for some of the seasonality in the flight movements data and make a distinct relationship with airport-level factors more evident.

Another possibility for future research would be to see if the percent changes we found in income when activity was increase by 10% is correlated to another variable. One idea would be to see if these percentages are correlated with the activity per capita of the region. Something like Figure 5 would be an exciting research endeavour.

It is worth mentioning that other algorithms were attempted after the regression analysis that is presented in this paper. However, all of these other algorithms produced unrealistic or nonsensical results and, therefore, are not included in the report. From the *R* package `party`, a conditional inference tree algorithm called `ctree` was applied to the data. This algorithm was promising as it is promised to be "applicable to all kinds of regression problems, including nominal, ordinal, numeric, censored as well as multivariate response variables".[8] This was not the case: for the Toronto dataset, the algorithm returned a 1-node tree and provided no interpretable results. It is most likely the case that this *R* function did not work because of the lack of information in our data, since each city's dataset only has 16 rows (the years 2001-2016). With more rows (years), this algorithm may have worked.

Along with the unsuccessful decision tree algorithm, a K-means algorithm was applied to the Toronto dataset in *Weka* using an 80/20 train-test scheme. From this method, two clusters were formed, each taking 50% of the rows in the test set. The mean values for clusters did not provide any insight. Like the decision tree algorithm, this method likely failed because of only having 16 rows in our datasets.

References

[1] Richard K Green. Airports and economic development. *Real estate economics*, 35(1):91–112, 2007.

[2] Douglas Baker, Rico Merkert, and Md Kamruzzaman. Regional aviation and economic growth: cointegration and causality analysis in australia. *Journal of Transport Geography*, 43:140–150, 2015.

[3] Volodymyr Bilotkach. Are airports engines of economic development? a dynamic panel data approach. *Urban Studies*, 52(9):1577–1593, 2015.

[4] Christoph Hanck, Martin Arnold, Alexander Gerber, and Martin Schmelzer. *Introduction to Econometrics with R*. Department of Business Administration and Economics University of Duisburg-Essen, 2019.

[5] Treasury Board Secretariat of Canada Open Government. Open data — open government, government of canada, 2017.

[6] Jason Gilmore and Sébastien LaRochelle-Côté. Canada's employment downturn, Dec 2009.

[7] Jan K Brueckner. Airline traffic and urban economic development. *Urban Studies*, 40(8):1455–1469, 2003.

[8] Torsten Hothorn, Kurt Hornik, and Achim Zeileis. ctree: Conditional inference trees. *The Comprehensive R Archive Network*, page 1, 2015.

Appendix

Cleaned data samples

MOV in absolute number of flights:

	Date	Airport	Movements
1	2001	Abbotsford, British Columbia	141098
2	2001	Boundary Bay, British Columbia	215442
3	2001	Calgary International, Alberta	236285
4	2001	Calgary/Springbank, Alberta	161457
5	2001	Chicoutimi, Quebec	65481
6	2001	Edmonton City Centre, Alberta	97038

ACT in thousands of passengers:

	Date	Airport	Activity
1	2001	TORONTO/LB PEARSON INTL ON	25517
2	2001	VANCOUVER INTL BC	13671
3	2001	MONTREAL/DORVAL INTL QC	8036
4	2001	CALGARY INTL AB	7447
5	2001	EDMONTON INTL AB	3594
6	2001	OTTAWA INTL ON	3312

INC in 2017 Canadian dollars:

	Date	Geo	Income
1	2001	Canada	22600
2	2001	Newfoundland and Labrador	16800
3	2001	St. Johns, Newfoundland and Labrador	21000
4	2001	Prince Edward Island	19600
5	2001	Nova Scotia	19600
6	2001	Halifax, Nova Scotia	24100

EMP in thousands of individuals:

	Date	Geo	Employed
1	2001	Canada	14932.3
2	2001	NL	203.7
3	2001	Avalon Peninsula, NL	108.3
4	2001	Notre Dame-Central Bonavista Bay, NL	52.4
5	2001	Northern Peninsula-Labrador, NL	43.0
6	2001	Prince Edward Island	63.8

Example of *GDP* in billions of 2010 U.S. dollars for Toronto:

	Date	GDP
1	2001	243.08
2	2002	253.28
3	2003	252.56
4	2004	256.00
5	2005	259.02
6	2006	262.23