Modeling Housing Prices in Canadian Cities in a Supply and Demand Framework

by

Karl Skogstad

An essay submitted to the Department of Economics
in partial fulfillment of the requirements for
the degree of Master of Arts

Queen's University
Kingston, Ontario, Canada
July 2009

Abstract

There has been much research on why housing prices vary within a city. The usual reasons cited include location within a city, square footage of the house, number of bathrooms, etc. On the other end of the spectrum, there have been many papers written about why housing prices vary between Canadian provinces or between different states in the U.S. These studies focus more on macro economic factors such as employment rate and income levels between these regions. This paper follows up on recent research, which suggests that provincial housing markets do not really exist. The focus should instead be on comparing housing markets of different cities. This paper develops a supply and demand model which attempts to pinpoint which economic factors cause housing price variations between different census metropolitan areas (CMAs) in Canada. The second contribution to this paper is an explanation on how a data set can be obtained to create such a model. There is a lack of data at the CMA level for certain variables, and this paper explains how to solve this potential problem.

Acknowledgments

I would like to thank everyone who has made this work possible. First and foremost, my greatest thanks go to Dr. Huw Lloyd-Ellis, who supervised me throughout this process. His insight and willingness to sit down and talk about the paper for hours on end was a definite help with this paper. Thanks also to Dr. James Mackinnon who took the time to help me with various econometric problems which arose in this paper.

Thanks also to my colleague Kris Shaw for his insights, and to David Byrne for his help in data procurement. Thank you also to employees of the MLS, CMHC, and Statistics Canada for their help during the data gathering phase. My thanks also go to the Social Science Research Council of Canada and Queen's University for funding concerns.

Finally, thank you to Shanna Skogstad and Fiona Skogstad for their support with this paper. Their encouragement and insight have made the process all the easier.

All errors within this paper are, of course, that of the author.

Contents

List of Tables

List of Figures

1 Introduction

The largest single purchase in the lifetime of most Canadians is that of a house. Yet at this time, there still does not exist an adequate model to explain how housing prices and quantities vary between regions in Canada. In an attempt to rectify this problem, this paper estimates a supply and demand model of housing across Canadian census metropolitan areas (CMAs) in the hopes of replicating the success that similar models have had with U.S. data. Past papers have focused on housing price variations within a city, while others still have focused on differences in housing prices between provinces. The first main contribution of this paper is to further the research in housing prices, by creating a model where pricing differences between cities is examined.

Where most Canadian papers in the past have focused on provincial level data, the regions that were chosen for this paper are CMAs. It is known that economic conditions between cities within a province may vary quite a lot. Thus to properly examine the Canadian housing markets the model must be applied on a more local level. This belief is reinforced by recent research by Byrne et al. (2008), who find that Canadian housing markets appear to be inherently local in nature.

The second main contribution of this paper is the construction of a data set that can be applied at the level of Canadian CMAs. Due to the large amount of data required, it was necessary to combine various data sources to obtain the desired variables at the CMA level. Although provincial level data was used for certain variables due to a lack of data at the CMA level, an effort was made to construct CMA level data whenever possible. Many of the variables used in this paper have, until now, been unavailable publicly in Canada.

There are many interesting relationships which we hope this model can explain, as the following figures show, certain variables seem to be highly correlated with the movement in housing prices. From Figure 1, it is obvious that a strong relationship between housing prices and rental rates clearly exists across time and space. This makes intuitive sense since it would be likely that shelter costs would be influenced by the same underlying factors such as wages and costs.

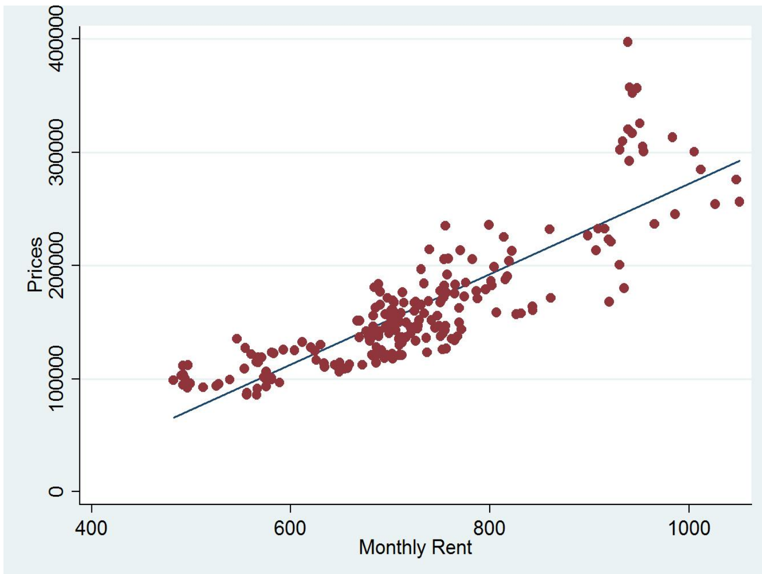

Figure 1: Relationship between Housing Prices and Rent

Though, perhaps not as strong as other relationships, a negative relationship between the unemployment rate and housing prices clearly exists as shown in Figure 2. Once again this is to be expected, as unemployment causes a reduction in income which shifts buyers out of the market, lowering overall housing demand. Along the same line, a positive relationship exists between housing prices and income. Figure 3 shows that more income allows more people to own their homes which shifts up demand. Note that the cluster of points above the overall trend, represent the city of Vancouver and have been denoted with squares instead of circles in Figures 2 and 3.

Finally, there clearly exists a strong positive correlation between current housing prices and those from the previous year, as shown in Figure 4. In fact this proves to be

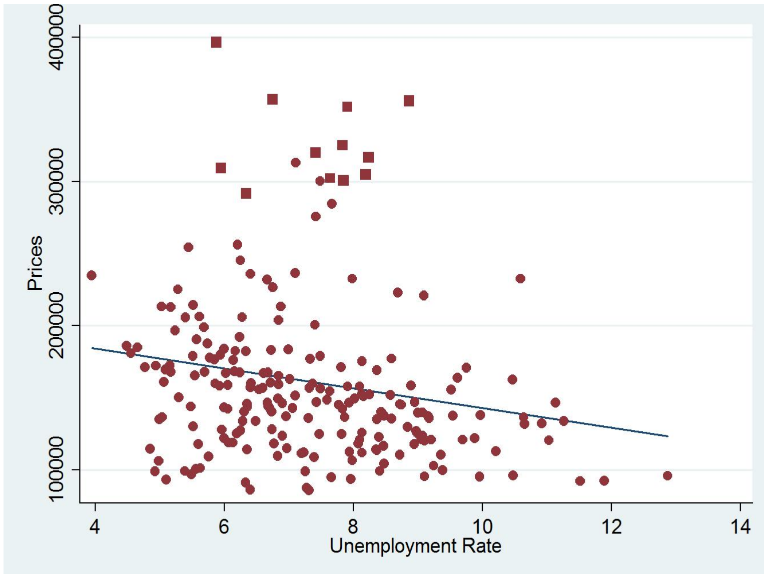

Figure 2: Relationship between Housing Prices and the Unemployment Rate

the strongest and most significant effect on housing prices. Figures 5 and 6 show the relationship between prices and quantity and prices and user cost, respectively. Clearly the first relationship is positive, while the second is negative.

Unfortunately, these relationships may be misleading. Though all variables are in real terms, the data is not detrended. Over time it is known that both income and housing prices rise, thus a positive relationship is all but guaranteed. In order to get around this problem, the models used in the body of this paper will be based on the one year differences in logged values of the data. Thus it will be possible to compare the growth rates of variables over time to determine their impact on housing prices. Figures 7 through 11 represent the same data as figures 1 through 5, except that each point represents the growth rate of the variables.

From these figures we see that a positive relationship still exists between prices and rent, but is not as strong as it previously was. The same can be said about the relationship between prices and lagged prices. On the other hand, price and unemployment

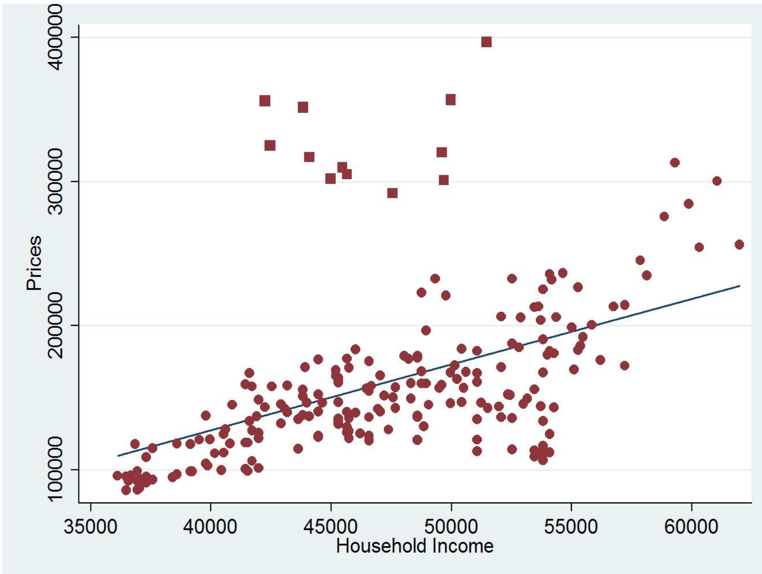

Figure 3: Relationship between Housing Prices and Mean Income

seem to have a very slight positive relationship, and price and income have a negative one. These last two results are quite different from what theory would otherwise predict. Also of note is the fact that Vancouver no longer stands out when the data is presented in growth rate form. For further graphical analysis, Appendix 3 includes a number of figures which illustrate the relation between prices and other variables on a CMA by CMA basis.

The relationships depicted in these figures and others will be further explored in this paper. Part 2 of this paper is concerned with the previous work undertaken in this field. The major papers which developed the model used here are reviewed. In addition, other papers which have looked at Canadian data are analyzed. The particular specification adopted here differs from other papers in the literature, and we will discuss the reasons for this. Finally, papers which have discussed specific variable effects are reviewed. Those which had significant results are used to modify our model.

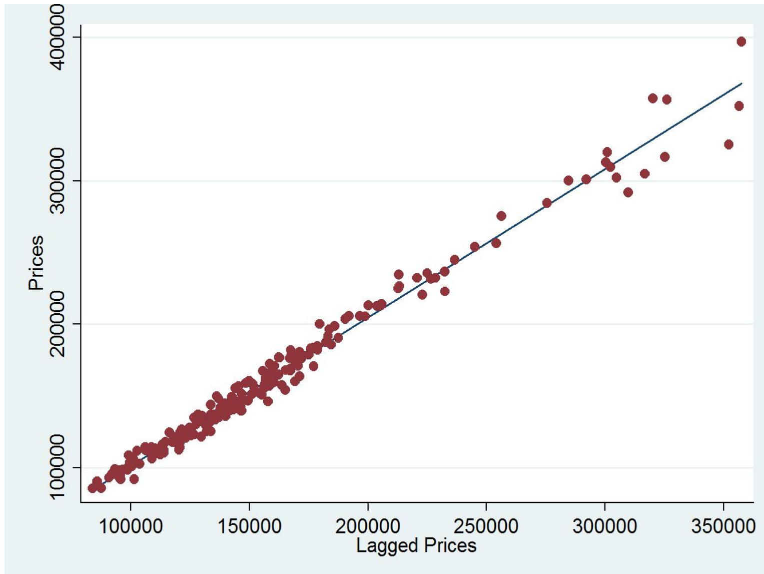

Figure 4: Relationship between Housing Prices and Lagged Housing Prices

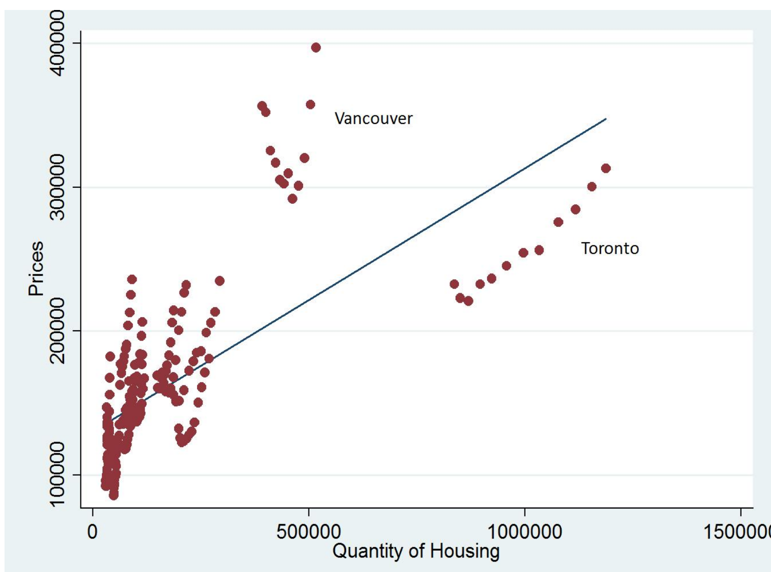

Figure 5: Relationship between Housing Prices and Quantity of Housing

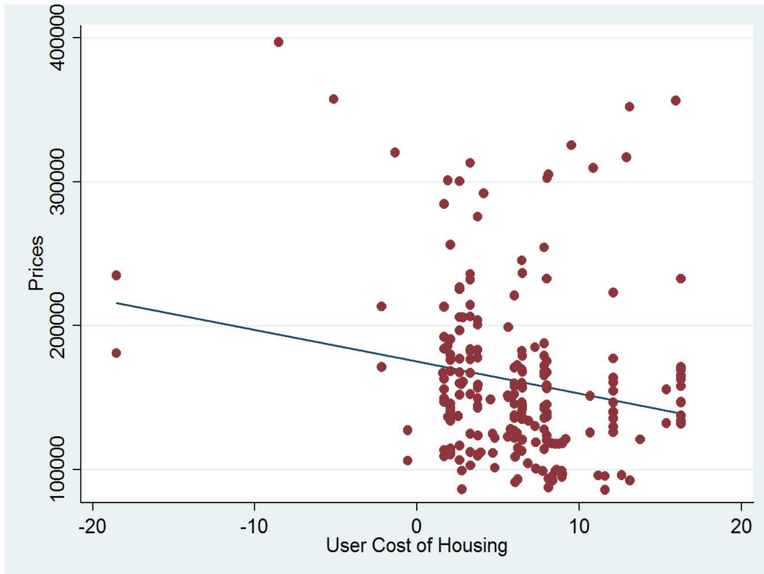

Figure 6: Relationship between Housing Prices and the User Cost of Housing

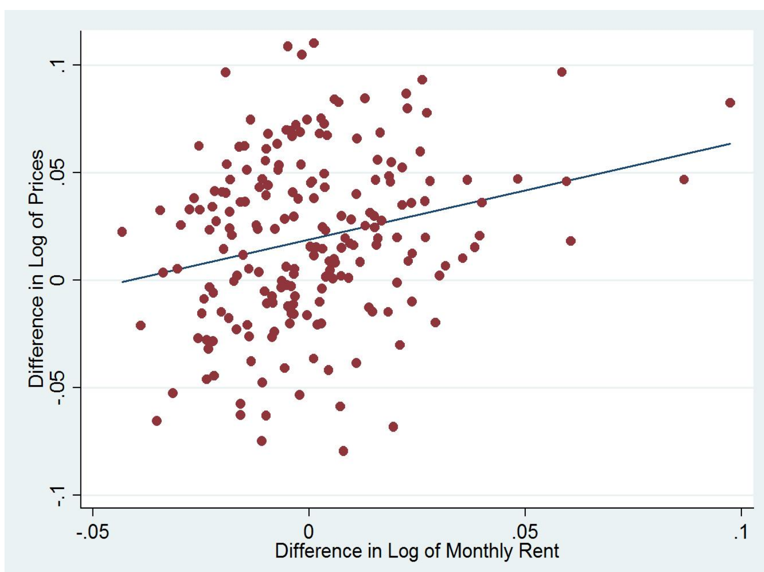

Figure 7: Relationship between the growth in Housing Prices and the growth in Rent

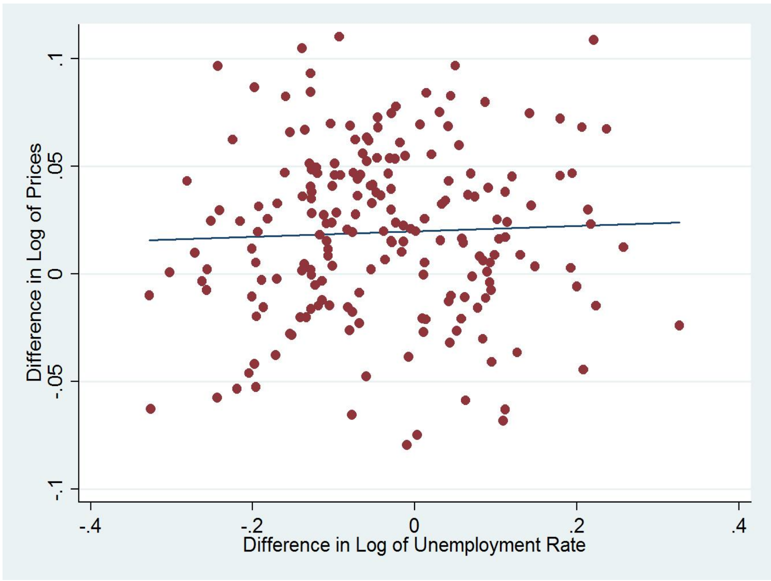

Figure 8: Relationship between the growth in Housing Prices and the growth of the Unemployment Rate

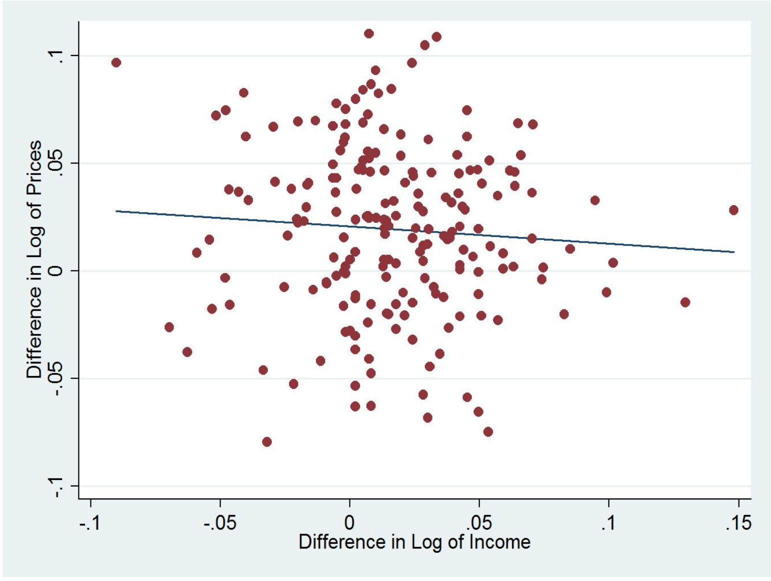

Figure 9: Relationship between the growth in Housing Prices and the growth in Mean Income

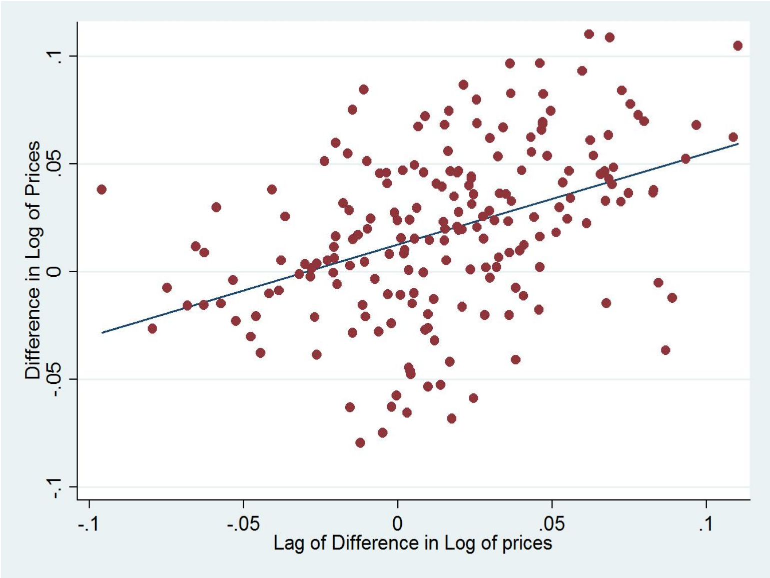

Figure 10: Relationship between the growth in Housing Prices and the growth in Lagged Housing Prices

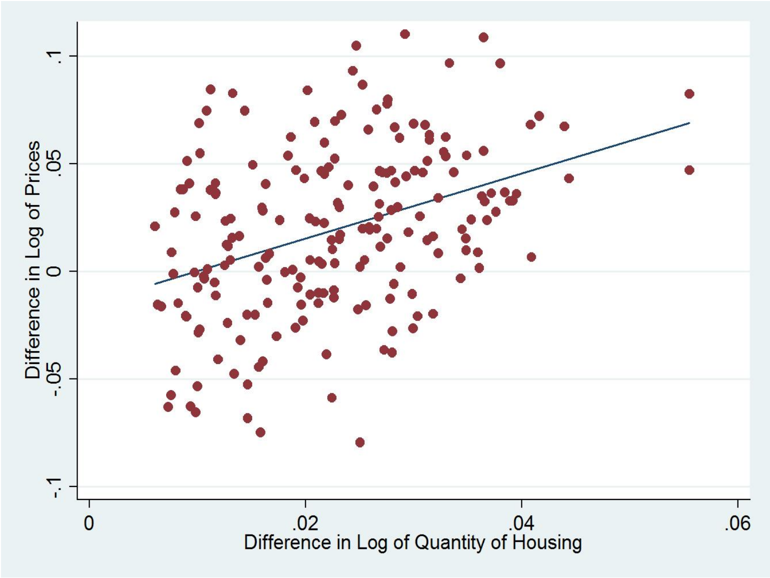

Figure 11: Relationship between the growth in Housing Prices and the growth in the Quantity of Housing

Section 3 of this paper is concerned with the models used in this paper. The general forms of our supply and demand equations are discussed. Moreover, theoretical predictions of the signs of coefficient estimates are discussed, and a closer look at certain variables is undertaken.

Section 4 is set aside to discuss the data collection process. Since one of this paper's key contributions is the construction of various new data series, it is necessary to mention all the assumptions and methods that went into data collecting. It is hoped that this will prove useful for other researchers studying housing markets in Canada.

Next, section 5 discusses the regression results from the model. An instrumental variables method is used to estimate the model. A number of variations of the model will be looked at in order to fully understand the significance of certain variables.

Finally, section 6 summarizes our findings and suggests future avenues of research. Three appendices have also been included in the paper. The first includes regressions performed on the level of the data rather than the difference in logs. The second contains the math behind the creation of the quantity of housing data series. The third contains additional graphical representations of the data.

2 Predecessors

The process of modeling housing prices has a long history. In the post-war era, interest in housing price modeling began in the 1960s. Early models were designed in the hope of forecasting future growth rates in the housing market. As research progressed, the study shifted to the study of housing prices. One of the first economists to do so was the

Canadian economist Lawrence B. Smith. He introduced a simple supply and demand equation to study the Canadian housing and mortgage market, with special attention to the role of financial variables (Smith 1969).

At the end of the 1970s, Kearl introduced the idea of user cost of housing (Kearl 1979). This was found to be an essential element in all further research on housing prices. The use of user cost in models extended throughout the 1980s and 1990s and is included in research today. Although the method of calculating costs has changed over time, it still maintains the basic elements introduced by Kearl.

Poterba presents a supply and demand model which uses Kearl's user cost of housing to more fully explore the housing market (Poterba 1984). He focused on developing the supply side of the model, especially the effects increased housing prices have on household construction. This paper represents the generally accepted form of housing models used throughout the 1980s. These models lacked lagged variables which were later shown to be of great importance. They also failed to model vacancies, which would become very important to models of the 1990s and 2000s.

This paper focuses on the newest generation of housing models developed by researchers such as Hwang, Quigley, DiPasquale, and Wheaton. Hwang and Quigley successfully applied a three equation model to the US housing market at a city level (Hwang and Quigley 2006). Besides a supply and demand equation, which will also be used in this paper, they had a third equation dealing with vacancies. In their opinion, vacancies occur when a person who recently bought a new home holds off on selling their old home in the hopes of obtaining a higher price in the future. This friction in their model was designed to capture the observed fact that it can often take several months, if not over a year for the housing market to clear. Unfortunately, this type of information is not publicly available in Canada, so at this stage it is not possible to replicate their third equation.

However, were the data available, it would not necessarily be prudent to add it to the model. Research shows that it is quite costly to own two houses for any period of time, thus it would be unlikely that a homeowner would purchase a second home before ensuring a buyer for the first home was obtained. Instead of waiting for a buyer to arrive that will pay the asking price, sellers will often lower the price in order to ensure a faster sale and avoiding paying higher taxes and user costs. However, evidence suggests that this lowering of prices only occurs after a period of time on the market. It is important that this market friction be captured in a full housing model, but at this time it is not entirely evident to this author that vacancy rates are the optimal way to accomplish this.

The supply side of the model in this paper is heavily influenced by the work of Mayer and Sommerville (2000). Before these two authors wrote their article, the supply side of housing models related the level of prices and the level of costs to the level of housing output. It was felt by earlier authors that this relationship satisfactorily explained the supply side of the economy. However, these two authors showed that it was more reasonable to link the supply of housing to the change of prices and the change of costs. This paper will use Mayer and Sommerville's results in the body of this paper when calculating the supply side of the model. However, the data will be the difference in the logged value of the variable rather than the simple difference. This modification still falls within the theory presented by the two authors.

DiPasquale and Wheaton were the first two to implement the gradual price adjustment nature of the housing market. Instead of looking at vacancies, they instead add a third equation of their own which states that change in prices are a function of a variable which represents the rate at which actual prices converge to equilibrium prices

(DiPasquale and Wheaton 1994). The authors also departed from previous supply side equations by making supply closely related to the land market. They were successful with their model in the sense that they showed, rather conclusively, that there is a need to model lags in housing price adjustments. Their research, which shows the significance of lags, influenced the model developed in this paper. Lagged changes in price and lagged changes in quantity will be added to various regressions to obtain a better fit to the data.

Clearly from these two papers it is essential that a model factor in some sort of mechanism which accounts for the frictions present in the housing market. It is also important to model the supply side of the economy in a theoretically correct manner, and one which also satisfactorily fits the data.

Though these two major papers focus on American data, there are housing market papers which look at Canadian data. Watuwa and Johnstone apply Canadian data to a housing market model in order to explain the movement in housing prices. Their paper discusses many of the problems that arise in obtaining data at the CMA level in Canada (Watuwa and Johnstone 2007). They were unable to find data on the stock of housing, a problem which is resolved in this paper. Where they proxy the data by using residential investment figures, our paper will explore an alternate route by constructing a data series. These authors also find that certain signs on some coefficients are the reverse of what is to be expected, a problem which, as will be made clear, was also faced in this paper. Watuwa and Johnstone also find that commodity prices play a large role on housing prices in certain resource rich provinces. This is a variable which is not included at this time, but in future papers might be added to obtain a more precise model.

Fortin and Leclerc's paper discussed the importance of demographics in housing price models at the provincial level. Again, using a supply and demand framework, the researchers discovered that any housing price model should include a variable for the age group between the age of 25 and 54(Fortin and Leclerc 2000). Their research shows that this proportion of the population is the only statistically significant contributor to the change in housing prices, since they are by far the largest group of people who purchase homes. Thus, the model in this paper includes a variable to capture this fraction, so as to better explain housing prices.

Finally, this paper draws upon the work of Allen, Amano, Byrne and Gregory (2006). They suggest that housing prices are determined locally, and it is for this reason that this paper focuses on CMAs as opposed to provinces. They also look into whether or not migration of labour into a city has a net positive or net negative effect on housing prices. We continue their work here by looking at the effect that net migration has on housing prices. Their research suggests that future models should include variables which look at land availability, institutions and future economic activity in the region. Again, as with the case of Watuwa and Johnstone's suggestion, this was not possible at this stage but could be added to models at a future date.

Most papers which have dealt with Canadian housing data have encountered a common problem. It is difficult to obtain data for all the variables that one would wish to include in their model. In order to overcome this problem, authors are forced to use various variables as proxys for the desired variable, or are forced to apply their model to provinces as opposed to CMAs. One of the big contributions of this paper is overcoming this problem. Part 4, discusses the data sources and assumptions used to obtain most of the desired variables at a CMA level. With all this previous research on hand, it is now possible to move on to the discussion of the model which will be used in this paper.

3 The Model

The model in this paper is a familiar two-equation supply and demand framework. A third equation, which has been used in the recent literature, could not be included in this paper since the vacancy rates for Canadian housing are not available at the CMA level. This paper's model simplifies into one where the owner-occupied housing stock is equal to the quantity of the housing stock in the home ownership market. As usual it is assumed that markets clear at an equilibrium price where $P_D = P_S$ and $D_t = S_t$. Where P_D is the price on the demand side of the economy, P_S is the price on the supply side of the economy, D_t is the demand for houses at time t, and S_t is the supply of houses at time t.

The demand equation for this model at time 't' is as follows:

$$D_t = D(P_t^*, X_t^D) \tag{1}$$

D_t is the quantity demanded for owner occupied houses, which is expected to be negatively related to prices. P_t^* is the market clearing price of owner-occupied houses in a CMA and is an endogenous variable in the equation. X_t^D represents a number of exogenous variables (demand shifters) which will be included in the model.

$$X_t^D = (UC_t, R_t, D_t, M_t, Y_t, U_t, P_{t-1}, PR_{t-1}, Dummy, Q_{t-1}) \tag{2}$$

Two variables in X_t^D are the user cost of owning a house (UC_t) and the rental rate of apartments (R_t). The user cost of housing is evaluated as a percentage of the value of the home. It is expected to be negatively related to prices. As the cost of owning increases, fewer people will purchase homes and demand will shift down. The rental price of apartments is included in the model as a representation of available substitutes.

It expected that it will be positively related to the price of houses.

One of the other variables in X_t^D is the demographics of the CMA (D_t). Following the advice of Fortin and LeClerc (2006), it was decided that the percentage of a population between the ages of 25 and 54 is an important indicator of the demand for houses in a model. It is expected that the larger the percentage of people between the ages of 25 and 54, the higher prices will be to clear the market. Along similar lines, a net migration variable (M_t) is included which represents increased demand for housing due to growth in population. A positive net migration would be expected to translate into a higher price for homes as the demand for accommodations shifted upwards.

Median household income (Y_t) and the unemployment rate (U_t) in a CMA are two variables that will be included in some regressions to capture the impact of the macro economy on housing prices. Income is expected to be positively related to prices. One would anticipate that a larger income should translate into a greater demand for higher priced homes. Unemployment on the other hand is expected to be negatively related to housing prices. The higher unemployment is in a city, the less money is available to spend on homes. Of equal importance is the increased ncertainty associated with unemployment. These two variables are designed to capture the varying economic conditions between CMAs in a given year.

Additionally, a variable which is the lagged ratio of housing prices to rental rates (PR_{t-1}) is included in the model. Hwang and Quigley(2006) had success including the ratio in their regression, as it represents the potential stream of earnings that owning a house could produce if rented out. It is expected to have a negative relationship with housing prices as the larger the ratio could indicate that housing is becoming relatively expensive compared to renting. As such, consumers may shift away from owning, and begin renting in order to avoid overpaying.

Finally, research seems to show that housing price growth is positively related to housing price growth from the previous period. As was discussed in Part 2, other papers in the past have also used the lagged price of housing (P_{t-1}) in their models with much success (DiPasquale and Wheaton 1994). It is for this reason that lagged prices are included in the model to capture this effect. In this case it is used to model the existence of frictions in the market. Research has shown that prices today are heavily dependent on previous prices in a CMA. Research has also shown that lagged quantity of housing (Q_{t-1}) is a potentially key variable in explaining housing prices, and thus it too will be included in the model.

Moving onto the supply side of the model, the second equation to be examined is based on the following general form:

$$S_t = S(P_t^*, X_t^S) \tag{3}$$

S_t is the quantity of houses supplied given the exogenous variables. P_t^* is the market clearing price of owner occupied homes at time t. Price is expected to be positively related to housing quantity, reflecting the incentive to increase productions of homes when prices increase.

$$X_t^S = (W_t, MC_t, I_t, P_{t-1}, Q_{t-1}, Dummy) \tag{4}$$

X_t^S represents exogenous supply shifters. Included amongst these variables are three variables which represent the costs involved in constructing new homes. A wage variable (W_t) captures the required pay of the workers who construct new homes. The material cost of new homes (MC_t) represents the physical materials required in building a home. Finally, the real interest rate (I_t) represents the financing cost of new home construction. All three of these variables are expected to be negatively related to housing quantity.

If the costs of new homes are higher, then fewer will be built.

Though there are not as many supply shifters as demand shifters in this model, one which may prove to be very important is lagged quantity (Q_{t-1}) Past results clearly create a strong case for including the variable in the regression. (Hwang and Quigley 2006). A second lag included in the supply equation is lagged prices (P_{t-1}). Unlike in the case of the demand equation, it is expected that lagged prices here will be negatively related to housing prices.

Finally, one variable which is included in both X_t^S and X_t^D is a dummy variable ($Dummy$) which takes the value of 1 for observations from 2001 onwards. In these years it has been shown that the housing market began to boom in North America. We may view this as a "bubble" that was unrelated to fundamentals. In an attempt to control for the price effect of this boom, this dummy will be used to ensure that accurate results are obtained for the other variables in the regressions.

Table 1 summarizes the variables presented in this section and the expected sign of their coefficients in a regression.

4 Data Collection

In order to apply a supply and demand model to Canadian data, it is important to be aware of what data exists. The nature of certain variables requires that data is procured from a variety of different sources. Another potential difficulty is that certain data is only available at the provincial level, rather than at the desired CMA level. In cases such as this, the available provincial data was applied to each city within that province.

Table 1: Expected sign of the coefficients of each variable in the regressions

Demand-side Variable	**Expected Sign**
Quantity (Q_t)	-
User Cost of Housing (UC_t)	-
Rental (R_t)	+
Demographics (D_t)	+
Migration (M_t)	+
Income (Y_t)	+
Unemployment (U_t)	-
Lagged Prices (P_{t-1})	+
Lagged Price to Rent Ratio (PR_{t-a})	-
Dummy variable for years 2001 and onwards ($Dummy$)	+
Lagged Quantity (Q_{t-1})	-
Supply-side Variable	**Expected Sign**
Price (P_t)	+
Construction Wages (W_t)	-
Material Costs (MC_t)	-
Real Interest Rates (I_t)	-
Lagged Prices (P_{t-1})	-
Lagged Quantity (Q_{t-1})	+
Dummy variable for years 2001 and onwards ($Dummy$)	+

Due to this unfortunate circumstance, certain variables are mainly useful at explaining the time series component of housing prices rather than the difference between CMAs.

The two most important variables when examining a supply and demand system of equations are of course quantity and price. The pricing data in this paper was generously provided by the Multiple Listing Service (MLS), a private organization of realtors. MLS provided their monthly index of housing prices for 18 Canadian cities from six provinces. These cities which are: Halifax, Saint John, Ottawa, Kingston, Oshawa, Toronto, Hamilton, St. Catharines, Kitchener, London, Windsor, Sudbury, Thunder Bay, Regina, Saskatoon, Calgary, Edmonton, and Vancouver. In order to transform the data to a yearly variable, the mean of the 12 monthly variables was calculated.

Owner Occupied Housing

Surprisingly, one of the most difficult pieces of data to find was the quantity of owner occupied housing in a CMA in a given year. The quantity was provided only in census years by Statistics Canada. In order to acquire information on the intervening years, a number of different calculations were attempted. The Canadian Mortgage and Housing Corporation (CMHC) provides yearly data on housing starts by intended sector as well as total housing completions. Simply adding the completions in each of the intervening years to the stock in the most recent census year failed to yield the stock of housing in the following census year. For instance in the case of Toronto between 1991 and 1996 the stock increased by approximately 80,000 houses, but according to data from the CMHC 95,000 houses were completed. This difference may be accounted for by homes which are destroyed or accommodations that are rented instead of owned. When data on starts for owner occupied houses were used instead of the data on completions, the results were also different from those which were reported in the census.

Table 2: Summary statistics of the data to be used in the model

Variable	No. of Obs.	Mean	Standard Deviation	Minimum	Maximum
Quantity	216	170486.9	226223.6	30221.92	1188205
User Cost of Housing	216	6.134613	4.893251	-18.5254	16.28872
Rental	216	721.7059	123.7064	482.1803	1050.102
Demographics	216	45.64851	1.999364	41.20697	49.84588
Migration	216	0.6165778	0.7972537	-1.733421	2.800984
Income	216	47419.97	5930.781	36113.66	61961.5
Unemployment	216	7.324035	1.633485	3.941667	12.875
Price	216	161223.9	58152.32	85554.74	396999.2
Wage	216	21.92927	3.795881	13.56301	28.23059
Material Cost of Housing	216	113067.4	15368.39	79110.27	159968.4
Interest Rate	216	2.538656	1.734548	0.4637646	5.649147
Lagged Price to Rent Ratio	198	215.0129	41.94772	151.1761	380.3339

After examining the data and attempting many different methods, a final decision was made on how to construct the data for the intervening years. Using the data on housing starts for the home ownership market, the sum of the starts for the five years between census periods was obtained. Each year's respective contribution to starts was divided by the sum to obtain a percentage of housing starts that occurred in each year compared to the whole five year period. Next, the quantity of housing stock between two census years was calculated. The difference between this value and the sum of housing starts over the same period was taken. If the difference was positive it meant that the housing starts did not properly account for all of the new houses that did in fact exist in the CMA. Therefore each year was assigned a proportion of this difference based on the relative proportion of housing starts in each given year. A year with more housing starts would be assigned a relatively larger proportion of the difference. If the difference had been negative, an appropriately weighted proportion would have been subtracted from each year. In this way the intervening years match up with the data provided by the census while still utilizing the data provided by housing starts. It was assumed that the census information was correct and that the housing starts simply needed an adjustment to bring it into line with the census data. In this respect, we are utilizing the growth rate of housing according to the census as well as CMHC's housing starts information. The formulas described above are included in Appendix 2.

Material Cost of Homes

The rest of the data was acquired either from Statistics Canada, the Bank of Canada or the CMHC. Many of the variables were readily available, while others required the combination of various data sources. For instance the material cost variable was calculated using two separate Statistic Canada data series. Using Cansim table 029-0005, the capital expenditures in the housing industry was calculated. This measure excluded expenditures on repairs, so was taken to be an accurate representation of expenditures on new homes. Then using Cansim table 027-0009, data on the number of housing starts, completions, and units still under construction was obtained. Since the sum of these three series would represent all the potential sources of expenditure, it was simple enough to divide one figure by the other to obtain an approximation of the material cost of new homes.

User Cost of Housing

Another piece of data required for the regressions was information on the user cost of housing. The user cost of housing was computed using the following equation:

$$UC_t = M_t * (1 - T_t^P) + D + MN_t - E_t(P) \tag{5}$$

Where M_t is the yearly average mortgage rate, T_t^P is the average provincial property tax rate, D is depreciation, MN_t is maintenance costs and $E_t(P)$ is the expected capital gains on the house. Since most of this data was only available on the provincial level, the user costs are not unique to a CMA but rather for each province. Mortgage rates were provided by the Bank of Canada on a monthly basis. The mean of the twelve data points in each year was calculated to achieve a yearly rate. Property tax rates were based on the consumer price index. The Canadian convention of a 1.5% depreciation rate on housing was used for all provinces.[1] The maintenance costs of housing were calculated by dividing the aggregate expenditures on all housing repairs in a province by the dollar value of housing within that province. Finally, expected capital gains were replaced by realized capital gains, thus assuming perfect foresight. This is clearly a rather strong assumption to be sure, but one which should have a limited impact on the results. In the case of the user cost of housing, a negative value would indicate

[1]This rate was taken from Statistic Canada's CPI calculations. A discussion on the logic behind the choice can be found at this web address: `http://www.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=62F0014MIE2001015&lang=eng`

that the expected increase in housing prices would offset the cost of maintaining the home. This would be the case in CMAs such as Edmonton and Calgary which were experiencing an economic boom in the time frame covered by the data.

With all of these data sets from a variety of sources it was difficult to find a continuous number of years where data from all data sets were available. A lack of data for a few key variables forced the data series to include only the years 1994 to 2005. With data for 18 CMAs over 12 years, this translates into 216 data points which was deemed sufficient to gauge the strength of the model. Note however, that once the differences in logged values are taken and lagged values are included, the resulting number of observations falls to either 198 or 180 depending on the presence of lags.

Other Variables

Certain data was much easier to come by. The real interest rate is simply the nominal yearly bank rate minus the CPI inflation from that year. Both pieces of data are available from Statistics Canada. The wage rate used in the supply equation was also easy to obtain. Statistics Canada has union wage rates for the construction industry going back many years. The data on workers classified as labourers was chosen as an appropriate proxy for the wage cost of new homes construction. Unfortunately, data on four cities were missing from the statistics Canada source, so data from the Ontario Ministry of Finance was used in conjunction with Statistics Canada data for the cities of Oshawa and Kingston. Data from the Saskatchewan Bureau of Statistics was used for the cities of Regina and Saskatoon.

Data for the rental cost of accommodation in each city was obtained from the CMHC. It is simply the average rent for a two bedroom apartment in each city. This was felt to be an appropriate representation for the cost of alternative forms of accommodation. For data on the median income for families, Cansim did contain data for most CMAs in the study. Under the assumption that there is one family per household, this data would provide the income per household figure that was sought. Unfortunately, it was not available for every CMA. For seven CMAs, only provincial level data was available and thus was used in place of CMA level data. These seven CMAs were: Saint John, Kingston, Oshawa, Sudbury, Thunder Bay, Regina and Saskatoon.

The migration variable was easily determined as the net flow of people entering or leaving a CMA for each year divided by the population of the city. A positive value represents a net inflow of migrants. This data was readily available from Statistics Canada. Demographic data for the CMAs was also obtained from Statistics Canada. It is simply the percentage of the population between the ages of 25 and 54 in each CMA for a given year. Unemployment rates for each CMA were also acquired from Statistics Canada. The unemployment rate in this paper is the generally accepted definition used by the government of Canada.

5 Results

With the basic model complete and the data collected, it is now possible to begin examining the data using econometric techniques. Unfortunately, in our model, it is not possible to simply regress prices and quantity on the exogenous variables. The fact that we wish to estimate two equations simultaneously means that we must use a more advanced technique. Our supply and demand equations constitute what is known as a linear simultaneous equation model (Davidson and MacKinnon 2004). This type of models does not allow us to use the simple OLS method of regressions since the dependent variables are correlated with the error terms. This is easily shown as follows. Our two equation model can be expressed as follows:

$$p_t = \alpha^D q_t + X_t^D \beta^D + \epsilon_t^D \tag{6}$$

$$p_t = \alpha^S q_t + X_t^S \beta^S + \epsilon_t^S \tag{7}$$

Where p_t is the price in period t, q_t is the quantity of housing in period t, X_t^D and X_t^S are defined above, ϵ_t^S and ϵ_t^D are error terms, and the α and β terms are the parameters to be estimated. It is possible to solve for p_t and q_t by putting the two equations into matrix form. We obtain the following result.

$$\begin{bmatrix} p_t \\ q_t \end{bmatrix} = \begin{bmatrix} 1 & -\alpha^D \\ 1 & -\alpha^S \end{bmatrix}^{-1} \left(\begin{bmatrix} X_t^D \beta^D \\ X_t^S \beta^S \end{bmatrix} + \begin{bmatrix} \epsilon_t^D \\ \epsilon_t^S \end{bmatrix} \right) \tag{8}$$

The above equation clearly shows how prices depend on all exogenous variables and the error terms from both equations. As such running and OLS regression would yield results which are both biased and inconsistent (Davidson and MacKinnon 2004).

To solve this problem, the method of instrumental variables (IV) will be used. This regression technique, also known as two staged least squares, relies on running two regressions to obtain results which are unbiased and consistent. For instance, if one wished to estimate a demand equation to obtain unbiased and consistent estimators, the first step would be to regress quantity on all available exogenous variables. Using the results of this regression, the fitted values of quantity would then be obtained by multiplying each exogenous variable by its estimated coefficient. Finally, we can then regress prices on these fitted quantity values as well as all demand shifters included in X_t^D. The result would be a demand equation which expressed price as a function of quantity and demand shifters. The estimates would be unbiased and consistent. Calculating the appropriate supply equation would then proceed in a similar manner, but prices would be regressed on all exogenous variables, then its fitted values calculated. Finally, quantity of housing would be regressed on fitted prices to obtain the unbiased and consistent estimate of the supply equation.

The reason that this IV method is able to deliver unbiased estimates relies on the exogeneity of the variables in the first regression. We can reasonably assume that the demand shifters and supply shifters are uncorrelated with the error terms. Since this is the case, the first regression of the two is able to correct for the impact of the error terms on price and quantity. Obviously, it is not possible to go into all the details behind the IV method of estimation in this paper. For more detail on the process, consult any modern econometric text book.

The basic approached used while deciding what the final model would be went as follows. A very basic equation was used for each of the supply and demand equation in the first model. Variables would be removed in a subsequent stage if they were insignificant or to simply examine the effect that such a removal would have. As more models were examined, it was possible to determine which variables were the most significant to determining housing prices by the strength of the t-statistic in each model. This process continued until it was believed that the model which best fit the data was obtained. Such a model would have a significant R-squared value, and would have statistically significant coefficients.

There is a concern as to whether the variables in the regression should be the level of data in a given year, or whether the differences in logged values should be used. The results presented in the body of the text are those obtained when using logged differences, which is in fact growth rates. The results from using levels will be presented in Appendix 1. With many of the variables such as housing prices and income trending

Table 3: Summary statistics of the growth rate of the variables used in the model

Variable	No. of Obs.	Mean	Standard Deviation	Minimum	Maximum
Quantity	198	0.0226553	0.0099175	0.0060663	0.0555429
Rental	198	0.0006664	0.0209373	-0.0432363	0.0974989
Demographics	198	-0.0003339	0.0035482	-0.0089207	0.0106556
Income	198	0.0165262	0.0350218	-0.0902176	0.1480532
Unemployment	198	-0.0391891	0.1272862	-0.3275151	0.3267081
Price	198	0.019382	0.0391616	-0.0794268	0.1101408
Wage	198	-0.0034421	0.0172468	-0.0276153	0.0914636
Material Cost of Housing	198	0.002531	0.0642398	-0.1441669	0.2044601
Interest Rate	198	-0.1815089	0.5227007	-1.296394	0.7940613

upwards over time, spurious correlation may be present in the regression output. By using growth rates, we hope to avoid this potential problem but at the cost of a smaller number of observations. Note that all variables used are in constant 2002 dollars.

It should also be noted that all the variables in the following regressions are growth rates except for three. Firstly, user cost is sometimes negative and thus would create obvious problems when attempting to apply the natural logarithm to it. Therefore, it is simply the difference between the two years rather than the difference of logged values. Secondly, migration is simply the stock variable rather than the logged differences. It suffers from the same problem as user cost, but since it is already a flow variable rather than a level, it was deemed unnecessary to take a difference. Finally, lagged price to rent ratio is kept as a level.

The first model estimated is a basic one, which will be built upon one step at a time. The following two equations make up model 1.

$$P_t = \beta_0 + \beta_1 Q_t + \beta_2 UC_t + \beta_3 R_t + \epsilon_t \tag{9}$$

$$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 W_t + \alpha_3 MC_t + \alpha_4 I_t + \upsilon_t \tag{10}$$

Table 4: Demand equation regression results for variation 1 to 5

Model	(1)	(2)	(3)	(4)	(5)
	P_t	P_t	P_t	P_t	P_t
Q_t	0.806	-3.429	-5.834	-2.238	-1.289
	(0.42)	(2.15)*	(4.41)**	(1.41)	(0.86)
R_t	0.365	0.319	0.434	0.378	0.342
	(2.26)*	(2.42)*	(3.51)**	(3.24)**	(3.02)**
UC_t	0.003	0.002	0.002	0.003	0.003
	(2.58)**	(2.63)**	(2.36)*	(3.63)**	(3.97)**
Q_{t-1}		4.134	6.179	2.898	2.029
		(3.00)**	(5.38)**	(2.08)*	(1.54)
Y_t			-0.145	-0.046	
			(2.13)*	(0.69)	
Dummy				0.026	0.032
				(4.43)**	(5.62)**
U_t					-0.038
					(2.14)*
Constant	0.005	0.011	0.023	-0.000	-0.007
	(0.11)	(1.31)	(2.90)**	(0.01)	(0.81)
Observations	198	180	180	180	180
Number of Cities	18	18	18	18	18
R^2 of 1st stage	.0541	.8100	.8179	.8191	.8195
R^2 of 2nd stage	.1278	.2131	.2863	.3879	.3998

Absolute value of z statistics in parentheses
*significant at 5%; **significant at 1%

If we examine the results from these regression in tables 4 to 7, it is obvious that this simple equation is unsatisfactory. First of all, the first stage regression on the demand size has an extremely low R-squared value, indicating that the instruments are not valid in this case. The result is a poor second stage fit. Few variables are significant, and some obviously have a theoretically incorrect sign, most notably the user cost of capital and the quantity demanded in the demand curve. According to these results, there is an upward sloped demand curve for housing! On the supply side of the equation the wage and interest rate prove to be insignificant while material costs is significant. All variables display the theoretically correct sign, yet there is a rather low value for R squared. As such, the supply equation will be developed further by dropping the two insignificant variables and adding other exogenous variables in an effort to improve the model.

In an attempt to correct these problems, a second model was set up which would include more instruments to obtain a better first stage fit. Without a better first stage

Table 5: Demand equation regression results for variation 6 to 10

Model	(6)	(7)	(8)	(9)	(10)
	P_t	P_t	P_t	P_t	P_t
Q_t	1.052	0.971	1.038	1.048	2.106
	(4.17)**	(3.83)**	(3.83)**	(3.72)**	(1.26)
R_t	0.274	0.278	0.279	0.279	0.340
	(2.57)*	(2.59)**	(2.59)**	(2.58)**	(2.36)*
UC_t	0.003	0.003	0.003	0.003	0.003
	(4.55)**	(4.60)**	(4.52)**	(4.43)**	(2.82)**
U_t	-0.040	-0.042	-0.040	-0.041	-0.040
	(2.25)*	(2.34)*	(2.22)*	(2.28)*	(2.17)*
$Dummy$	0.038	0.040	0.038	0.038	0.039
	(8.36)**	(7.76)**	(8.30)**	(8.26)**	(6.85)**
D_t		0.675			
		(0.88)			
M_t			-0.002		
			(0.65)		
PR_{t-1}				-0.000	
				(0.59)	
Constant	-0.018	-0.017	-0.017	-0.011	-0.044
	(2.75)**	(2.53)*	(2.51)*	(0.94)	(1.17)
Observations	180	180	180	180	180
Number of Cities	18	18	18	18	18
R^2 of 1st stage	.8135	.8228	.8213	.8201	.0825
R^2 of 2nd stage	.3991	.3959	.3927	.3924	.3930

Absolute value of z statistics in parentheses
*significant at 5%; **significant at 1%

Table 6: Supply equation regression results for variation 1 to 5

Model	(1)	(2)	(3)	(4)	(5)
	Q_t	Q_t	Q_t	Q_t	Q_t
P_t	0.110	0.418	0.262	0.205	0.136
	(4.60)**	(12.94)**	(6.42)**	(5.77)**	(4.16)**
W_t	-0.020				
	(0.91)				
MC_t	-0.018	-0.053	-0.022	-0.007	-0.020
	(3.18)**	(6.12)**	(2.91)**	(0.89)	(2.21)*
I_t	-0.001				
	(1.16)				
P_{t-1}			-0.071	-0.025	-0.012
			(3.41)**	(1.43)	(0.59)
$Dummy$				-0.005	
				(4.16)**	
Constant	0.020	0.013	0.018	0.021	0.020
	(9.74)**	(13.58)**	(16.79)**	(21.80)**	(20.63)**
Observations	198	180	180	180	180
Number of Cities	18	18	18	18	18
R^2 of 1st stage	.1451	.2131	.3304	.4201	.4267
R^2 of 2nd stage	.0539	.5516	.3611	.3252	.2060

Absolute value of z statistics in parentheses
*significant at 5%; **significant at 1%

Table 7: Supply equation regression results for variation 6 to 10

Model	(6)	(7)	(8)	(9)	(10)
	Q_t	Q_t	Q_t	Q_t	Q_t
P_t	-0.020	0.002	-0.004	-0.005	0.075
	(1.12)	(0.10)	(0.19)	(0.23)	(5.63)**
MC_t	-0.026	-0.024	-0.024	-0.024	-0.021
	(4.53)**	(4.22)**	(4.20)**	(4.19)**	(4.00)**
Q_{t-1}	0.914	0.924	0.928	0.929	
	(22.90)**	(23.87)**	(23.95)**	(23.97)**	
P_{t-1}		-0.025	-0.023	-0.023	
		(2.02)*	(1.82)	(1.81)	
Constant	0.003	0.003	0.003	0.003	0.021
	(3.48)**	(3.06)**	(3.07)**	(3.08)**	(10.39)**
Observations	180	180	180	180	180
Number of Cities	18	18	18	18	18
R^2 of 1st stage	.3993	.4302	.4268	.4278	.3930
R^2 of 2nd stage	.8085	.8143	.8143	.8143	.0740

Absolute value of z statistics in parentheses
*significant at 5%; **significant at 1%

fit, the effect of quantity and price on one another would just be random noise. The second model adds lagged quantity to the demand equation and removes wages and interest rates from the supply equation.

Adding the lagged value of quantity into the regression, clearly is a useful instrument since it increases the R squared on the first stage regressions to values over .8. A negatively sloped demand curve is also obtained thanks to the addition of lagged quantity on the demand side of the equation. Removing the insignificant variables of wage and interest rate kept the supply equation the way theory would predict, albeit with a low R-squared value.

Some problems however still exist with the demand equation. The user cost and lagged quantity terms both have the opposite sign to what would be expected. It is difficult to believe that as the cost of owning a home increased, more people would wish to own one. However, it might be the case that current prices predict future prices. At the same time it is odd that an increase in the growth rate of homes in the previous year would cause the demand for homes to increase. The theory behind the addition of lagged quantity in the demand equation is that there is a delay in the time it takes for prices to adjust to past information. A large change in the quantity of housing in the previous period may indicate that a housing boom is in effect. Thus prices increase on the demand side perhaps from a speculative purpose. Clearly this model leaves something to be desired.

The 3rd model adds a variable to each equation in order to capture some more information. Income is added to the demand equation to pick up macroeconomic factors which affect the market, and lagged prices are added to the supply equation to model the slow adjustment of housing prices. Adding income to the demand equation results in it taking a statistically significant negative value. This fact once again clashes with theory, which would predict that an increase in income would shift the demand curve out causing an increase in housing prices. With three of the five variables showing the opposite sign to what would be expected, it is clear that this demand equation is not adequate.

However, on the supply side, the addition of lagged prices has resulted in the expected outcome. There is a negative correlation between lagged prices and current quantity on the supply side.

The 4th model adds a simple dummy term to both equations for the years 2001 onwards to reflect the housing boom of the early 21st century. The dummy term performs as expected in the demand equation, being both positive and significant. It also reveals that the income variable is insignificant when this dummy is added, and thus will be removed from further progressions of the model. The supply side however, does not react favourably to the addition of the dummy variable. Theory would also indicate that the boom in housing prices was a demand side phenomenon, so adding the dummy to the supply equation would not be a valid approach.

For this reason, model 5 removes the dummy variable on the supply side. In order to capture the effect of the macro economy that income had previously been representing, the unemployment rate will be added to the demand side in its stead.

The regression results indicate that the unemployment term is both negative and significant as was expected. A higher unemployment rate will clearly shift the demand curve down, driving down housing prices. The unemployment rate however causes the lagged quantity to become insignificant in the demand equation and lagged prices to become insignificant in the supply side equation. In an attempt to model the market better, we follow the lead of Hwang and Quigley and move lagged quantity to the supply equation. Lagged prices are removed from the supply equation but are added once again in a future model.

The results in tables 5 and 7 indicate that lagged quantity is clearly significant in the supply side equation with a massive t-statistic. Unfortunately, the removal of lagged quantity from the demand side has resulted in the demand curve once again being positively sloped. What is worse is that the supply curve became negatively sloped, although the coefficient is statistically insignificantly. Clearly, these results are not desirable, so a seventh model is required to hopefully correct for these unfortunate effects. It is quite clear however, that lagged quantity does belong in the supply equation, since it is quite significant and raises the R-squared significantly.

Thus the 7th model keeps lagged quantity in the supply side but also reintroduces lagged prices to the supply equation. Another shift variable is also added to the demand equation in this model, that of demographics. Unfortunately, demographics is insignificant and is removed from future regressions. Lagged prices are significant in the supply equation, but the coefficient on price is insignificant.

Model 8 attempts to improve the model by adding migration in place of demographics in the demand equation in the hopes that it will be a useful instrument. The supply equation remains unchanged. Unfortunately, the addition of migration to the equation does not have the desired results. It is insignificant, and thus will be removed from the next model. The same problems remain in model 8 as in model 7.

Since Hwang and Quigley had success adding a price to rent ratio term to their demand equation, model 9 includes the same variable in the hope of replicating their success.(Hwang and Quigley 2006) The results show that it clearly does not have the same effect in Canadian data as it does in U.S. data. The same problems remain from the two previous regressions.

Finally, model 10, is simply the supply equation from model 2 combined with the demand equation from the 6th model. It was felt that these two equations represented the best supply and demand equation from the various models. However, removing the lagged values clearly causes the first stage regression to become insignificant, so this model is also a failure.

In the end, none of these models seems to fit the data as we would expect it to. It is difficult to reconcile the fact that the demand curve will often appear to be upward sloping when theory clearly points against this. The fact that the user cost of capital is always positive is also peculiar. A higher cost of ownership should drive demand down, not up. Clearly many unanswered questions remain from these results.

The problem is that the data seems to reflect a horizontal demand curve. The movements of price and quantity seem to be mostly demand driven. The result is that a positive relationship between price and quantity are revealed in the regression for the demand side of the economy. The horizontal demand curve will shift up and down causing prices and quantity to move in the same direction. Data on supply shifts seems to be rather limited. The result is that we can get a clear picture of the supply curve, by observing shifts in the demand curve, but with no shifts in the supply curve, the demand curve remains ambiguous. The only way to solve this problem would be to introduce additional instruments which capture shifts in the supply curve. But at this time, it is unknown what these variables may be. Another problem may be the limitations of the available data. At this time the data set consists of only an eleven year period over eighteen CMAs. The result is one hundred and eighty observations. Although that is not an insubstantial amount, it may still be insignificant to obtain a clear picture of the housing market.

Despite these shortcomings there is a lot that can be taken away from these regressions. Clearly certain variables perform well and should be included in future models. On the demand side, the cost of renting is almost always positive and significant. Its role as a substitute for housing is clearly an important variable to potential homeowners. The unemployment rate seems to be a better predictor of housing prices then income. This is not entirely unexpected as unemployment is more sensitive to business cycles. If the transition to unemployment is interpreted as an enormous decrease in income, it is easy to see why the demand for housing would drop significantly. A reduction in income on the other hand would not necessarily cause homeowners to sell their homes, but perhaps cut back in other areas instead. Finally, the dummy variable for 2001 onwards is quite useful at showing the break that occurred at the turn of the century, where confidence in housing increased at a substantial rate.

On the supply side of the model, material costs seemed to be more significant than wages and interest rates in capturing the cost of building a home. Lagged prices and lagged quantity are also significant in most supply side equation, indicating that they both should be included. Unfortunately, the inclusion of these variables, specifically lagged quantity, seems to affect the significance of price in the supply equation. This problem can hopefully be resolved in the future.

A similar process of modeling was conducted with levels as opposed to logged differences. Some interesting results were obtained, but due to some theoretical problems which have yet to be resolved, these results are relegated to Appendix 1 and should not be considered an accurate representation of the Canadian economy at this time.

6 Conclusions

The current supply and demand model for housing prices being used in the literature has been successfully applied to data for the United States. The purpose of this paper was to apply this model to the data on Canadian CMAs in an attempt to replicate this success. Unfortunately, at this time with the current data available, it must be concluded that this model does not successfully explain the Canadian housing market.

The regression models explored in part 5 of this paper clearly leave much to be desired. Results are obtained where the demand curve is upward sloping in certain regressions. Also, certain key variables, such as the user cost of housing, have a statistically significant positive coefficient, when theory clearly states that a negative value is expected. At this stage, more data is believed to be necessary in order to obtain more concrete results.

This upward sloping demand curve is a puzzle that has been found in other papers as well such as in certain models in DiPasquale and Wheaton (1994). The data would seem to suggest that housing price growth is positively related to quantity growth in the demand curve. This result could be an indication that current price growth is a predictor of future price growth. This would imply that there is some sort of bubble effect driving market prices, as demand would be driven by speculative motives. The other possibility is that the demand and supply framework is insufficient when it comes to modeling the housing market. Since houses act as both a durable good and an investment good, we may need a more sophisticated model to explain how prices are determined.

These negative outcomes aside, there are many positive outcomes to be had from this paper. This is the first time data on some of these variables have been calculated on the CMA level for Canada. With the availability of this data, it is believed that future research on the CMA level of housing prices can be conducted. Papers such as Allen et al.(2006) clearly show that house pricing models must be approached at the city level in Canada do to the lack of a national housing market.

The regressions performed in part 5 also show that certain variables are consistently more significant than others, and thus provides an obvious starting point for future regressions. Certainly, with the availability of this new data, more research on the subject is to follow.

7 Bibliography

Allen, Amano, Byrne, and Gregory. 2006. "Canadian City Housing Prices and Urban Market Segmentation." *Bank of Canada Working Paper 2006-49:* 1-21.

Anonymous, 2002. "Pre-Budget Submission: The Opportunity for Action." *Canadian Home Builders' Association:* 1-38.

Bruce, David, and Tom Carter. 2003. *Literature Review of Socio-economic Trends Affecting Consumers and Housing Markets*, Ottawa, Canadian Housing and Mortgage Corporation.

Canadian Press. 2006. "Population boom creating problems in Alberta." *CTV.ca* April 2nd 2006: 1-2.

Davidson, Russell, and James G. Mackinnon. 2004. *Econometric Theory and Method*, New York, Oxford University Press.

DiPasquale, Denise, and William C. Wheaton. 1994. "Housing Market Dynamics and the Future of Housing Prices." *Journal of Urban Economics 35:* 1-27.

Duhamel, Bruno. 2005. "Variations in Housing Prices in Canada." *Canadian Mortgage and Housing Corporation Socio-Economic Series 05-014:* 1-3.

Engelhardt, Gary V., and James M. Poterba. 1991. "House prices and demographic change Canadian evidence." *Regional Science and Urban Economics 21:* 539-546.

Fortin, Mario, and Andr Leclerc. 2000. "Demographic Changes and Real Housing Prices in Canada." *Universit de Sherbrooke Working Paper 00-06:* 1-36.

Gupta, Kamal. 2001. "Housing Transition in Single Industry 'Instant Towns'." *Canadian Mortgage and Housing Corporation Socio-Economic Series 96:* 1-3.

Hosios, Arthur J., and James E. Pesando. 1991. "Measuring Prices in Resale Housing Markets in Canada: Evidence and Implications." *Journal of Housing Economics 1:* 303-317.

Hwang, Min, and John M. Quigley. 2006. "Economic Fundamentals in Local Housing Markets: Evidence from U.S. Metropolitan Regions." *Journal of Regional Science Vol 46, No. 3:* 425-453.

Kearl, J.R. 1979. "Inflation, Mortgages, and Housing." *Journal of Political Economy Vol 87, No 5:* 1115-1138.

Ley, David, and Judith Tutchener. 1999. "Immigration and Metropolitan House Prices in Canada." *Research on Immigration and Integration in the Metropolis Working Paper 99-09:* 1-34.

Ley, Tutchener, and Cunningham. 2001. "Immigration, Polarization, or Gentrification? Accounting for Changing Housing Prices and Dwelling Values in Gateway Cities." *Research on Immigration and Integration in the Metropolis Working Paper 01-20:* 1-29.

Mayer, Christopher J., and C. Tsuriel Somerville. 2000. "Land Use Regulation and New Construction." *Regional Science and Urban Economics 30:* 639-662.

Li, Prud'homme, Yu. 2008. "Studies in Hedonic Resale Housing Price Indexes." *Working Paper:* 1-24.

Poterba, James. 1984. "Tax Subsidies to Owner-Occupied Housing: An Asset-Market Approach." *The Quaterly Journal of Economics November 1984:* 729-752.

Smith, Lawrence B. 1969. "A Model of the Canadian Housing and Mortgage Markets." *Journal of Political Economy Vol 77, No. 5:* 795-816.

Watuwa, Richard, and Harvey Johnstone. 2007. "House Prices in Canada: An Empirical Investigation." *CEA 2007:* 1-20.

A Appendix 1

This appendix deals with regressions of the data as levels rather than the growth rate. As stated earlier, there is the potential problem that as time series data moves along, certain variables are naturally increasing. As a result, there may appear to be a positive correlation between certain variables when none in fact exist.

In the case of this data, it is not entirely clear if this problem exists. Certainly some variables such as income and prices would trend upwards together, but as the regression output shows, this does not seem to be the case. Income is not statistically significant in any model in this section. The model in this section is the same as that in the core of the paper, except that the data is not the difference of logs. The two regressions equations are:

$$P_t = \beta_0 + \beta_1 Q_t + \beta_2 X_t^D + \epsilon_t \tag{11}$$

$$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 X_t^S + \upsilon_t \tag{12}$$

Where P_t and Q_t represent prices and quantity respectively, and X_t^D and X_t^S represent exogenous shift variables for demand and supply. A basic model was regressed and variables were either removed if they were insignificant or damaging to the overall regressions, or kept if they added significantly to the model. Tables 8 through **??** summarize the regression results

Certain facts stand out when comparing these results to those of differences in logs. Firstly, certain variables are not significant and are not present in the regressions. Income, unemployment rate, and user cost of housing are all insignificant in these models. These variables are added in certain models, but are never significant. This stands in stark contrast to the regression of logged differences. In those cases both user cost and unemployment rate were significant in explaining housing prices. Rental rates and demographics are both significant in this model. Rental rates being significant is not surprising considering the results from Part 5. It was unexpected to see demographics in a successful model considering their failure to explain anything in the differences of logs model. The lagged price to rent term that Hwang and Quigley found useful in their regressions, also seem to perform quite well in model 13 of this section. Certain problems concerning the signs of variables occurred in the level models. For instance, a positively sloped demand curve was often the findings in early models, until additional variables revealed its true negative slope. Finally, in terms of the demand equation, it is noteworthy that the models in this section display high R-squared values compared with the differences in logs equation. Using the level variables seems to be a better fit for the data.

On the supply side of the model the results were equally as promising. It was determined that unlike in the differences of logs model, the wages of workers were more significant than the cost of materials when it came to deciding how many houses to construct. The lagged values of quantity and prices which had worked well in the growth version of the model do not have the same impact as they do in the level regressions. Although the removal of these variables does lower the R squared value of the supply regressions, the addition of other variables are either insignificant or have unfortunate effects on the coefficients and standard errors of other variables.

In the end, the model that best seems to fit the data is model 13. All but one variable is highly significant and the signs of the coefficients are all in the theoretically correct direction. A high R squared value is also present for the demand equation. Despite these seemingly positive results, one must remember to be cautious when dealing with housing data as levels as opposed to growth rates. Without fully understanding the implications, it is difficult to draw appropriate conclusions from the results.

Table 8: Demand side regression results using levels for variation 1 to 5

Model	(1)	(2)	(3)	(4)	(5)
	P_t	P_t	P_t	P_t	P_t
Q_t	0.444	0.343	0.449	0.222	0.212
	(3.59)**	(3.57)**	(2.55)*	(1.73)	(1.56)
R_t	139.173	185.084	137.000	208.567	208.003
	(2.40)*	(3.78)**	(2.15)*	(3.66)**	(3.60)**
UC_t	-283.810				
	(0.72)				
U_t		-1,578.139			-458.163
		(1.61)			(0.53)
Y_t			0.167		
			(0.22)		
$Dummy$				11,486.851	11,217.723
				(3.18)**	(3.12)**
Constant	-13,107.733	-19,221.307	-22,171.811	-31,968.557	-26,278.713
	(0.51)	(0.72)	(0.46)	(1.22)	(0.98)
Observations	216	216	216	216	216
Number of Cities	18	18	18	18	18
R^2 of 1st stage	.3397	.2202	.4292	.3206	.3172
R^2 of 2nd stage	.7202	.7052	.7083	.7473	.7471

Absolute value of z statistics in parentheses
∗significant at 5%; ∗∗significant at 1%

Table 9: Demand side regression results using levels for variation 6 to 10

Model	(6)	(7)	(8)	(9)	(10)
	P_t	P_t	P_t	P_t	P_t
Q_t	-0.320	0.226	0.187	0.186	0.221
	(0.29)	(1.99)*	(1.49)	(1.49)	(1.72)
R_t	428.789	206.250	237.902	236.394	207.470
	(0.91)	(3.95)**	(4.19)**	(4.13)**	(3.63)**
$Dummy$	24,124.310	11,404.354	10,547.619	10,780.329	11,741.014
	(0.94)	(3.46)**	(2.90)**	(3.00)**	(3.29)**
D_t			-5,564.775	-5,468.272	
			(2.37)*	(2.32)*	
M_t				-2,400.009	-2,861.019
				(1.01)	(1.19)
Constant	-103,705.565	-30,948.845	207,304.311	205,489.782	-29,264.350
	(0.64)	(1.22)	(1.97)*	(1.95)	(1.10)
Observations	216	216	216	216	216
Number of Cities	18	18	18	18	18
R^2 of 1st stage	.4795	.3473	.3245	.3337	.3318
R^2 of 2nd stage	.7355	.7483	.6712	.6707	.7500

Absolute value of z statistics in parentheses
∗significant at 5%; ∗∗significant at 1%

Table 10: Demand side regression results using levels for variation 11 to 14

Model	(11)	(12)	(13)	(14)
	P_t	P_t	P_t	P_t
Q_t	0.912	0.166	0.104	0.176
	(16.57)**	(6.54)**	(3.66)**	(1.17)
R_t	-69.606	211.690	208.176	240.555
	(2.30)*	(6.39)**	(18.93)**	(3.60)**
Dummy	-7,044.767	11,462.546	3,709.102	8,274.979
	(3.74)**	(5.45)**	(2.77)**	(2.76)**
D_t	-8,950.049	-7,572.033	-1,119.865	-6,117.799
	(5.54)**	(3.16)**	(1.78)	(2.59)**
PR_{t-1}			807.282	
			(28.07)**	
UC_t				-363.719
				(0.90)
Constant	467,510.332	320,818.915	-130,625.186	235,667.654
	(6.42)**	(3.02)**	(4.98)**	(2.27)*
Observations	216	198	198	216
Number of Cities	18	18	18	18
R^2 of 1st stage	.4117	.9999	.2907	.3641
R^2 of 2nd stage	.6612	.6359	.9801	.6457

Absolute value of z statistics in parentheses
*significant at 5%; **significant at 1%

Table 11: Supply side regression results using levels for variation 1 to 5

Model	(1)	(2)	(3)	(4)	(5)
	Q_t	Q_t	Q_t	Q_t	Q_t
P_t	1.442	1.363	1.473	1.480	1.326
	(8.10)**	(8.57)**	(9.91)**	(6.93)**	(9.31)**
W_t	-5,192.205	-6,518.457	-4,459.531	-6,495.732	-5,411.409
	(1.82)	(2.29)*	(1.61)	(2.19)*	(1.93)
MC_t	-0.029				
	(0.15)				
I_t	157.667				
	(0.11)				
Dummy				-6,224.441	
				(1.00)	
Constant	54,712.950	93,670.033	30,797.624	76,862.227	75,400.337
	(0.66)	(1.09)	(0.37)	(0.92)	(0.90)
Observations	216	216	216	216	216
Number of Cities	18	18	18	18	18
R^2 of 1st stage	.7318	.7052	.7083	.7473	.7471
R^2 of 2nd stage	.3444	.2250	.4091	.3206	.3231

Absolute value of z statistics in parentheses
*significant at 5%; **significant at 1%

Table 12: Supply side regression results using levels for variation 6 to 10

Model	(6)	(7)	(8)	(9)	(10)
	Q_t	Q_t	Q_t	Q_t	Q_t
P_t	1.396	1.279	1.283	1.272	1.308
	(9.95)**	(7.12)**	(9.10)**	(9.02)**	(9.20)**
W_t		-4,907.914	-5,773.000	-5,870.023	-5,649.160
		(1.68)	(2.05)*	(2.08)*	(2.01)*
MC_t	-0.132				
	(0.66)				
I_t		-799.102			
		(0.54)			
Constant	-39,645.125	73,989.612	90,262.406	94,070.836	83,500.340
	(0.74)	(0.89)	(1.06)	(1.10)	(1.00)
Observations	216	216	216	216	216
Number of Cities	18	18	18	18	18
R^2 of 1st stage	.7355	.7483	.6712	.6707	.7500
R^2 of 2nd stage	.4778	.3404	.2373	.2212	.3043

Absolute value of z statistics in parentheses
*significant at 5%; **significant at 1%

Table 13: Supply side equation regression results using levels for variation 11 to 14

Model	(11)	(12)	(13)	(14)
	Q_t	Q_t	Q_t	Q_t
P_t	2.419	0.024	0.850	1.338
	(3.28)**	(2.33)*	(7.45)**	(9.82)**
W_t	-5,442.284	-160.572	-10,488.277	-5,158.972
	(1.77)	(2.16)*	(3.31)**	(1.87)
P_{t-1}	-1.566			
	(1.89)			
Q_{t-1}		1.030		
		(388.44)**		
Constant	146,930.677	-687.832	264,828.001	67,878.971
	(1.85)	(0.35)	(3.03)**	(0.81)
Observations	216	198	198	216
Number of Cities	18	18	18	18
R^2 of 1st stage	.9877	.6387	.9801	.6457
R^2 of 2nd stage	.3550	.9999	.1383	.2653

Absolute value of z statistics in parentheses
*significant at 5%; **significant at 1%

B Appendix 2

To calculate the stock of housing we use the following formulas.

Define the following:

H_0 = Housing stock according to the census in year 0

H_5 = Housing stock according to the next census in year 5

C_i = Construction of houses in year i, where $i = (1, 5)$

S_i = Stock of houses in year i, where $i = (0, 5)$

As stated above in Part 4, we set $S_0 = H_0$ and $S_5 = H_5$. We then use the following four formulas to find S_1 through to S_4:

$$S_1 = S_0+C_1+([(H_5-H_0)-(C_1+C_2+C_3+C_4+C_5)]*C_1/(C_1+C_2+C_3+C_4+C_5)) \quad (13)$$

$$S_2 = S_1+C_2+([(H_5-H_0)-(C_1+C_2+C_3+C_4+C_5)]*C_2/(C_1+C_2+C_3+C_4+C_5)) \quad (14)$$

$$S_3 = S_2+C_3+([(H_5-H_0)-(C_1+C_2+C_3+C_4+C_5)]*C_3/(C_1+C_2+C_3+C_4+C_5)) \quad (15)$$

$$S_4 = S_3+C_4+([(H_5-H_0)-(C_1+C_2+C_3+C_4+C_5)]*C_4/(C_1+C_2+C_3+C_4+C_5)) \quad (16)$$

Of course using these formulas it is easy to see that it will work out such that:

$$S_5 = S_4+C_5+([(H_5-H_0)-(C_1+C_2+C_3+C_4+C_5)]*C_5/(C_1+C_2+C_3+C_4+C_5)) = H_5 \quad (17)$$

C Appendix 3

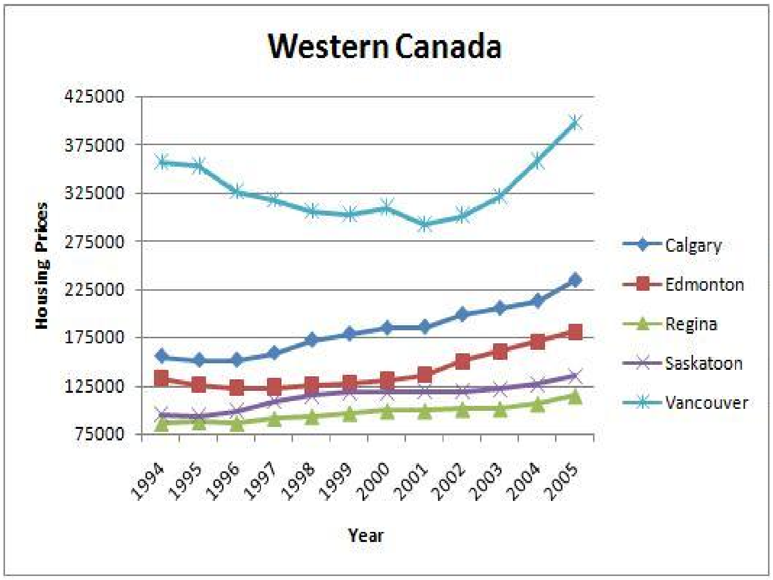

Figure 12: Relationship between Housing Prices and Time in Western Canada

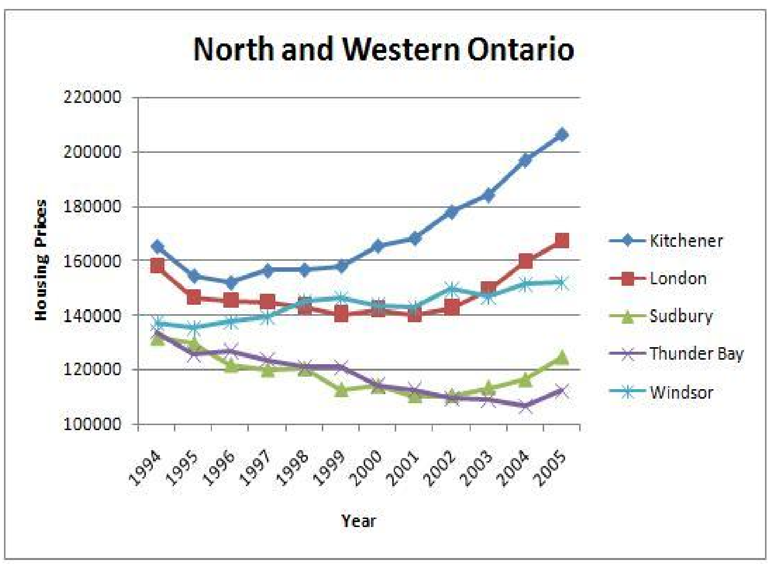

Figure 13: Relationship between Housing Prices and Time in North and Western Ontario

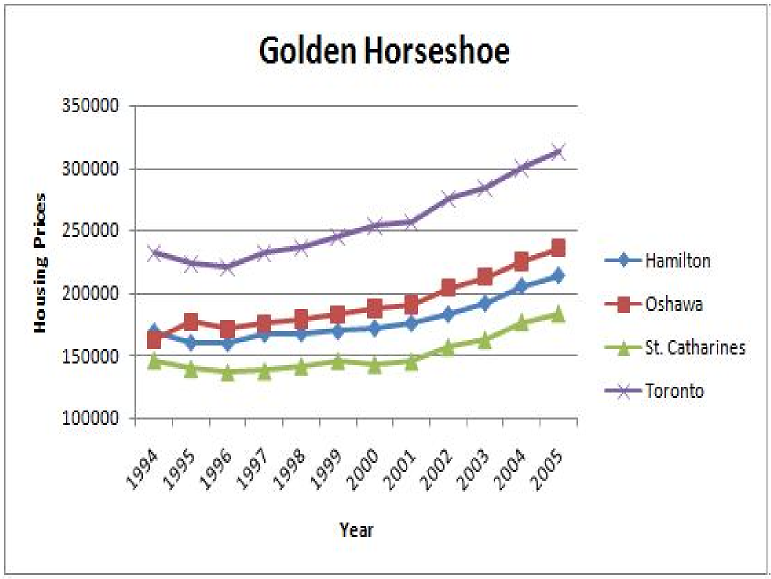

Figure 14: Relationship between Housing Prices and Time in the Golden Horseshoe

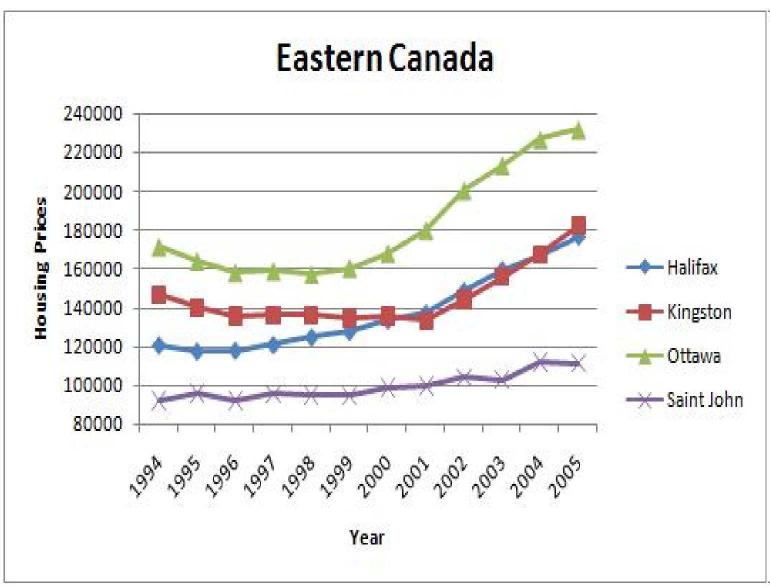

Figure 15: Relationship between Housing Prices and Time in Eastern Canada

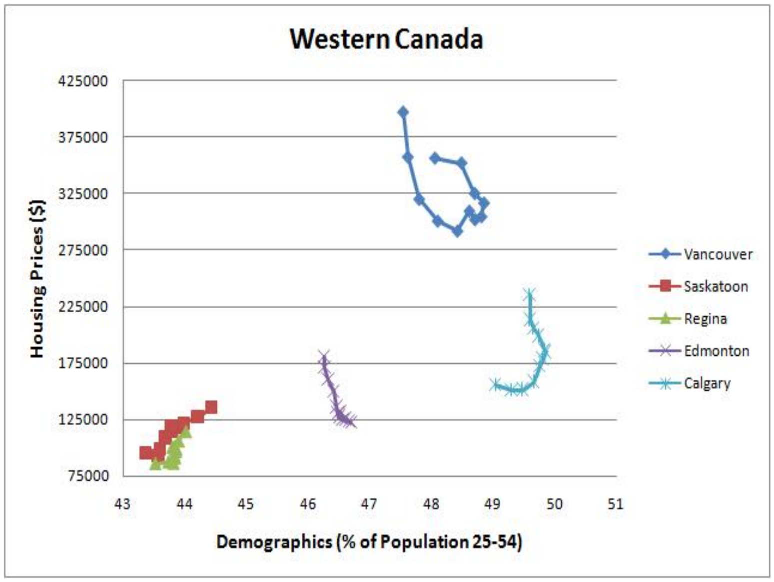

Figure 16: Relationship between Housing Prices and Demographics in Western Canada

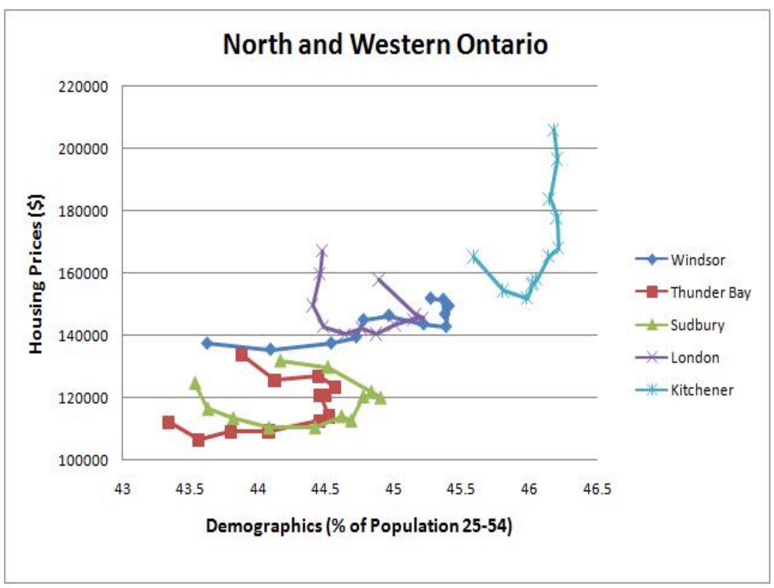

Figure 17: Relationship between Housing Prices and Demographics in North and Western Ontario

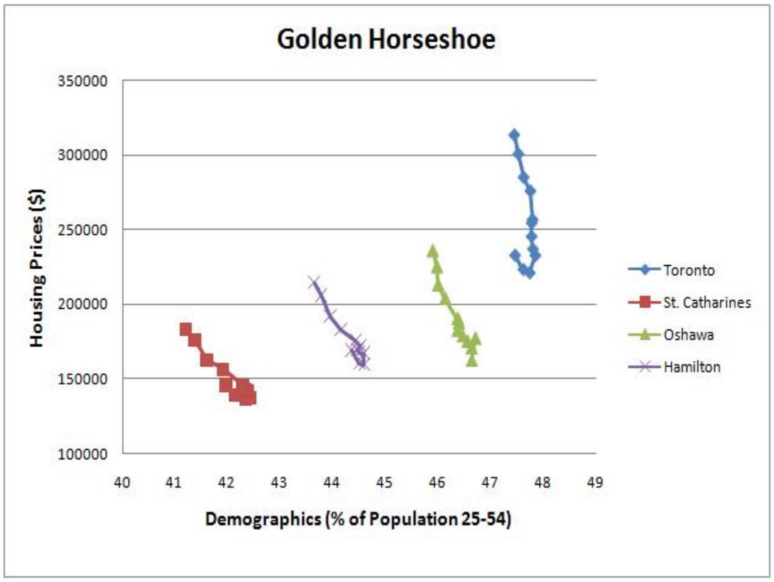

Figure 18: Relationship between Housing Prices and Demographics in the Golden Horseshoe

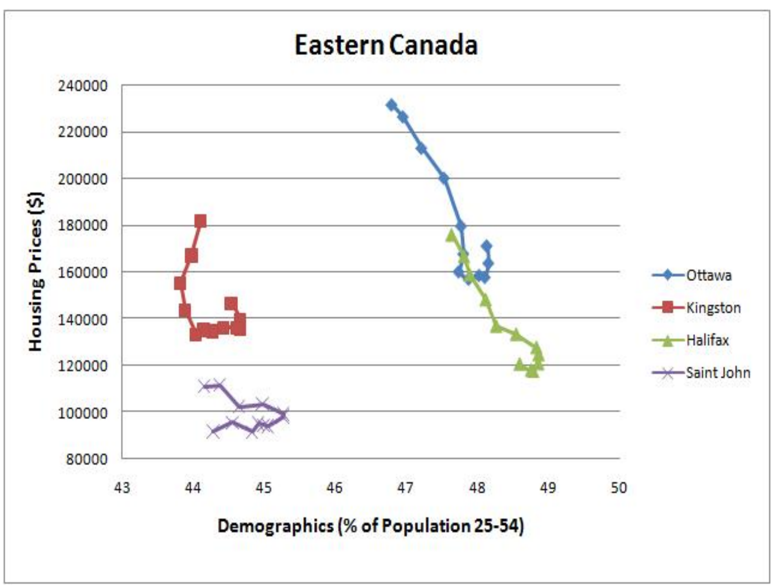

Figure 19: Relationship between Housing Prices and Demographics in Eastern Canada

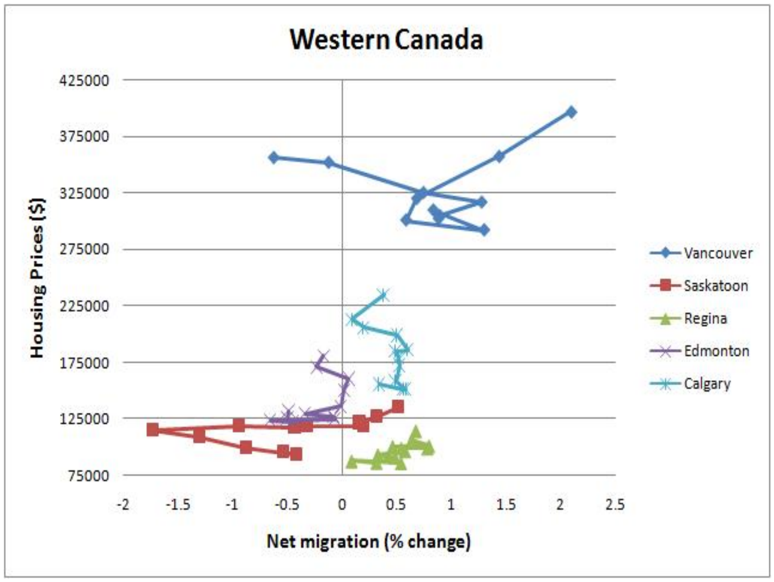

Figure 20: Relationship between Housing Prices and Migration in Western Canada

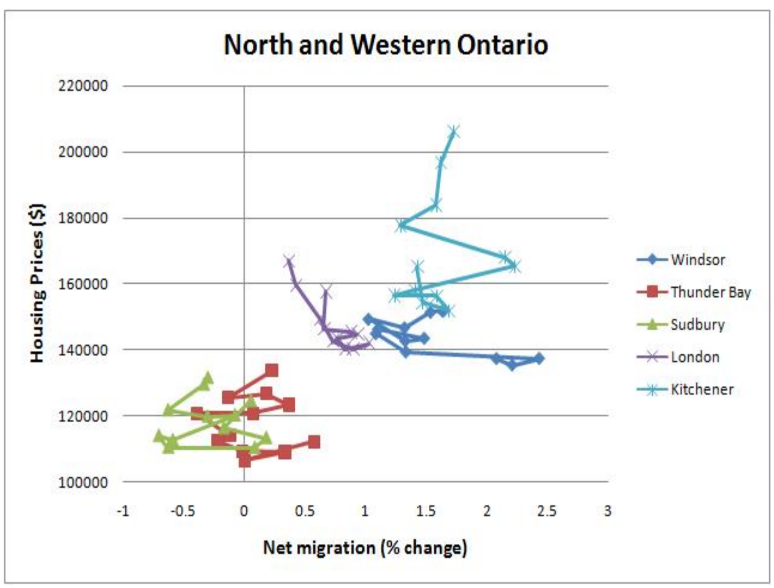

Figure 21: Relationship between Housing Prices and Migration in North and Western Ontario

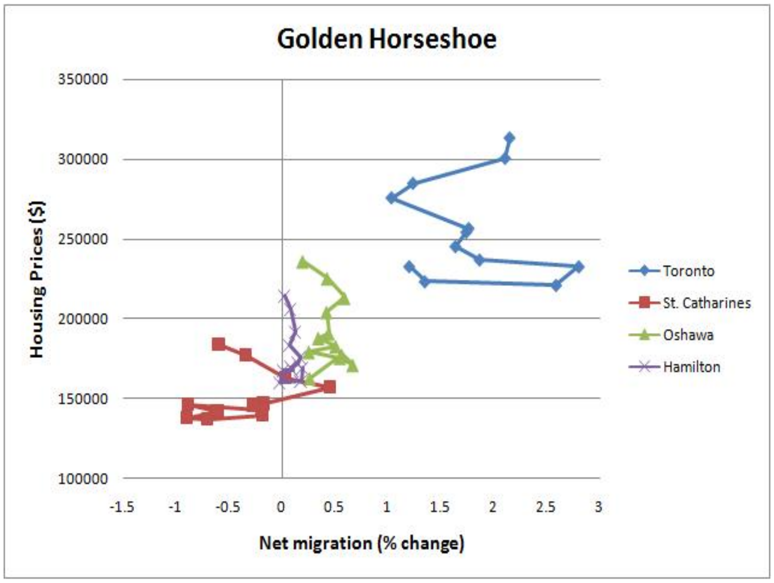

Figure 22: Relationship between Housing Prices and Migration in the Golden Horseshoe

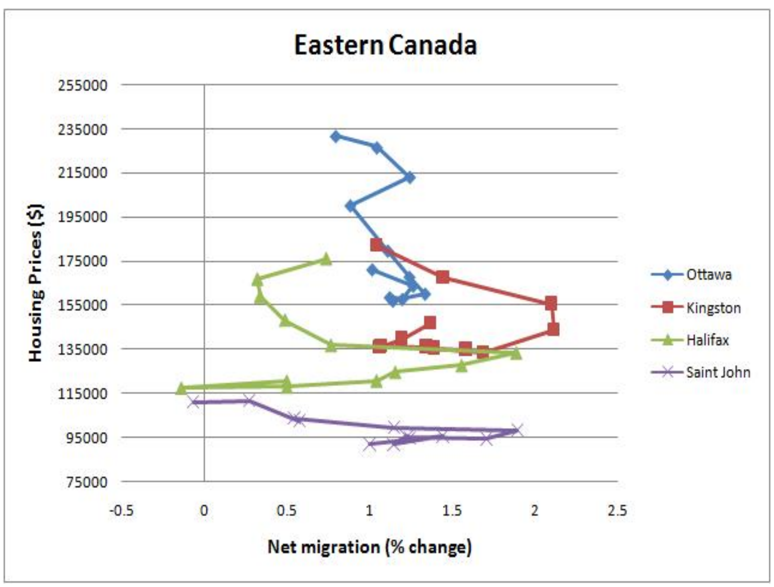

Figure 23: Relationship between Housing Prices and Demographics in Eastern Canada

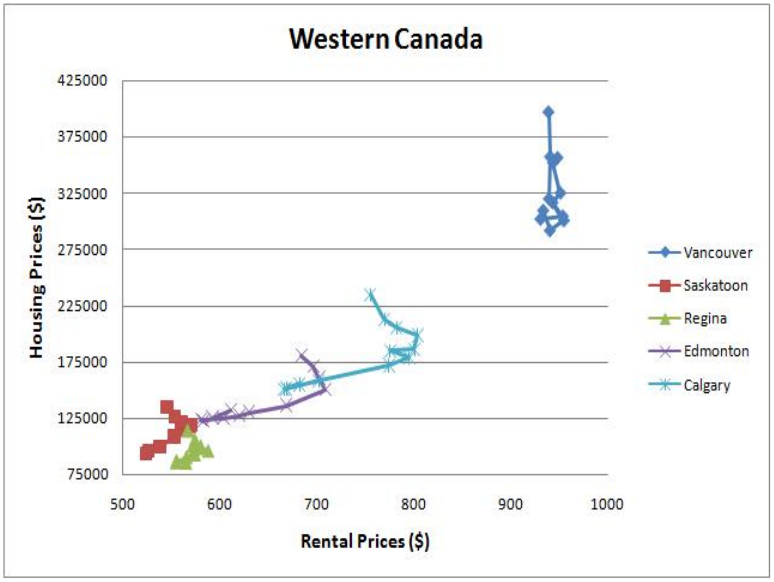

Figure 24: Relationship between Housing Prices and Rent in Western Canada

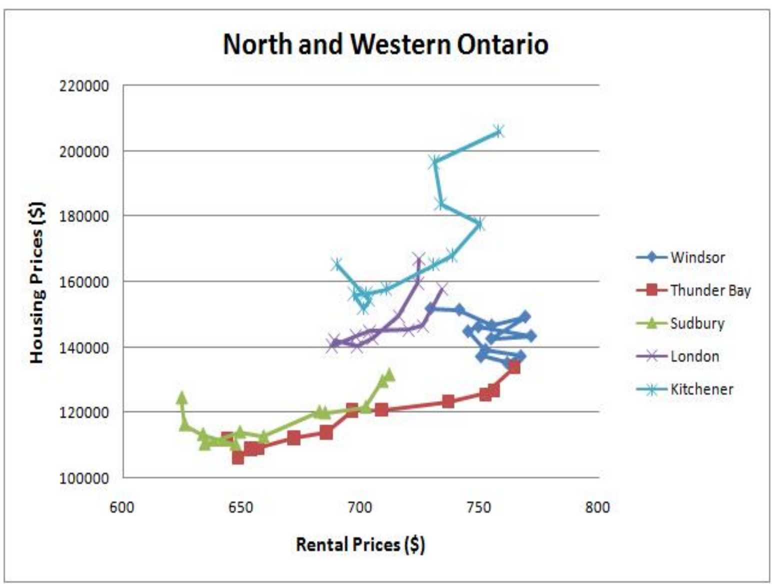

Figure 25: Relationship between Housing Prices and Rent in North and Western Ontario

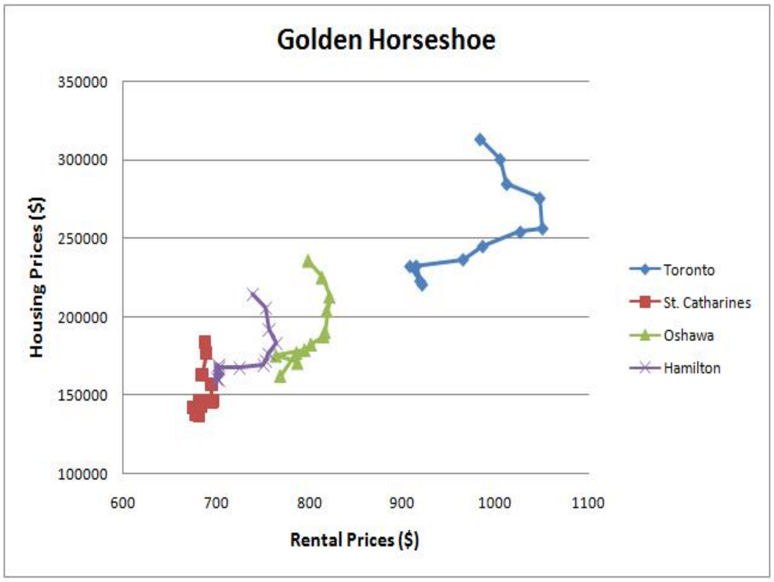

Figure 26: Relationship between Housing Prices and Rent in the Golden Horseshoe

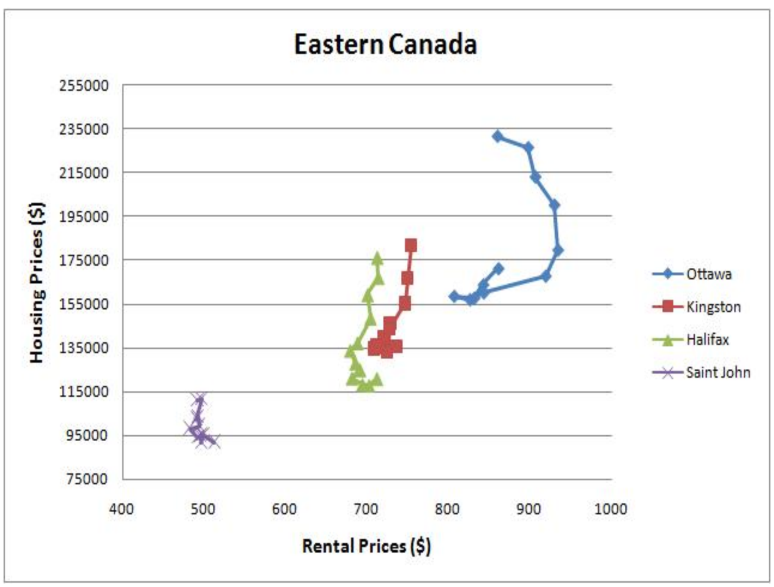

Figure 27: Relationship between Housing Prices and Rent in Eastern Canada

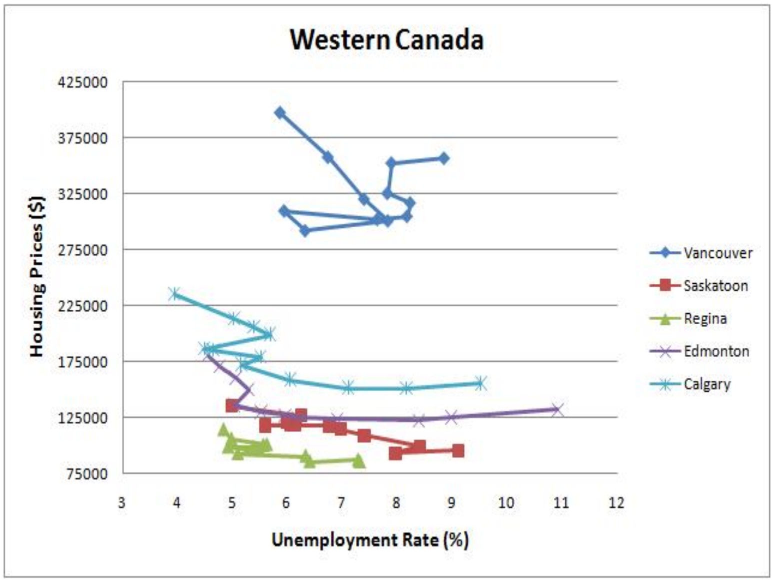

Figure 28: Relationship between Housing Prices and the Unemployment Rate in Western Canada

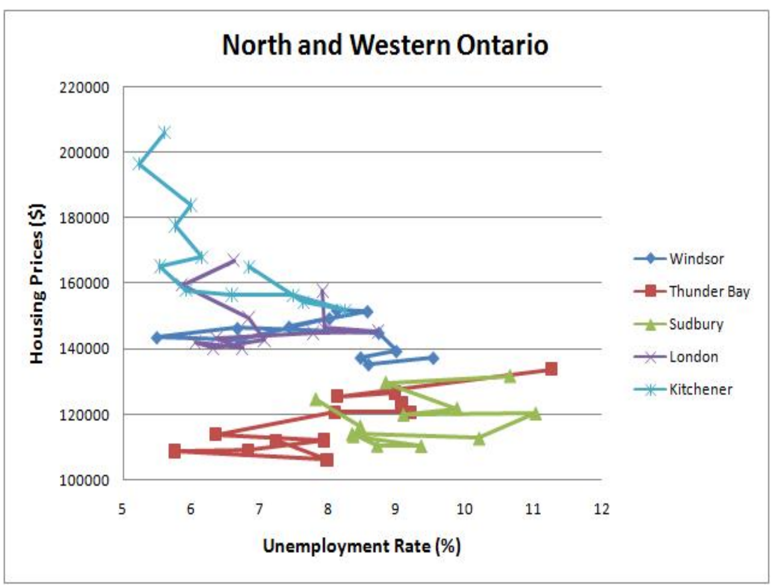

Figure 29: Relationship between Housing Prices and the Unemployment Rate in North and Western Ontario

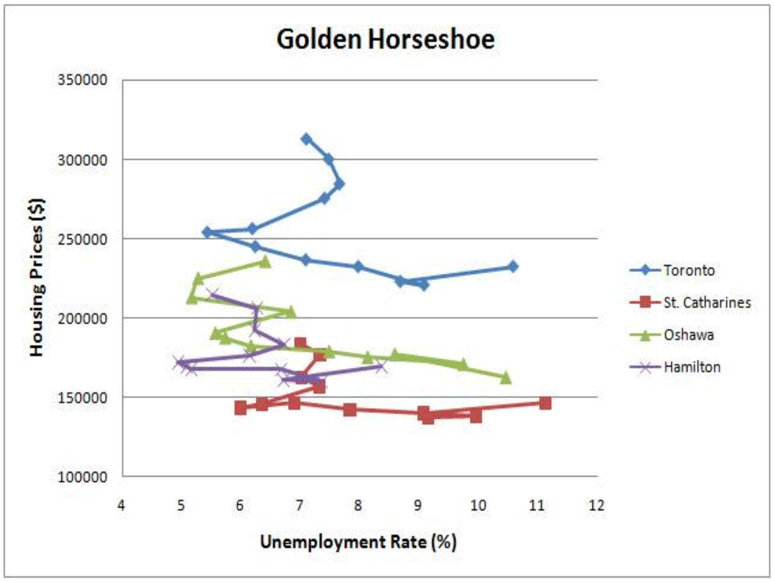

Figure 30: Relationship between Housing Prices and the Unemployment Rate in the Golden Horseshoe

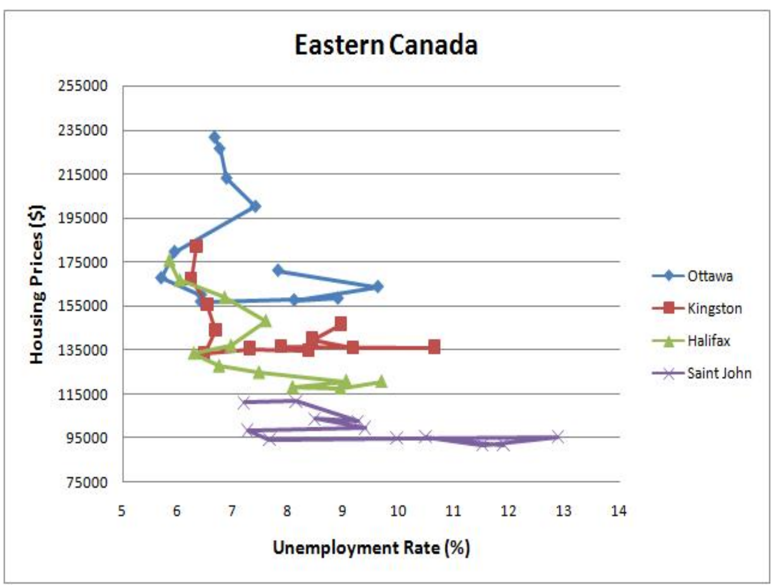

Figure 31: Relationship between Housing Prices and the Unemployment Rate in Eastern Canada

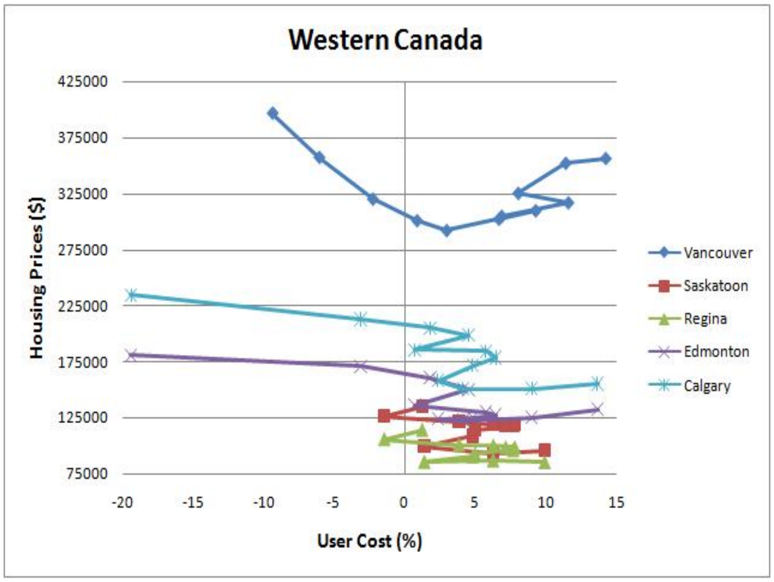

Figure 32: Relationship between Housing Prices and the User Cost of Housing in Western Canada

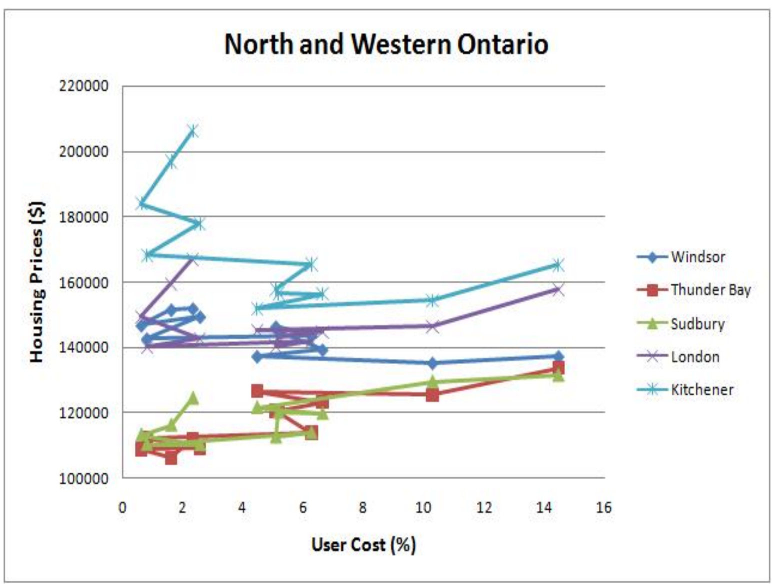

Figure 33: Relationship between Housing Prices and the User Cost of Housing in North and Western Ontario

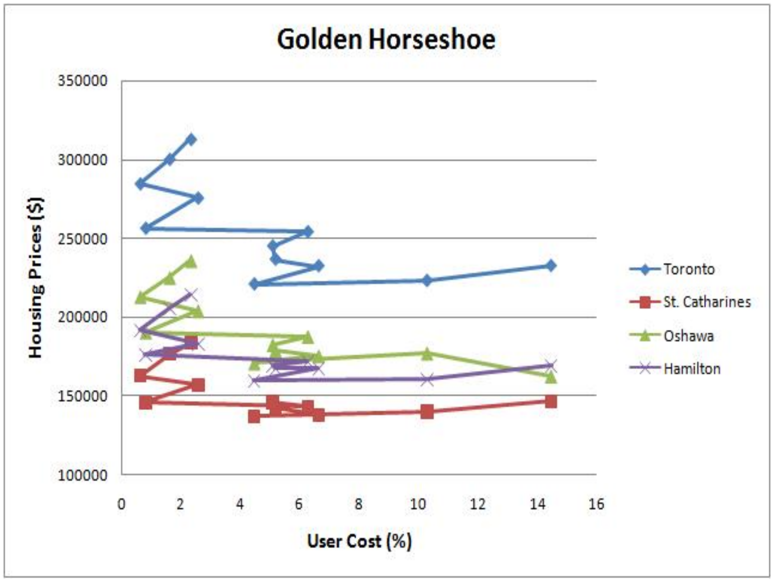

Figure 34: Relationship between Housing Prices and the User Cost of Housing in the Golden Horseshoe

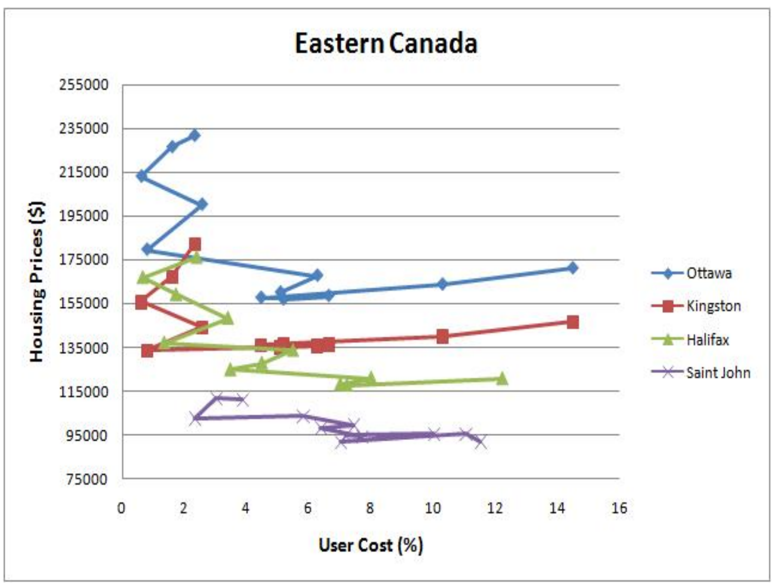

Figure 35: Relationship between Housing Prices and the User Cost of Housing in Eastern Canada

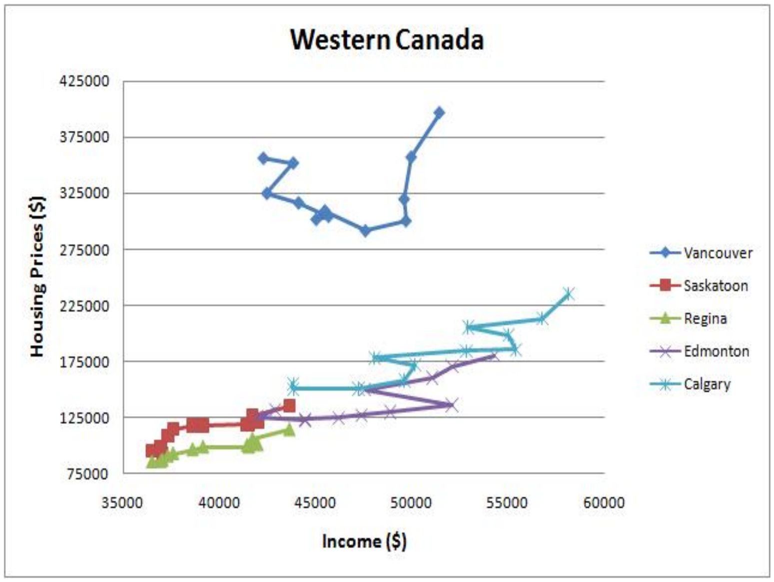

Figure 36: Relationship between Housing Prices and Income in Western Canada

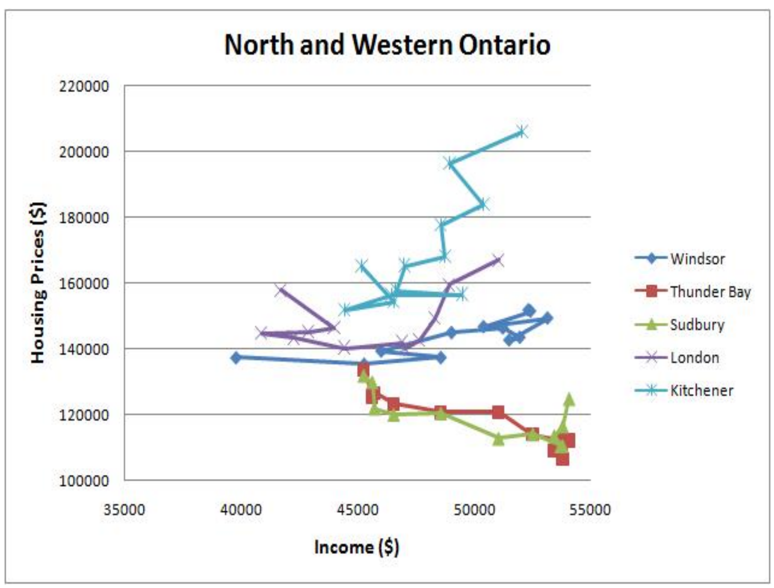

Figure 37: Relationship between Housing Prices and Income in North and Western Ontario

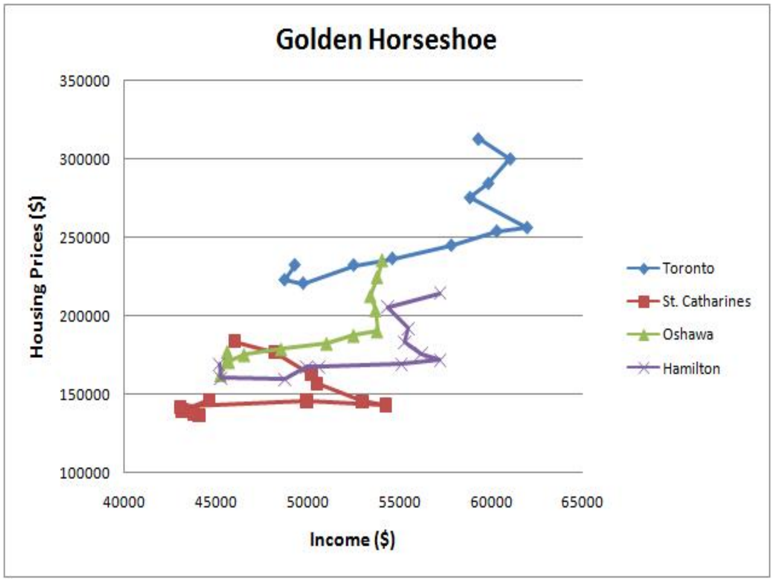

Figure 38: Relationship between Housing Prices and Income in the Golden horseshoe

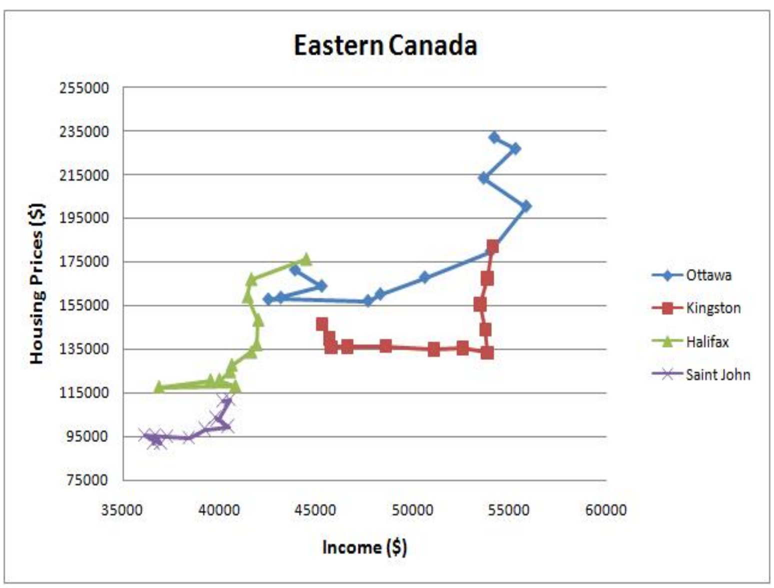

Figure 39: Relationship between Housing Prices and Income in Eastern Canada