Does air pollution feature lower housing price in Canada?

By
Sheng Hong

A Research Project Submitted to
Saint Mary's University, Halifax, Nova Scotia
in Partial Fulfillment of the Requirements for
the Degree of Master of Applied Economics.

August 26, 2015 Halifax, Nova Scotia

Approved:	Dr. Yigit Aydede Supervisor
Approved:	Dr. Ather Akbari Coordinator, MAE Program
Date:	August 26, 2015

Does air pollution feature lower housing price in Canada?

By Sheng Hong

Abstract

The purpose of this study is to analyze the pollution-housing price relationship within a Canadian context, with a special focus on the effects of particular matter 2.5 ($PM_{2.5}$) in terms of air pollution. The objective is important because environmental quality acts as a relatively less tangible characteristic to housing compared with other physical characteristics (such as number of bedrooms), but it has its own implicit price. Once we make explicit the implicit cost of air contaminant, it will guide public policy decisions on the measures that should be taken to reduce air pollution. This study examines the subject between 1997 and 2013 across Canada's benchmark cities. In order to provide an accurate analysis, two types of housing price index data are used: CANSIM New Housing Price Index (NHPI) and Teranet-National Bank Housing Price Index (THPI). $PM_{2.5}$ is employed as the proxy of air pollution and the data are collected from Environment Canada. First difference, lagged values, fixed effect and random effect models are the methods being used to produce an accurate and robust analysis. As a result, as this study improves the specifications with better HPI (which is THPI), the negative association between housing prices and air pollution surfaces. The results from the specifications that applied first-difference, logarithmic function, year dummies and fixed effects or random effects methods, suggest that air pollution has a negative effect on housing prices with a two-year lag.

August 26, 2015

Table of Contents

List of Tables

List of Figures

1. Introduction

In recent years, there has been growing concern about the adverse effects of air pollution, not only among those emerging countries with severe air quality issues such as China and India, but also developed countries such as the United States and European nations. [1] Evidence has been mounting that air pollution can have a detrimental influence upon human health through irritating the eyes, nose and throat if concentrations become elevated (Environment Canada, 2013). Moreover, it affects the appearance of buildings, and in general makes a neighborhood look shabby.[2] Since housing plays a key role in human life and a house acts as a durable good to consumers, homebuyers will consider its characteristics cautiously, including its environmental amenities. It is instructive to ask how consumer well-being is influenced by changes in air quality in terms of housing price. Rosen (1974) was the first to give this association an economic interpretation using a Hedonic model. A Hedonic method is a revealed preference approach which decomposes the total housing expenditure into the values of individual components, each of which has its own implicit price. The hedonic model can be represented as:

(1) Housing Price = f (Physical Characteristics, Other Factors),

where the housing price is a function of its physical factors (such as location, lot size, bedrooms, age) and other determinants. Expenditures on other less tangible characteristics, such as local public services and air quality, also contribute to dwelling price. Rosen also proposed a two-step approach for estimating the marginal willingness-to-pay (MWTP)

---

1. See Ridker and Henning (1967), Zheng, Cao and Kahn (2011) and Raaschou-Nielsen, Andersen and Beelen (2013) for details.
2. See Malpezzi, S. (1996). Housing prices, externalities, and regulation in US metropolitan areas. *Journal of Housing Research*, *7*, 209-242.

function, as well as the supply curve. According to his analysis, homebuyers' MWTP for a house depends on the amenity level, consumer tastes and demand shifters. [3] Based on Rosen's model, Chay and Greenstone (1998) state that "The gradient of the implicit price function with respect to air pollution gives the equilibrium differential that allocates individuals across locations and compensates those who face higher pollution levels" (P.4). Chay and Greenstone also state that locations with worse environmental amenities must have lower property prices in order to attract potential homebuyers. In addition, according to the law of demand, a decline in the demand for a good will result in a lower price for that good. Thus, It seems theoretically reasonable to reveal that consumers are willing to pay less for a house with a relatively worse environmental amenities compared to a house that located in a better environmental amenity. This negative air pollution-housing price relationship is confirmed by many of the existing studies (Ridker and Henning, 1967; Harrison and Rubinfeld, 1978; Malpezzi, 1996; Chay and Greenstone, 1998; Jerrett, Burnett, and Kanaroglou, 2001; Kim, Phipps and Anselin, 2003; Brasington and Hite, 2005; Brasington and Hite, 2005; Bayer, Keohane and Timmins, 2009; Zheng, Cao and Kahn, 2011; Raaschou-Nielsen, Andersen and Beelen, 2013).

While it is reasonable to assume that the detrimental influences of air pollution are reflected in property values, reliable statistical evidence bearing on this hypothesis has been nearly non-existent in Canada. Only Jerrett Burnett, and Kanaroglou, (2001) revealed that dwelling values are significantly and negatively associated with pollution exposure in Hamilton, Canada, a robust result by applying geographical information systems (GIS) analysis and spatial statistical method. Therefore, it is important to investigate the implicit

3. A detailed discussion on Rosen (1974) model will be provided in Section 2.

cost of air pollution in terms of housing price. Once we make explicit the implicit cost of air contaminant, it will guide public policy decisions on the measures that should be taken place to reduce air pollution. It will also be beneficial to narrow the type of air pollution ($PM_{2.5}$) if it has severe social cost and should be given top priority by policy makers. Last but not least, environmental justice has emerged as an important dimension of environmental and public health policy in North America. Canada and the United States have carried out a joint transboundary particulate matter science assessment report in support of the Canada-U.S. Air Quality Agreement. Thus, credibly measuring the economic value of clean air to the housing market is a remarkable topic to policy makers, economists and environmentalists. This is the first study that investigates the air pollution-housing price relationship within a Canadian context.

$PM_{2.5}$ is used as a proxy of air pollution and the data about it is obtained from Environment Canada. Two types of housing price index are employed: CANSIM new housing price index (NHPI) and Teranet-National Bank housing price index (THPI). The full discussion of these indices is provided in Section 3.

The structure of this study is as follows. Section 2 summarizes the literature on the air pollution-housing price relationship. Section 3 presents the data sets employed and the steps that were taken to make them comparable including merging them into a master panel dataset. Section 4 evaluates different model specifications by applying fixed effect and random effect methods, as well as discusses what econometric problems are resolved. Finally, Section 5 discusses the outcome and the policy implications of this paper.

2. Literature Review

Economists have estimated the association between housing prices and air pollution at least since Ridker and Henning (1967). However, Rosen (1974) was the first to give this correlation an economic interpretation. In the Rosen model, a differentiated good can be described by a vector of its characteristics, K = (k_1, k_2,…, k_n). In the case of a house, these characteristics may include structural attributes (e.g., type of heating systems), the provision of neighborhood public services (e.g., distance to work), and local amenities (e.g., air quality). Thus, the price of the ith house can be written as:

(2) p_i = P(k_1, k_2,…, k_n).

The partial derivative of P(·) with respect to the nth characteristic, $\partial p_i/\partial k_n$, is referred to as the marginal implicit price. It is the marginal price of the nth feature implicit in the overall price of the property.

In this case, the welfare effects of non-marginal changes can be calculated. Rosen proposed a two-step approach for estimating the MWTP function, as well as the supply curve. In the first step, equation (2) is estimated and employed to predict the household-specific marginal implicit price, $\partial p_i/\partial k_n$. In the second step, the function of demand and supply are represented as:

(3) $k_{ni}^d = \partial p_i/\partial k_n$ = f (k_n, μ), and

(4) $k_{ni}^s = \partial p_i/\partial k_n$ = g (k_n,η),

where the estimated implicit prices from equation (2) are used as observations on actual prices, and k_{ni}^d and k_{ni}^s are the demand and supply marginal prices of characteristic k_n.

Equation (3) is the MWTP function, which depends on the amenity level, k_n, and on consumer tastes and demand shifters, μ. Equation (4) is the inverse supply curve, which is a function of k_n and production technologies/cost shifters, η. A credible estimation of this system has tremendous practical importance. For example, one could estimate individuals' WTP for the large improvements in air quality induced by the Clean Air Act Amendments of the 1970s.

However, based on the discussions of misspecification of the hedonic pricing model from Halvorsen and Pollakowski (1981) and Cropper et al. (1988), Chay and Greenstone (1998) found that there are two econometrics problems that plague the hedonic method to estimate the WTP for clean air. First, the predicted changes of air pollution-housing price relationship may be biased if we neglect some variables that should be included in the function. In cross-sectional studies, there may be unobserved factors that covary with both air pollution and housing values. For instance, areas with higher level of pollution tend to be more urbanized and have higher population density and higher total income. Second, if the homebuyers' preferences in terms of clean air is heterogeneous, they will choose the houses based on their criteria. Therefore, credible estimation need to be applied.

The Chay and Greenstone (1998) study provides important groundwork. They use the declines in air pollution induced by the 1970 and 1977 Clean Air Act Amendments in United States to reveal new evidence on the capitalization of air quality into property values. They estimate the hedonic price schedule (HPS) in first-differences and apply the county-level regulations as instrumental variables for changes in total suspended particulates (TSPs) pollution. Their findings show that TSPs dropped substantially more in regulated than in unregulated counties during the 1970s. Meanwhile, housing prices rose more in regulated regions. They employed fixed effect method to investigate whether housing prices fall with air pollution, and they applied random effect model to estimate the average MWTP across individuals while accounting for self-selection bias arising from negative assortive matching. They estimate that a 1-mg/m3 decline in particulate levels results in 0.4 to 0.5 increase in housing price, which is an elasticity between 0.3 and 0.4. This appears to be a robust estimate of the average MWTP for clean air across individuals. For example, the estimates from this design are remarkably stable across specifications, while the estimates based on conventional HPS designs are 6-7 times smaller and very sensitive to model specification.

Zheng, Cao and Kahn (2011) provide new hedonic estimates of the implicit price of air pollution in 2006 to 2008 across 85 major Chinese cities. The core question of this paper is: Does air pollution affect housing price in China? According to the core question, the authors use a cross-city hedonic pricing equation and an air pollution production function. The Air Pollution Production Function is expressed as:

(5) $\ln(PM10_{it}) = \alpha_0 + \alpha_1 X_{it} + \alpha_2 \cdot \ln(NEIGHBOR_{it}) + \alpha_3 \cdot \ln(SANDSTORM_i) + \alpha_4 \cdot NORTH_i + \alpha_5 \cdot NORTH\_BORDER_t + \varepsilon_{it}$

Where $PM10_{it}$ is the Particulate Matter 10 (PM_{10}) concentration in city i in year t, X_{it} is a vector of city attributes that affect the city's PM_{10} concentration, such as city population (POP). The last four explanatory variables (NEIGHBOR, SANDSTORM, NORTH and NORTH_BORDER) are the instrumental variables in the hedonic pricing equation that are reported below (Equation (6)). These instrumental variables indicate "imports" of pollution from nearby sources. Moreover, these instrumental variables determine a city's PM_{10} level but are uncorrelated with the error term (ε_{it}) and address endogeneity concerns such as cities population density. The Hedonic Home Price Equation is estimated as:

(6) $\ln(HP_{it}) = \beta_0 + \beta_1 \cdot \ln(POP_{it}) + \beta_2 \cdot A_{it} + \beta_3 \cdot \log(PM10_{it}) + \mu_{it}$

Where HP_{it} is home price in city i in year t, POP_{it} illustrates population density in city i in year t, A_{it} are a vector of amenities in city i in year t, which may include number of hospital beds per capita, number of school teachers per pupil, PM_{10} and the temperature discomfort index,[4] and μ_{it} represents the error term. This regression allows authors to test for the size and statistical significance of amenity effect and the city's population scale effect. The result shows that on average, a 10 per cent decline of the imported pollution from neighbors will cause a 1.8 per cent increase in home values. This paper's main empirical contributions is that it provides new hedonic estimates of the implicit price of air pollution from 2006 to 2008 across 85 major Chinese cities. Furthermore, it is the first study to use PM_{10} as a proxy of air pollution and examine its effect on housing market within a Chinese Context. It is very useful as a reference to other related studies. However, the weaknesses are that they only use the data from 2006 to 2008, which is a short period for an accurate analysis; in addition, the $PM_{2.5}$ may be a better application than PM_{10} based on its criteria and effect of environmental and health damage. PM_{10} are the particles that are between 2.5 and 10 micrometers. On the contrary, the particles that are smaller than 2.5 micrometers are called $PM_{2.5}$, and have more directly adverse effects on human health because it can penetrate deeper into the gas exchange regions of the lung (alveolus), and it may also may pass through the lungs to influence other organs (Environment Canada, 2013). Raaschou-Nielsen (2013) revealed that there was no safe level of particulates in nine European

4. See Zheng, et. al. (2010) for definition of temperature discomfort index.

countries and that for every increase of 10 μg/m^3 in PM_{10}, the lung cancer rate increased by 22 percent. $PM_{2.5}$ were even more deadly, every increase of 10 μg/m^3 in $PM_{2.5}$ resulted in a 36 percent increase in lung cancer. However, the data limitations and time constraints may make it unfeasible to attempt to apply $PM_{2.5}$ for China between 2006 and 2008.

Another study done by Smith and Huang (1995) estimated the association between property values and air pollution, using total suspended particulates (TSPs) as a proxy for the air quality. However, their meta-study of 37 cross-sectional studies indicates that the cross-sectional correlation between housing prices and particulates pollution appears weak: a 1-mg/m^3 decline in TSPs results in 0.05 to 0.10 per cent increase in housing price, which shows only an elasticity between 0.04 and 0.07. Furthermore, there are other studies that support fundamental evidence for our study. For example, Malpezzi (1996) showed the relationship of housing market and environmental costs. Building additional house units may reduce the local supply of greenspace, reduce air quality and increase pressure on local water, sanitation, and solid waste collection systems.

3. Data Set Description

3.1 Air Pollution Data

The air contaminants that are being used as a proxy for air quality is $PM_{2.5}$, due to the following reasons. First, PM are the deadliest form of air pollution because their ability to penetrate deep into lungs and blood streams unfiltered, causing permanent DNA mutations, heart attacks, and premature death (Environment Canada, 2013). Studies have linked particles to aggravate respiratory and cardiac diseases such as bronchitis, asthma and emphysema, as well as various forms of heart disease. [5] PM can also have adverse effects

5. See Harrison and Yin (2000), Raaschou-Nielsen. et al (2013), Katsouyanni, Touloumi and Spix (1997).

on vegetation and structures, and leads to visibility deterioration and regional haze. Second, the size of PM particles mainly identify the extent of environmental and health damage caused. Environment Canada (2013) identifies three types of particles: Total Particulate Matter (TPM), PM_{10} and $PM_{2.5}$. TPM states "airborne particulate matter with an upper size limit of about 100 microns (μm) in aerodynamic equivalent diameter" (Environment Canada, 2013). Particles' mass median diameter smaller than about 10 microns, referred to as PM_{10} and $PM_{2.5}$ are the particles with a mass median diameter less than 2.5 micrometres (μm). Both PM_{10} and $PM_{2.5}$ are associated with health effects. However, $PM_{2.5}$ is more deadly than PM_{10} because it can penetrate deeper into the lungs. Although there are more data available for PM_{10} in the 1990s, $PM_{2.5}$ has had a wider distribution in the late 1990s to today.

$PM_{2.5}$ data from 1995 to 2013 is obtained from Environment Canada. I use the annual average of $PM_{2.5}$ that is based on the annual average concentrations recorded at 64 monitoring stations across Canada. Six types of annual $PM_{2.5}$ datasets from 1995-2013 are provided: $PM_{2.5}$ BAM, $PM_{2.5}$ BAM35, $PM_{2.5}$ SHARP5030, $PM_{2.5}$ TEOM, $PM_{2.5}$ TEOM-FDMS and $PM_{2.5}$ TEOM-SES. By definition, BAM and BAM 35 stand for Beta Attenuation Monitoring, which is a widely used air monitoring technique employing the absorption of beta radiation by solid particles extracted from air flow; SHARP 5030 incorporates two different measuring techniques, of both nephelometry and beta attenuation, to obtain highly accurate particulate monitoring; TEOM, TEOM_FDMS and TEOM-SES are Tapered Element Oscillating Microbalance samplers that operate by drawing air through a filter attached at the tip of a glass tube, while TEOM_FDMS stand for Tapered Element Oscillating Microbalance with Filter Dynamics Measurement System and TEOM-SES represents Tapered Element Oscillating Microbalance fitted with a sample equilibration system.[6] These six different codes are based on technology used to detect $PM_{2.5}$, but they are all considered valid $PM_{2.5}$ methods of measurement. Since technologies have been improved over time, newer technologies confirm the result of old measuring method. Moreover, more measurements produce more accurate $PM_{2.5}$ results.

These datasets are manipulated in three ways in order to get our desired $PM_{2.5}$ dataset. First, each dataset for different years contains 6 different types of land: residential, commercial, industrial, undeveloped rural, forest and agricultural. Since this paper examines the air-pollution-housing price relationship, we only use "residential" type of land. This leaves us with 136 cities across Canada but the earliest year changed to 1996 instead of 1995 due to data cleaning. Second, since each dataset is classified based on different year and different technology use, I merged all 62 dataset files into one master $PM_{2.5}$ dataset that includes all the years and all technology types. Third, we removed all the irrelevant information that we do not need, such as latitude, longitude and elevation. Thus, this leaves the master dataset with only Year, City and annual average $PM_{2.5}$. However, this is an unbalanced dataset due to not all the cities having data available starting from the same year; they have different beginning time. For instance, Saint John has the data available since 1996 while Calgary has it available only from 2002. [7]

6. For a full description, see Taylor, E., & McMillan, A. (2013). *Air Quality Management: Canadian Perspectives on a Global Issue*. Springer Science & Business Media.
7. See Appendix for a full list of the $PM_{2.5}$ dataset.

3.2 Housing Price Data

There are only two sources that can be sued to compile housing price indices. In order to provide an accurate analysis, both types are used in this study: CANSIM New Housing Price Index (NHPI) and Teranet-National Bank Housing Price Index (THPI).

3.2.1 CANSIM Housing Data

The New Housing Price Index (NHPI) is obtained from CANSIM[8]. The NHPI measures the selling price of newly constructed houses with a focus on the changes between measurement periods (monthly). NHPI data are obtained originally from survey respondents and derived from other Statistics Canada surveys. The following housing types are covered by the survey: detached house, semi-detached and row dwellings (town house and garden home). Contractors' estimates of the current value (evaluated at market price) of the land are also collected by the survey. These estimates are independently indexed to provide the published series for land. Moreover, a residual value (total selling value less land value), which largely associates to the current cost of the structure (house) is independently indexed and is presented as the estimated dwelling series (Statistics Canada, 2015). [9]

Among the cities with NHPI data available, only 11 also have $PM_{2.5}$ data for the period of 1996 to 2013: Calgary, Edmonton, Ottawa-Gatineau, Hamilton, Montreal, Toronto-Oshawa, Quebec, Vancouver, Victoria, Windsor and Winnipeg.

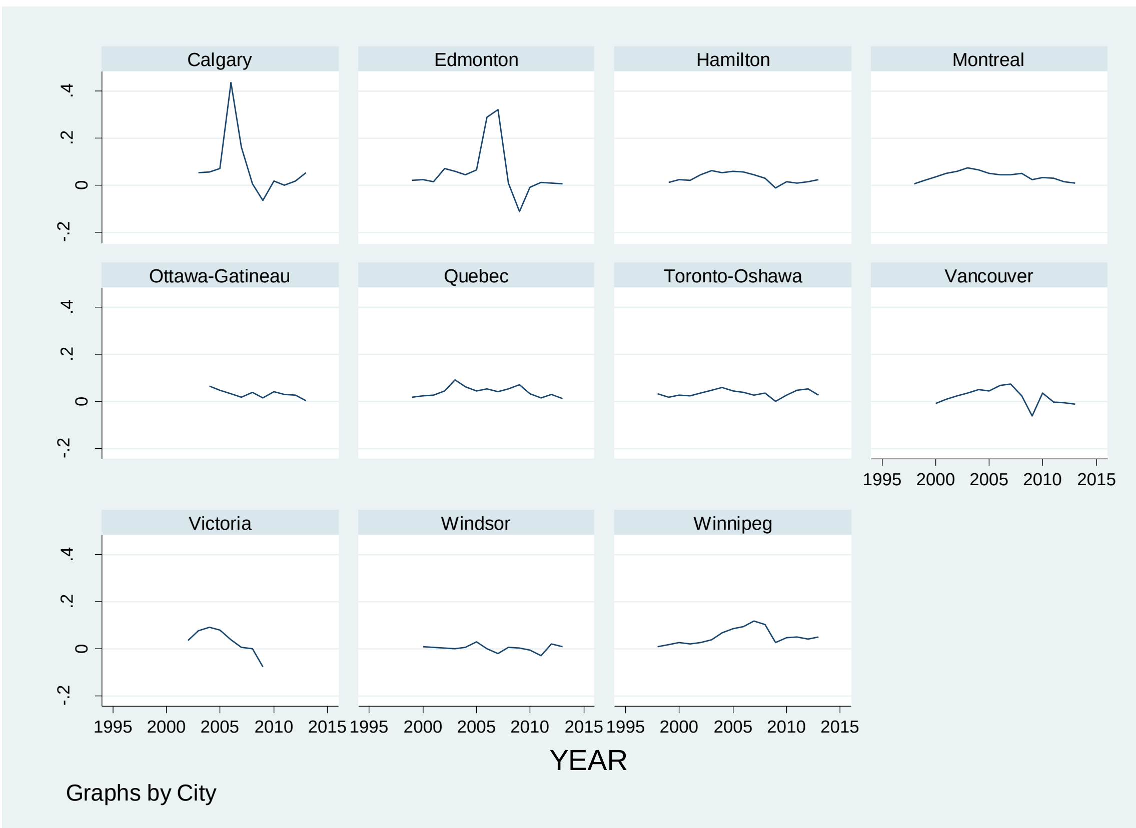
---

8. CANSIM is Statistics Canada's main socioeconomic time series database. It contains most of the aggregate data collected by Statistics Canada on a regular basis such as data from the Consumer Price Index Survey, the Labour Force Survey, or the National Income and Expenditure accounts (University of Toronto, 2013).
9. For a detailed description of NHPI, see Statistics Canada. (2015). *New Housing Price Index (NHPI)*.

Since NHPI only tracks the changes over time in the contractors' selling prices of new residential dwellings, it does not capture the resale prices over time, which may result in an inaccurate analysis. Therefore, Teranet – National Bank House Price Index is also used.

Figure 1 shows the growth rate of NHPI in terms of 11 cities from 1997 to 2013. The NHPI spike in Calgary and Edmonton from 2005 to 2007 may be due to a boom in the oil industry.

Figure 1 Growth Rates of NHPI by Major Canadian Cities, 1997-2013



Source: Statistic Canada (2014)

3.2.2 Teranet-National Bank Housing Data

The THPI is obtained from Teranet – National Bank House Price Index. The THPI is measured by tracking the registered property sale prices over time. Repeat sales methodology were used so at least two sales of the same property are considered in the calculation of the index. Such a "sales pair" evaluates the increase or decrease of the house price in the period between the sales in a linear regression algorithm. Properties that are affected by endogenous factors are not considered in the estimation. These factors include: non-arms-length sale, change of type of property (for instance after renovations), high turnover frequency and data error. Once the unqualified sales pairs have been minimized or eliminated, the estimation of the index in a certain jurisdiction can be started by compiling all qualified sales pairs in a linear regression estimator.[10]

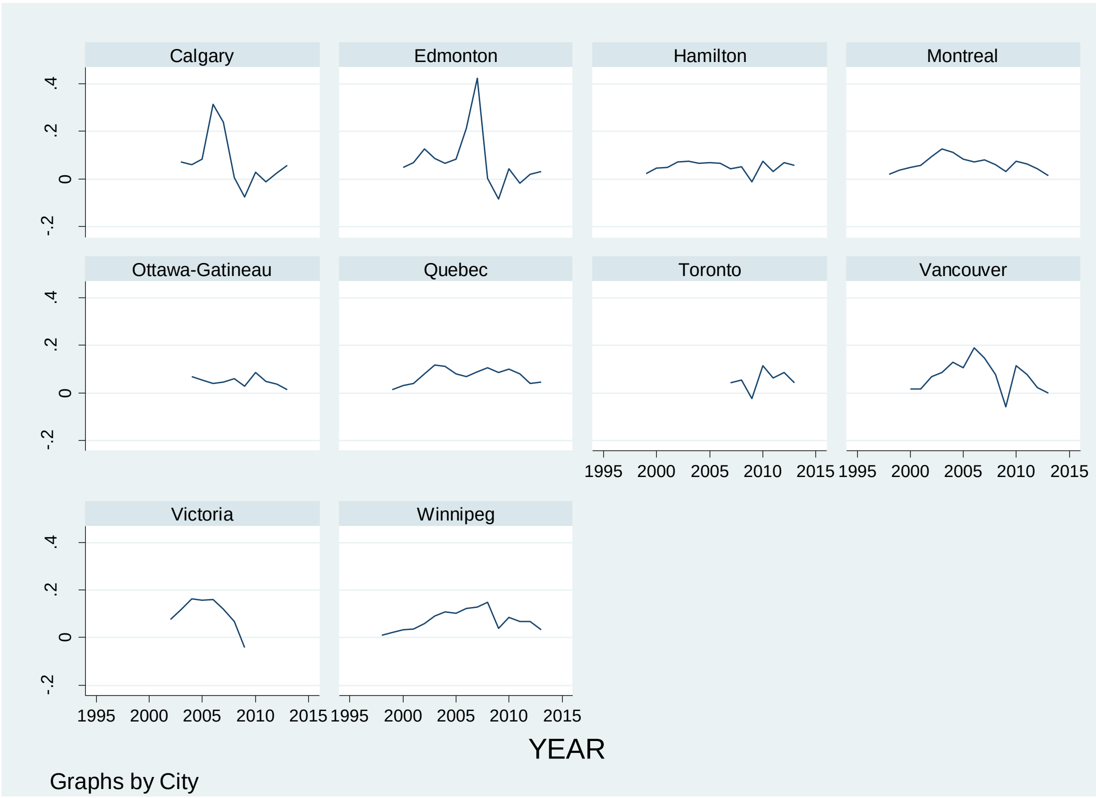
Due to the limited of cities in THPI, in order to match the $PM_{2.5}$ data, 10 major Canadian cities from 1996 to 2013 were collected: Calgary, Edmonton, Ottawa-Gatineau, Hamilton, Montreal, Toronto, Quebec, Vancouver, Victoria and Winnipeg.

Overall, the $PM_{2.5}$ dataset covers 136 Canadian cities between 1996 and 2013, the CANSIM NHPI covers only 11 cities out of 136 and THPI contains only 10 cities out of 136 cities. In addition, after combining $PM_{2.5}$ data and housing price index data into one massive dataset, our analysis is restricted to the period between 1997 and 2013. Therefore, this study uses an unbalanced panel data.

10. For a full description of the methodology used to calculate the Teranet – National Bank House Price Index™, see Teranet – National Bank House Price Index™, Methodology.

Figure 2 demonstrates the growth rate of THPI in terms of 10 cities from 1997 to 2013. As before, we can see the spike in Calgary and Edmonton from 2005 to 2007, due to a boom in the oil industry.

Figure 2 Growth Rates of THPI by Major Canadian Cities, 1997-2013



Source: Teranet- National Bank Housie Price Index (n.d.)

3.3 Other Control Variables

Population density, unemployment rate and family average total income are the control variables for our estimation. They are collected from CANSIM, matching the period from 1997 to 2013. However, there is a limitation in regard of the family average total income data. It only includes the period from 1997 to 2011 instead of from 1997 to 2013.

Population density is selected as a control variable because it captures major differences in environments and amenities. Burda (2014) illustrates that population growth impacts supply and demand — and subsequently affects home prices. As population boosts, demand and competition for desirable housing increases, which can lead to rising prices for certain property types. In addition, as shown by Katz and Rosen (1987), population of a metropolitan area may be perfectly mobile across municipalities. This implies the population is likely to migrate to those cities that are highly productive and that have high amenities; as a result, such a migration may drive up the housing price.

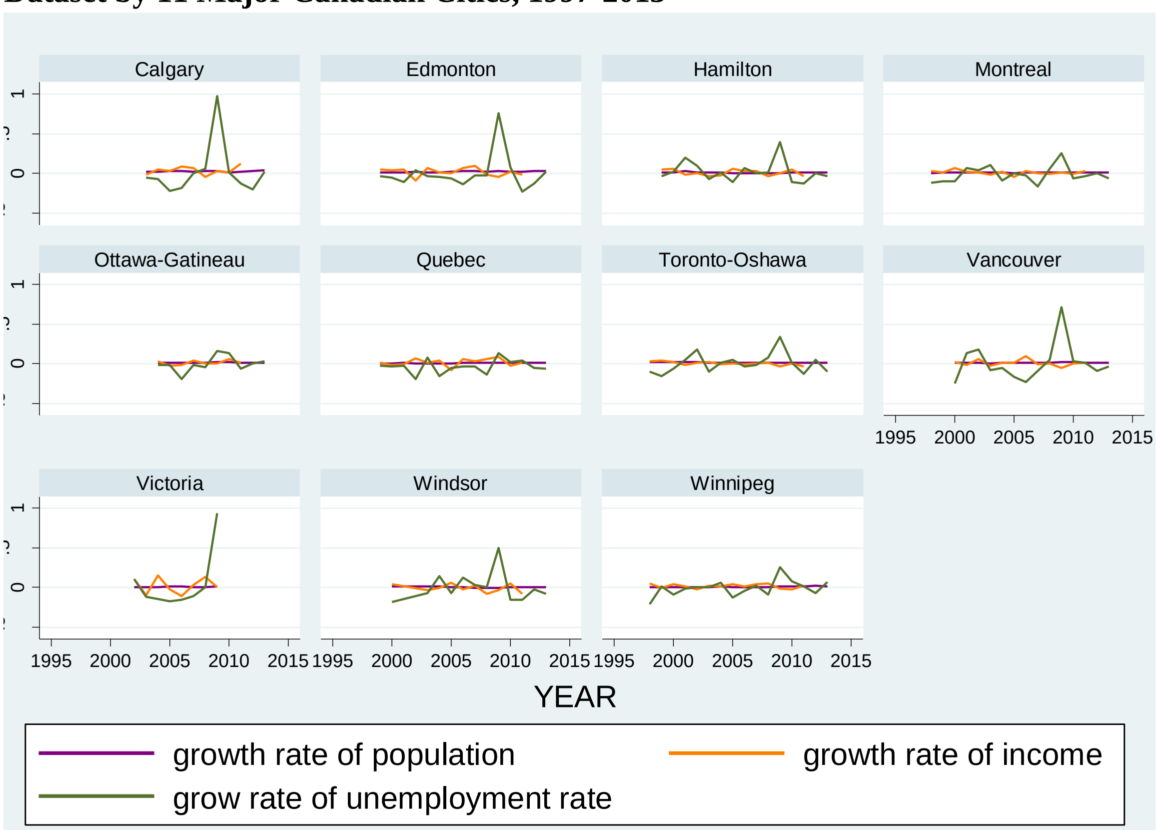
Income are an important control variable, because that increases in income are well-recognized demand-side factors can influence long-term price trends in Canada. Average real income in the Greater Toronto Area increased by 18 per cent over the past 30 years (in 2012 dollars), which tends to increase demand, but real home prices increased by 80 percent during the same time period. However, the upward pressure on housing prices is not caused by income alone; other medium- and short-term factors such as population growth and mortgage financing impact housing price too. (Burda, 2014). Moreover, since urban population is enjoying increased income and the average urbanite is increasingly well-educated, such households will be increasingly willing to pay more for a house with good environmental amenities, thus, willing to pay more to avoid urban air pollution (Zheng, Cao and Kahn, 2011).

Likewise, unemployment affects the housing market too. As shown by Oswald (1999), there was a strong increase in house prices between 1993 and 1999 in the United Kingdom. This was due to a combination of low unemployment rate, high population growth and low interest rates. On the contrary, when unemployment is increasing, fewer people will be able to afford a house. However, even the fear of unemployment may discourage people from entering the property market.

Overall, after redefining the dataset, we are left with a 16-year (1997-2013) comprehensive panel with 11 cities in CANSIM dataset and 10 cities in Teranet dataset, for a maximum dimension of 176 observations and 160 observations, respectively.

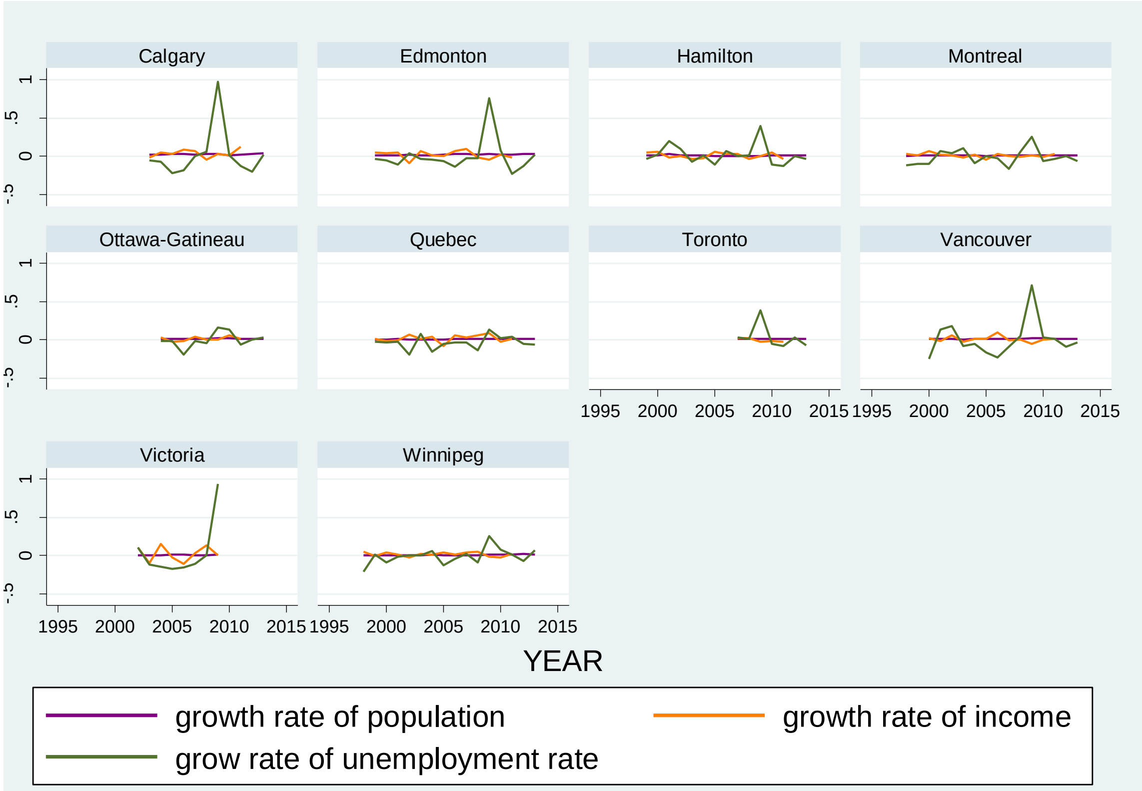
Figure 3 and Figure 4 show the growth rates of population, income and unemployment from 1997 to 2013, for the 11 cities that are used in CANSIM dataset and 10 cities that are selected in Teranet dataset, respectively. The growth rate of population remained stable between 1997 and 2013, and the growth rate of income fluctuated slightly. However, the growth rate of unemployment fluctuated dramatically over time.

Figure 3 Growth Rates of Population, Income and Unemployment for CANSIM Dataset by 11 Major Canadian Cities, 1997-2013



Source: Statistic Canada (2011, 2014, 2015a, 2015b)

Figure 4 Growth Rate of Population, Income and Unemployment for the 10 Major Canadian Cities (that are used) in Teranet Dataset, 1997-2013



Source: Statistic Canada (2011, 2014, 2015a, 2015b)

4. Model Results and Discussion

This section explicates different methodologies that are used for both CANSIM and Teranet datasets. The fixed effect approach is used on each specification in both datasets, for the purpose of removing the unobserved differences between cities and between years. A fixed effects (FE) model that represents the observed quantities in terms of independent variables that are treated as they were non-random. FE controls the unobserved heterogeneity when this heterogeneity is constant over time and correlated with explanatory variables. This constant can be removed through differencing, for instance by taking a first difference (Gujarati, 2008, p.596).

4.1 Table 1 (CANSIM) vs Table 2 (Teranet)

Table 1 and Table 2 illustrate the regression results by using the FE model for both CANSIM and Teranet datasets. In terms of both datasets, specification (1) is the a FE model in the present time (t) with year dummies, while specifications (2) and (3) add lagged values on top of that. The reason to include the lagged values is due to the fact that the change in air pollution on housing prices may not occur immediately. Therefore, the specification (2) and (3) introduce 1 lag and 2 lags respectively, to capture the full impact of air pollution on housing prices. In addition, year dummies are included for these three specifications in order to remove unobserved differences between the years, such as policy changes in different years. Furthermore, robust regression methods are used to achieve almost the efficiency of ordinary least squares (OLS) with ideal data and substantially better-than-OLS efficiency in the face of non-normal situations (Hamilton, 2008, p.239).

We apply the following model for CANSIM and Teranet datasets in this comparison

(Table 1 vs. Table 2):

(7)

$$HPI_{it} = \alpha_i + \gamma_t + \sum_{j=0}^{2} \alpha_j\, PM2.5_{i,t-j} + \sum_{j=0}^{2} \alpha_j\, POP_{i,t-j} + \sum_{j=0}^{2} \alpha_j\, I_{i,t-j} + \sum_{j=0}^{2} \alpha_j\, UR_{i,t-j} + u_{it}$$

where HPI_{it} represents the NHPI or THPI for CANSIM and Teranet datasets, for the city i in the time period t. The lags are represented by j. This specification includes current change and 2 lags. PM2.5 is $PM_{2.5}$ for the city i in the time period t-j; POP_{it} stands for population for the city i in the same period; I represents average total income; UR is unemployment rate and u_{it} is a random error term. α_i indicates regional fixed effects and γ_t represents year dummies.

The results of both datasets are statistically insignificant, which suggests that the association of air pollution and housing prices is not strong. However, when I improve specifications using lagged values, the negative correlation between air pollution and housing prices surfaces. The specification (3) for both datasets appears to provide a negative association between air pollution and housing prices, especially the Teranet dataset, which shows that a 1 $\mu g/m^3$ increase in $PM_{2.5}$ on t-2 period will decrease the HPI by 1.65 unit point. Moreover, among all the control variables, the results of CANSIM dataset indicate that the change in income in the current time period, the t-1 period and the t-2 period all have a statistically significant, positive impact on housing prices. While in the Teranet dataset, the change in income in the current time period and the change in income the t-1 period has a positive impact on housing prices, both with a statistically significant coefficient. In both datasets, population has a positive effect on housing prices, and unemployment rate has a negative impact on housing prices; however, their coefficients are not statistically significant. Furthermore, the number of observations is not consistent with the maximum dimensions because of the use of unbalanced data, and the number of observations varies over different specifications due to the use of lagged values.

The fact that the findings are not robust may be due to the specification selection. In the next sections, I will apply different specifications in order to see if there is any improvement on the results.

Table 1 Regression Result for *CANSIM* Dataset by Fixed Effect method

	NHPI (t)	Coefficient	Robust Std. Err.	t-statistics	R-squared	# of observations
(1)	PM2.5 (t) POP (t) I (t) UR (t)	0.8619042 0.00000785 0.0010276 -1.513864	.469807 .00000636 .0002257 .89323	1.83 1.23 4.55 -1.69	within = 0.9037 between = 0.0063 overall = 0.3345	140
(2)	PM2.5 (t-1) POP (t-1) I (t-1) UR (t-1)	0.4227247 0.00000522 0.0009312 -01.777938	.6990823 .00000763 .0001918 1.140941	0.60 0.68 4.86 -1.56	within = 0.8903 between = 0.0031 overall = 0.4364	138
(3)	PM2.5 (t-2) POP (t-2) I (t-2) UR (t-2)	-0.1165756 0.00000176 .0007186 -2.027714	.682139 .00000916 .0001971 1.243219	-0.17 0.19 3.65 -1.63	within = 0.8662 between = 0.0010 overall = 0.5111	137

Table 2 Regression Result for *Teranet* Dataset by Fixed Effect method

	THPI (t)	Coefficient	Robust Std. Err.	t-statistics	R-squared	# of observations
(1)	PM2.5 (t) POP (t) I (t) UR (t)	.1540663 .0000281 .0010804 -2.438983	.9216106 .0000392 .0005444 2.032106	0.17 0.72 1.98 -1.20	within = 0.9346 between = 0.1137 overall = 0.3707	118
(2)	PM2.5 (t-1) POP (t-1) I (t-1) UR (t-1)	-1.137517 .0000286 .0009283 -4.285495	1.111998 .0000467 .0004735 2.44473	-1.02 0.61 1.96 -1.75	within = 0.9393 between = 0.1404 overall = 0.4060	118
(3)	PM2.5 (t-2) POP (t-2) I (t-2) UR (t-2)	-1.650057 0.00000705 .0005653 -5.035819	1.182144 .0000491 .0003532 2.645249	-1.40 0.14 1.60 -1.90	within = 0.9374 between = 0.4872 overall = 0.8390	117

4.2 Table 3 (CANSIM) vs Table 4 (Teranet)

A logarithmic function is used in this comparison to examine if the percentage change can provide a better result. The logarithmic function is represented as:

(8)

$$\ln(HPI_{it}) = \alpha_i + \gamma_t + \sum_{j=0}^{2} \alpha_j \ln(PM2.5_{i,t-j}) + \sum_{j=0}^{2} \alpha_j \ln(POP_{i,t-j}) + \sum_{j=0}^{2} \alpha_j \ln(I_{i,t-j}) + \sum_{j=0}^{2} \alpha_j \ln(UR_{i,t-j}) + u_{it}$$

Where $\ln(HPI_{it})$ represents growth rate of HPI, for the city i in period t-j; $\ln(PM2.5_{i,t-j})$ indicates growth rate of $PM_{2.5}$, for the city i in the period t-j; $\ln(POP_{i,t-j})$ is growth rate of population; $\ln(I_{i,t-j})$ is growth rate of income and $\ln(UR_{i,t-j})$ represents growth rate of unemployment in the same time period.

The logarithmic term for both datasets render statistically insignificant coefficients of air pollution; however, when I improve specifications using the lagged values, the t statistics improve and the negative association between air pollution and housing prices is surfaces again. In particular, the percentage change in $PM_{2.5}$ in t-2 period has a negative impact on housing prices. Especially the specification (3) in Teranet dataset, one percentage change in $PM_{2.5}$ will result in 0.079 percent decrease in housing prices. Furthermore, among the control variables, only the percentage change in income of CANSIM dataset has a statistically significant, positive impact on housing prices in time period t, t-1 and t-2.

Table 3 Regression Result for *CANSIM* Dataset with Percentage Changes

	ln NHPI (t)	Coefficient	Robust Std. Err.	t-statistics	R-squared	# of observations
(1)	ln PM2.5 (t) ln POP (t) ln I (t) ln UR (t)	0.0655508 1.259269 0.7601693 -0.127999	.0431251 .6511739 .2622133 .0815711	1.52 1.93 2.90 -1.57	within = 0.9032 between = 0.0202 overall = 0.0065	140
(2)	ln PM2.5 (t-1) ln POP (t-1) ln I (t-1) ln UR (t-1)	0.0068472 0.8778518 0.7801083 -0.1244648	.0612969 .6934799 .2689603 .0988152	0.11 1.27 2.90 -1.26	within = 0.8882 between = 0.0136 overall = 0.0243	138
(3)	ln PM2.5 (t-2) ln POP (t-2) ln I (t-2) ln UR (t-2)	-0.0398454 0.4530492 0.7053681 -0.1146456	.059801 .7265818 .2780504 .0998087	-0.67 0.62 2.54 -1.15	within = 0.8663 between = 0.0070 overall = 0.0934	137

Table 4 Regression Result for *Teranet* Dataset with Percentage Changes

	ln THPI (t)	Coefficient	Robust Std. Err.	t-statistics	R-squared	# of observations
(1)	ln PM2.5 (t) ln POP (t) ln I (t) ln UR (t)	.0075337 .9044599 .3122953 -.2112242	.0469646 .739065 .2593118 .1050111	0.16 1.22 1.20 -2.01	within = 0.9638 between = 0.1191 overall = 0.1881	118
(2)	ln PM2.5 (t-1) ln POP (t-1) ln I (t-1) ln UR (t-1)	-.0508314 .5086704 .3372754 -.262853	.0584964 .7651506 .2542378 .1187194	-0.87 0.66 1.33 -2.21	within = 0.9675 between = 0.2313 overall = 0.4414	118
(3)	ln PM2.5 (t-2) ln POP (t-2) ln I (t-2) ln UR (t-2)	-.0786798 -.1144864 .3314157 -.2426014	.0603188 .7333044 .1966987 .1228514	-1.30 -0.16 1.68 -1.97	within = 0.9646 between = 0.4464 overall = 0.8487	117

4.3 Table 5 (CANSIM) vs Table 6 (Teranet)

In order to remedy the nonstationary[11] issue, the data is first-differenced. In this comparison, FE model and first-difference[12] are employed but without year dummies, because first-difference method already make data stationary, so the year dummies are not necessary to apply. The first difference regression is shown as:

(9)

$$\Delta HPI_{it} = \alpha_i + \sum_{j=0}^{2} \alpha_j \, \Delta PM2.5_{i,t-j} + \sum_{j=0}^{2} \alpha_j \, \Delta POP_{i,t-j} + \sum_{j=0}^{2} \alpha_j \, \Delta I_{i,t-j} + \sum_{j=0}^{2} \alpha_j \, \Delta UR_{i,t-j} + u_{it}$$

Where ΔHPI_{it} represents differences of successive HPI (HPI_{it} – $HPI_{i,t-1}$); $\Delta PM2.5_{i,t-j}$ indicates differences of successive $PM_{2.5}$ ($PM2.5_{it}$ - $PM2.5_{i,t-1}$; or $PM2.5_{it}$ - $PM2.5_{i,t-2}$); $\Delta POP_{i,t-j}$ is differences of successive population; $\Delta I_{i,t-j}$ is differences of successive income and $\Delta UR_{i,t-j}$ represents differences of successive unemployment rate.

The results of CANSIM dataset suggest that $PM_{2.5}$ in all specifications has a negative impact on the housing price, which meets our expectation. This indicates that when I remove non-stationary factors, even CANSIM can have a negative result between air pollution and housing prices. The t-statistics of $PM_{2.5}$ and the robust standard error improves as the specifications improve. However, none of the control variables has a statistically significant coefficient.

In terms of the results in Teranet dataset (Table 6), differences of successive $PM_{2.5}$ in t-1 and t-2 periods produce relatively stronger t statistics, indicating that $PM_{2.5}$ has a negative

11. Non-stationary process has a variable variance and a mean that does not remain near, or returns to a long-run mean over time (Gujarati, 2008, p.741)
12. First difference is an approach by running the regression on the differences of successive values of the variables in order to address the problem of omitted variables with panel data (Gujarati, 2008, p.345).

impact on the housing price. For example, 1 μg/m^3 increase in $PM_{2.5}$ on t-2 period will result in a decrease in the THPI by 0.61 unit points. Even though the coefficients of $PM_{2.5}$ are not statistically significant, the t-statistics of $PM_{2.5}$ and the robust standard error are improve as the specifications improve.

Overall, we can conclude that Teranet is a better dataset since it provides better results from the three comparisons (Table 1 vs. Table 2; Table 3 vs. Table 4; Table 5 vs. Table 6) so far, mainly due to it capturing the resale value of houses. Meanwhile, as I increase the response period form t to t-2, the negative effect between air pollution and housing prices gets stronger (t-statistics is improving). Moreover, the time period t-2 produces a stronger negative relationship between air pollution and housing prices, due to air pollution likely not having an immediate effect on housing prices.

Table 5 Regression Result for *CANSIM* Dataset with first difference

	Δ NHPI (t)	Coefficient	Robust Std. Err.	t-statistics	R-squared	# of observations
(1)	Δ PM2.5 (t) Δ POP (t) Δ I (t) Δ UR (t)	-0.0371343 0.0000911 0.0003409 -1.215359	.143472 .0000916 .0002063 .7050985	-0.26 0.99 1.65 -1.72	within = 0.1655 between = 0.0511 overall = 0.0648	128
(2)	Δ PM2.5 (t-1) Δ POP (t-1) Δ I (t-1) Δ UR (t-1)	-.1512488 0.0000814 0.0001218 -0.6230995	.178383 .0000786 .000123 .4415426	-0.85 1.04 0.99 -1.41	within = 0.0411 between = 0.0338 overall = 0.0087	126
(3)	Δ PM2.5 (t-2) Δ POP (t-2) Δ I (t-2) Δ UR (t-2)	-0.374795 -0.0000385 -0.0001152 -0.4072774	.2741764 .0000591 .0000958 .2680859	-1.37 -0.65 -1.20 -1.52	within = 0.0360 between = 0.0821 overall = 0.0044	125

Table 6 Regression Result for *Teranet* Dataset with first difference

	Δ THPI (t)	Coefficient	Robust Std. Err.	t-statistics	R-squared	# of observations
(1)	Δ PM2.5 (t) Δ POP (t) Δ I (t) Δ UR (t)	.2355294 .0004317 .0005935 -3.528788	.25796 .0001704 .0004108 .8478369	0.91 2.53 1.44 -4.16	within = 0.2536 between = 0.0839 overall = 0.0413	108
(2)	Δ PM2.5 (t-1) Δ POP (t-1) Δ I (t-1) Δ UR (t-1)	-.4530716 .0003823 .0001881 -1.04873	.3465723 .0002061 .0003314 1.051462	-1.31 1.85 0.57 -1.00	within = 0.0818 between = 0.0246 overall = 0.0043	108
(3)	Δ PM2.5 (t-2) Δ POP (t-2) Δ I (t-2) Δ UR (t-2)	-.6100436 -.0000503 -.0002626 -.8056926	.3791404 .0001045 .0002095 .834557	-1.61 -0.48 -1.25 -0.97	within = 0.0338 between = 0.0560 overall = 0.0279	107

4.4 Table 7 (CANSIM) vs Table 8 (Teranet)

I employ the first-difference and logarithmic function in regard to both datasets, to investigate if the percentage changes through the stationary data can provide a statistically significant analysis. Since first-difference removes the unobserved fixed effects between regions, I also apply the random effect (RE) model in this comparison because the RE model is used in the analysis of hierarchical or panel data when one assumes no fixed effects (Gujarati, 2008, p.596). Since the evidence from previous specifications show that t-2 period provides a better result in terms of air pollution- housing price relationship, I only use t-2 period in this comparison. Therefore, the function can be expressed as:

(10) $$\Delta\ln(HPI_{it}) = \alpha_{1i} + \gamma_t + \alpha_2\Delta\ln(PM2.5_{i,t-2}) + \alpha_3\Delta\ln(POP_{i,t-2}) + \alpha_4\Delta\ln(I_{i,t-2}) + \alpha_5\Delta\ln(UR_{i,t-2}) + u_{it}$$

Where $\Delta\ln(HPI_{it})$ represents growth rate in differences of successive HPI; $\Delta\ln(PM2.5_{i,t-2})$ indicates growth rate in differences of successive $PM_{2.5}$ in time period t-2; $\Delta\ln(POP_{i,t-2})$ is growth rate in differences of successive population in time period t-2; $\Delta\ln(I_{i,t-2})$ is growth rate in differences of sequential income in period t-2 and $\Delta \ln(UR_{i,t-2})$ represents growth rate in differences of successive unemployment in the same time period.

The results are striking in both datasets. Form the FE model in CANSIM dataset, one percent increase in $PM_{2.5}$ in period t-2 results in 0.06 percent decrease in NHPI in period t, with a statistically significant coefficient (-1.97). While the FE method of Teranet dataset indicates that with one percent increase in $PM_{2.5}$ in period t-2, THPI decreases by 0.07 percent, along with a statistically significant coefficient (-2.45). In terms of the RE method, results in both data sets show a negative association between air pollution and housing prices, but only the coefficient in Teranet dataset is statistically significant. The coefficients of other control variables are not statistically significant in this comparison, but these coefficients catch the right signs.

We also use Hausman's Specification Test [13] to estimate the appropriate approach between fixed effect and random effect. The discussion and results of this test are shown in Appendix B and C.

Table 7 Regression Result for *CANSIM* Dataset with first difference and Percentage Changes, with Year Dummies

	Δ ln NHPI (t)	**Coefficient**	**Robust Std. Err.**	**t-statistics**	**R-squared**	**# of observations**
(1)FE	Δ ln PM2.5 (t-2) Δ ln POP (t-2) Δ ln I (t-2) Δ ln UR (t-2)	-.0597194 -.8940203 -.005816 -.025595	.0302541 .8326992 .037166 .0309319	-1.97 -1.07 -0.16 -0.83	within =0.3816 between=0.3809 overall = 0.2784	125
(2)RE	Δ ln PM2.5 (t-2) Δ ln POP (t-2) Δ ln I (t-2) Δ ln UR (t-2)	-.046417 .7591435 .0590773 -.0334449	.0301799 .5552277 .0483996 .0355535	-1.54 1.37 1.22 -0.94	within = 0.3576 between=0.1648 overall = 0.3367	125

13. The Hausman's Specification Test evaluates the consistency of an estimator when compared to an alternative, less efficient, estimator which is already known to be consistent (Gujarati, 2008).

Table 8 Regression Result for *Teranet* Dataset with first difference and Percentage Changes, with Year Dummies

	Δ ln THPI (t)	**Coefficient**	**Robust Std. Err.**	**t-statistics**	**R-squared**	**# of observations**
(1)FE	Δ ln PM2.5 (t-2) Δ ln POP (t-2) Δ ln I (t-2) Δ ln UR (t-2)	-.0661696 -.4190338 .0994848 -.0355158	.0269688 .8024631 .0611366 .0400959	-2.45 -0.52 1.63 -0.89	within = 0.5225 between=0.4814 overall = 0.5156	107
(2)RE	Δ ln PM2.5 (t-2) Δ ln POP (t-2) Δ ln I (t-2) Δ ln UR (t-2)	-.0642191 -.240878 .0943613 -.0496023	.0278359 .4163491 .0622844 .0426901	-2.31 -0.58 1.52 -1.16	within = 0.5218 between=0.5082 overall = 0.5165	107

4.5 Table 9 (CANSIM) vs Table 10 (Teranet)

Last, we still use the first-difference and logarithmic function in regard to both datasets, with FE and RE methods, but no year dummies are included. Hausman's Specification Test is applied for both datasets as well, and the discussion and results of Hausman's Test are shown in Appendix D and E. The function can be represented as:

(11) $$\Delta \ln(HPI_{it}) = \alpha_{1i} + \alpha_2 \Delta \ln(PM2.5_{i,t-2}) + \alpha_3 \Delta \ln(POP_{i,t-2}) + \alpha_4 \Delta \ln(I_{i,t-2}) + \alpha_5 \Delta \ln(UR_{i,t-2}) + u_{it}$$

The results from two datasets indicate $PM_{2.5}$ has a negative effect on housing prices, in terms of both FE and RE methods. Unfortunately, the coefficients are not statistically significant. In addition, the sign of population and income in CANSIM FE, Teranet FE and Teranet RE models show a negative impact on housing prices, which is against our expectation.

Table 9 Regression Result for *CANSIM* Dataset with first difference and Percentage Changes, without Year Dummies

	Δ ln NHPI (t)	**Coefficient**	**Robust Std. Err.**	**t-statistics**	**R-squared**	**# of observations**
(1)FE	Δ ln PM2.5 (t-2) Δ ln POP (t-2) Δ ln I (t-2) Δ ln UR (t-2)	-.0429783 -.5589795 -.0551754 -.0337198	.0328376 1.26731 .0644356 .0186899	-1.31 -0.44 -0.44 -1.80	within = 0.0423 between=0.5147 overall = 0.0095	125
(2)RE	Δ ln PM2.5 (t-2) Δ ln POP (t-2) Δ ln I (t-2) Δ ln UR (t-2)	-.031394 .8292146 .0074116 -.0417854	.0260174 .6320979 .116594 .033268	-1.21 1.31 0.06 -1.26	within = 0.0257 between=0.3410 overall = 0.0422	125

Table 10 Regression Result for *Teranet* Dataset with first difference and Percentage Changes, without Year Dummies

	Δ ln THPI (t)	**Coefficient**	**Robust Std. Err.**	**t-statistics**	**R-squared**	**# of observations**
(1)FE	Δ ln PM2.5 (t-2) Δ ln POP (t-2) Δ ln I (t-2) Δ ln UR (t-2)	-.0449547 -.9310981 -.0665336 -.0419323	.0308957 1.674709 .1028279 .0332764	-1.46 -0.56 -0.65 -1.26	within = 0.0462 between=0.1541 overall = 0.0466	107
(2)RE	Δ ln PM2.5 (t-2) Δ ln POP (t-2) Δ ln I (t-2) Δ ln UR (t-2)	-.0400782 -.4556429 -.0582134 -.0556732	.0281785 .6017417 .0894253 .0317043	-1.42 -0.76 -0.65 -1.76	within = 0.0445 between=0.2787 overall = 0.0499	107

4.6 Quadratic Regression

I apply the quadratic regression[14] for both datasets to see if the changes in $PM_{2.5}$ will cause any instantaneous rate of change in HPI. The quadratic formula can be expressed as:

14. A quadratic equation is any equation having the form:
$Y=\beta_0+\beta_1 x+\beta_2 x^2$,
where x represents an unknown, and β_0, β_1, and β_2 represent known numbers such that β_0 is not equal to 0. If $\beta_0 = 0$, then the equation is linear, not quadratic. For more description on quadratic formula, see (Gujarati, 2008)

(12) $HPI_{it} = \alpha_i + \gamma_t + \alpha_2 PM2.5_{i,t} + \alpha_3 PM2.5^2_{i,t} + \alpha_4 POP_{i,t} + \alpha_5 I_{i,t} + \alpha_6 UR_{i,t} + u_{it}$

Where $PM2.5^2_{i,t}$ is the quadratic term of $PM_{2.5}$, α_3 stands for the quadratic coefficient. In order to assess the change in HPI in terms of $PM_{2.5}$, we have to calculate the derivative of the above model as:

(13) $\frac{d(HPI)}{d(PM2.5)} = \alpha_2 + 2\alpha_3 PM2.5$

Which shows a partial effect of $PM_{2.5}$ on HPI. It also indicates the instantaneous rate of change in the expected value of HPI as $PM_{2.5}$ changes, all else being equal.

Figure 5 indicates the quadratic prediction between NHPI and $PM_{2.5}$ from 1997 to 2013; this graph suggests that $PM_{2.5}$ has a positive effect on NHPI until a turning point is reached. Figure 6 illustrates that $PM_{2.5}$ has a negative effect on THPI until a turning point is reached. However, only the result of Teranet is consistent. The result of Teranet dataset in Table 11 indicates that both coefficients of $PM_{2.5}$ and $PM2.5^2_{2.5}$ are statistically significant. We can use the equation (13) and the coefficients in specification (2) of Table 11 to calculate the turning point for THPI: d (HPI)/d (PM2.5) = -7.098 + 0.4612 PM. When we set this equation equals to zero, we can get the turning point value as 15.40, which indicates that $PM_{2.5}$ has a negative effect on THPI until a turning point (15.40) is reached; as the $PM_{2.5}$ exceeds this point, the effect of $PM_{2.5}$ on housing prices becomes positive. Even though there could be many reasons to explain why $PM_{2.5}$ has a positive effect on THPI after the turning point, one important reason may be due to that as $PM_{2.5}$ goes upon to a level, the income effect of pollution becomes more prominent. For example, higher pollutions often comes with high industrial activities, higher population densities and higher income groups.

Figure 5 Quadratic Prediction of NHPI and $PM_{2.5}$, 1997-2013

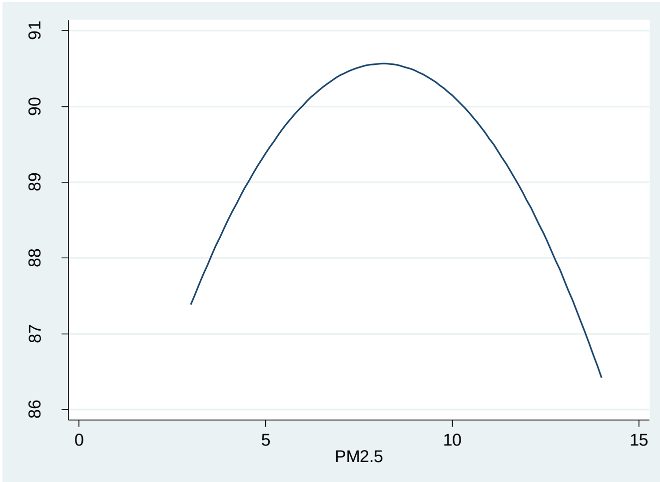


Figure 6 Quadratic Prediction of THPI and $PM_{2.5}$, 1997-2013

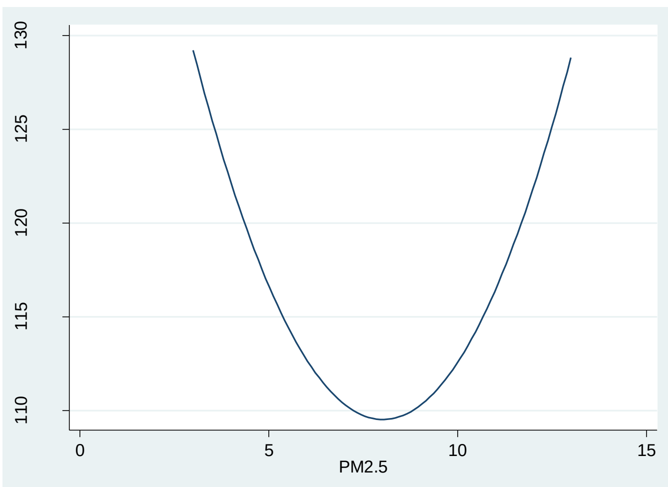


Table 11 Quadratic Regression Result for *CANSIM* and Teranet Dataset, by FE model

	HPI (t)	Coefficient	Robust Std. Err.	t-statistics	R-squared	# of observations
(1)CANSIM	PM2.5 (t) $PM2.5^2$ (t) POP (t) I (t) UR (t)	-.6809899 .0945928 .00000683 .0010341 -1.341107	1.733919 .0926887 .00000568 .0002196 .7762082	-0.39 1.02 1.20 4.71 -1.73	within = 0.9045 between=0.0060 overall = 0.3664	140
(2)Teranet	PM2.5 (t) $PM2.5^2$ (t) POP (t) I (t) UR (t)	-7.097664 .4612477 .0000185 .001134 -1.950786	2.308907 .1080166 .0000376 .0005046 1.763882	-3.07 4.27 0.49 2.25 -1.11	within = 0.9385 between=0.1985 overall = 0.5564	118

In summary, for all the specifications that are used in this study, as I increase the response period from t to t-2, the t-statistics and robust standard error improve, and the results get stronger and more robust. Our findings from FE model of CANSIM dataset (Table 7) and FE, RE models of Teranet dataset (Table 8) indicate that air pollution in time period t-2 has a statistically significant, negative impact on housing prices in period t. Therefore, by applying FE or RE, first-difference and logarithmic function can produce a stronger association between air pollution and housing prices than the other methodologies I tried. However, other specifications do not provide as strong a result as the specification in Table 7 and Table 8, and this may be due to the data limitation and time constraints of $PM_{2.5}$. By nature, in order to see a stronger association between air pollution and property prices, we may need to employ a longer time period, for example, a 30 year time period or even longer, because air pollution may not has an immediate effect on housing price. If we are able to extend the data, the result may be more accurate. In addition, attempting to estimate other pollution sources instead of single pollutant, may provide a stronger result.

5. Conclusions

This is the first study that examines air pollution-housing price relationship in a Canadian context through compiling a comparable and unique panel dataset under employing two different housing price index data: CANSIM New Housing Price Index and Teranet - National Bank House Price Index. In order to provide an accurate and robust analysis, the methodologies this study employs are fixed effect model, random effect model, first-difference approach, robust regression, lagged values and year dummies. The findings show that as I improve the specifications with better HPI (which is THPI), the negative association between air pollution and housing prices surfaces. Meanwhile, the t-statistics improve as the specifications improve. Furthermore, the results from the specifications that applied first-difference, logarithmic function, year dummies, fixed effects or random effects methods, suggest that air pollution has a statistically significant, negative impact on housing prices with a two-year lag. These compelling evidences are shown in the FE model of CANSIM dataset (Table 7) and FE, RE models of Teranet dataset (Table 8). Therefore, we can assume this methodology can produce a stronger association between air pollution and housing prices than the other methodologies. Hausman's Specification Test is also used to estimate the appropriate approach between fixed effect and random effect models. The discussion and results of this test are shown in Appendix 7.2. Overall, this study provides a groundwork for the future research that investigates the air pollution-housing price relationship.

However, the weak results from other specifications may be due to the data limitation and time constraint. If we are able to extend the data, such as by extending the time period to 30 years, we will be able to introduce more lagged values to investigate the non-immediate effect of air pollution to the housing market, and we may also be able to find stronger results when applying other specifications. In addition, quantifying air pollution is not easy. Even though $PM_{2.5}$ is the deadliest form of air pollution that is confirmed by environmentalists, the way people perceive it may be different than scientific measurement. Therefore, maybe $PM_{2.5}$ is not a strong proxy of air pollutants. Attempting to estimate other pollution sources instead of a single pollutant may provide a stronger result. For example, we could use an air quality index that is based on all the air pollutants, which associates with adverse effects on human health and the environment.

Future research could pay attention on these following aspects: In terms of methodology, applying additional indices may contribute to a better result. The closest index available in Canada is the Multiple Listing Service (MLS) Home Price Index. However, this index is only available for the benchmark cities and only starts from year 2005, which seems a short period. In addition, as shown before, in order to achieve a better result, the future study should employ a longer time period for air pollution and housing prices. Attempting to use other pollution sources instead of a single pollutant, may also provide a stronger association between air pollution and housing prices. However, determining which air pollutants are the most effective on influencing housing prices is a task that is left to environmentalists or architects and other experts in the field. Since our study proves that air pollution creates lower housing prices, measures should be taken by policy makers to reduce the air contaminant. For example, if a city has an unhealthy real estate market, the policy makers can address air quality as a possible solution. Last but not least, one benefit of this paper is that it connects environmental protection with economic success.

6. Reference

Bayer, P., Keohane, N., & Timmins, C. (2009). Migration and hedonic valuation: The case of air quality. *Journal of Environmental Economics and Management*, *58*(1), 1-14.

Brasington, D. M., & Hite, D. (2005). Demand for environmental quality: a spatial hedonic analysis. *Regional science and urban economics*, *35*(1), 57-82.

Burda, C. (2014). *Priced Out: Understanding the factors affecting home prices in the Greater Toronto Area*.

Buzzelli, M., Jerrett, M., Burnett, R., & Finklestein, N. (2003). Spatiotemporal perspectives on air pollution and environmental justice in Hamilton, Canada, 1985–1996. *Annals of the Association of American Geographers*, *93*(3), 557-573.

Canadian Council of Ministers of the Environment (2000). *CANADA-WIDE STANDARDS for PARTICULATE MATTER (PM) and OZONE*. Retrieved from: http://www.ccme.ca/files/Resources/air/pm_ozone/pmozone_standard_e.pdf

Chay, K. Y., & Greenstone, M. (1998). *Does air quality matter? Evidence from the housing market* (No. w6826). National Bureau of Economic Research

Cropper, M. L., Deck, L. B., & McConnell, K. E. (1988). On the choice of functional form for hedonic price functions. *The Review of Economics and Statistics*, 668-675.

DiPasquale, D., & Wheaton, W. C. (1992). The cost of capital, tax reform, and the future of the rental housing market. *Journal of Urban Economics*, *31*(3), 337-359.

Environment Canada. (2013). *NAPS Data Products*. Retrieved from National Air Pollution Surveillance Program (NAPS) Catalogue: http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx?lang=en

Environment Canada. (2013). *Particulate Matter*. Retrieved from Criteria Air Contaminants and Related Pollutants Catalogue: http://www.ec.gc.ca/air/default.asp?lang=En&n=2C68B45C-1

Follain, J. R., Hendershott, P. H., & Ling, D. C. (1992). Real Estate Markets since 1980: What Role Have Tax Changes Played? *National Tax Journal*, 253-266.

Gujarati, D. N. (2008). *Basic econometrics*. Tata McGraw-Hill Education.

Hamilton, L. (2008). *Statistics with Stata: Updated for version 8*. University of New Hampshire.

Halvorsen, R., & Pollakowski, H. O. (1981). Choice of functional form for hedonic price equations. *Journal of urban economics*, *10*(1), 37-49.

Harrison, D., & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of environmental economics and management*, *5*(1), 81-102.

Harrison, R. M., & Yin, J. (2000). Particulate matter in the atmosphere: which particle properties are important for its effects on health?. *Science of the total environment*, *249*(1), 85-101.

Jerrett, M., Burnett, R. T., Kanaroglou, P., Eyles, J., Finkelstein, N., Giovis, C., & Brook, J. R. (2001). A GIS-environmental justice analysis of particulate air pollution in Hamilton, Canada. *Environment and Planning A*, *33*(6), 955-974.

Katsouyanni, K., Touloumi, G., Spix, C., Schwartz, J., Balducci, F., Medina, S., & Anderson, H. R. (1997). Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project. *Bmj*, *314*(7095), 1658.

Katz, L., & Rosen, K. T. (1987). The interjurisdictional effects of growth controls on housing prices. *Journal of Law and Economics*, 149-160.

Kim, C. W., Phipps, T. T., & Anselin, L. (2003). Measuring the benefits of air quality improvement: a spatial hedonic approach. *Journal of environmental economics and management*, *45*(1), 24-39.

Malpezzi, S. (1996). Housing prices, externalities, and regulation in US metropolitan areas. *Journal of Housing Research*, *7*, 209-242.

National Bank House Price Index, Methodology. Retrieved from: http://www.housepriceindex.ca/documents/MethodologyEN.pdf

Oswald, A. (1999). The housing market and Europe's unemployment: a non-technical paper. *Homeownership and the labour Market in Europe*.

Raaschou-Nielsen, O., Andersen, Z. J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G., & Cesaroni, G. (2013). Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *The lancet oncology*, *14*(9), 813-822.

Ridker, R. G., & Henning, J. A. (1967). The determinants of residential property values with special reference to air pollution. *The Review of Economics and Statistics*, 246-257.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *The journal of political economy*, 34-55.

Sirmans, S., Macpherson, D., & Zietz, E. (2005). The composition of hedonic pricing models. *Journal of real estate literature*, *13*(1), 1-44.

Smith, V. K., & Huang, J. C. (1995). Can markets value air quality? A meta-analysis of hedonic property value models. *Journal of political economy*, 209-227.

StataCorp LP. (2013). *Stata programming: reference manual: release 13.*

Statistics Canada. (2011). *Average total income, by census family type, 2011 constant dollars, annual (dollars)* (CANSIM *Table 202-0410).* Retrieved from: http://www5.statcan.gc.ca/CANSIM/a26?lang=eng&retrLang=eng&id=2020410&paSer=&pattern=&stByVal=1&p1=1&p2=-1&tabMode=dataTable&csid=

Statistics Canada. (2014). *New housing price index, annually (index, 2007=100)* (CANSIM *Table 327-0046).* Retrieved from: http://www5.statcan.gc.ca/CANSIM/a26?lang=eng&retrLang=eng&id=3270046&paSer=&pattern=&stByVal=1&p1=1&p2=37&tabMode=dataTable&csid=

Statistics Canada. (2015). *New Housing Price Index (NHPI)*. Retrieved from Definitions, data sources and methods Catalogue: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&lang=en&db=imdb&adm=8&dis=2&SDDS=2310#a2

Statistics Canada. (2015a). *Labour force survey estimates (LFS), by census metropolitan area based on 2011 Census boundaries, sex and age group, annual* (CANSIM *Table 282-012).* Retrieved from: http://www5.statcan.gc.ca/CANSIM/a26?lang=eng&retrLang=eng&id=2820129&paSer=&pattern=&stByVal=1&p1=1&p2=-1&tabMode=dataTable&csid=

Statistics Canada. (2015b). *Estimates of population by census metropolitan area, sex and age group for July 1, based on the Standard Geographical Classification (SGC) 2011,annual (persons)* (CANSIM *Table 051-0056).* Retrieved from: http://www5.statcan.gc.ca/CANSIM/a26?lang=eng&retrLang=eng&id=3270046&paSer=&pattern=&stByVal=1&p1=1&p2=37&tabMode=dataTable&csid=

Taylor, E., & McMillan, A. (2013). *Air Quality Management: Canadian Perspectives on a Global Issue*. Springer Science & Business Media.

Teranet - National Bank House Price Index, (n.d.). Retrieved from: http://www.housepriceindex.ca/default.aspx

University of Toronto. (2013). *CANSIM Guide*, Retrieved from: http://data.library.utoronto.ca/cansim-guide

Zheng, S., Cao, J., & Kahn, M. E. (2011). *China's rising demand for" green cities": Evidence from cross-city real estate price hedonics* (No. w16992). National Bureau of Economic Research.

Zheng, S., Kahn, M. E., & Liu, H. (2010). Towards a system of open cities in China: Home prices, FDI flows and air quality in 35 major cities. *Regional Science and Urban Economics*, *40*(1), 1-10.

7. Appendix

7.1 Appendix A - $PM_{2.5}$ Dataset by Canadian City

YEAR	City	PM2.5
2002	Calgary	5.0
2003	Calgary	7.0
2004	Calgary	5.0
2005	Calgary	4.0
2006	Calgary	5.0
2007	Calgary	4.0
2008	Calgary	4.0
2009	Calgary	7.0
2010	Calgary	8.0
2011	Calgary	8.0
2012	Calgary	8.0
2013	Calgary	9.0
1998	Edmonton	12.0
1999	Edmonton	9.0
2000	Edmonton	8.0
2001	Edmonton	9.0
2002	Edmonton	7.0
2003	Edmonton	8.0
2004	Edmonton	7.0
2005	Edmonton	5.5
2006	Edmonton	5.0
2007	Edmonton	5.0
2008	Edmonton	6.0
2009	Edmonton	6.0
2010	Edmonton	13.0
2011	Edmonton	9.0
2012	Edmonton	8.0
2013	Edmonton	7.0
2003	Ottawa-Gatineau	3.0
2004	Ottawa-Gatineau	7.0
2005	Ottawa-Gatineau	8.0
2006	Ottawa-Gatineau	6.0
2007	Ottawa-Gatineau	6.0
2008	Ottawa-Gatineau	5.0
2009	Ottawa-Gatineau	5.5
2010	Ottawa-Gatineau	5.5
2011	Ottawa-Gatineau	6.0
2012	Ottawa-Gatineau	7.0
2013	Ottawa-Gatineau	8.0
1998	Hamilton	13.0
1999	Hamilton	10.0
2000	Hamilton	9.0
2001	Hamilton	8.0
2002	Hamilton	9.0
2003	Hamilton	10.0
2004	Hamilton	9.0
2005	Hamilton	10.0
2006	Hamilton	8.0
2007	Hamilton	8.0
2008	Hamilton	7.0

2009	Hamilton	6.0
2010	Hamilton	6.0
2011	Hamilton	7.0
2012	Hamilton	6.0
2013	Hamilton	9.0
1997	Montreal	7.0
1998	Montreal	7.0
1999	Montreal	7.0
2000	Montreal	7.0
2001	Montreal	8.0
2002	Montreal	8.0
2003	Montreal	8.0
2004	Montreal	8.0
2005	Montreal	9.0
2006	Montreal	7.0
2007	Montreal	7.0
2008	Montreal	10.5
2009	Montreal	10.3
2010	Montreal	10.3
2011	Montreal	9.3
2012	Montreal	9.0
2013	Montreal	9.3
1997	Toronto-Oshawa	10.0
1998	Toronto-Oshawa	10.0
1999	Toronto-Oshawa	10.0
2000	Toronto-Oshawa	8.0
2001	Toronto-Oshawa	9.0
2002	Toronto-Oshawa	9.0
2003	Toronto-Oshawa	8.0
2004	Toronto-Oshawa	8.0
2005	Toronto-Oshawa	8.0
2006	Toronto-Oshawa	8.0
2007	Toronto-Oshawa	7.7
2008	Toronto-Oshawa	6.9
2009	Toronto-Oshawa	5.5
2010	Toronto-Oshawa	6.5
2011	Toronto-Oshawa	5.5
2012	Toronto-Oshawa	5.5
2013	Toronto-Oshawa	7.5
1998	Quebec	7.0
1999	Quebec	11.0
2000	Quebec	8.0
2001	Quebec	8.0
2002	Quebec	8.0
2003	Quebec	7.5
2004	Quebec	8.0
2005	Quebec	9.0
2006	Quebec	8.3
2007	Quebec	7.7
2008	Quebec	7.3
2009	Quebec	9.7
2010	Quebec	8.7
2011	Quebec	9.0
2012	Quebec	10.0
2013	Quebec	9.3
1999	Vancouver	6.0

2000	Vancouver	6.0
2001	Vancouver	5.0
2002	Vancouver	6.0
2003	Vancouver	5.5
2004	Vancouver	5.7
2005	Vancouver	5.7
2006	Vancouver	4.7
2007	Vancouver	4.7
2008	Vancouver	4.7
2009	Vancouver	5.0
2010	Vancouver	4.0
2011	Vancouver	4.3
2012	Vancouver	4.0
2013	Vancouver	5.7
2001	Victoria	4.0
2002	Victoria	5.0
2003	Victoria	6.0
2004	Victoria	6.0
2005	Victoria	5.5
2006	Victoria	5.7
2007	Victoria	4.5
2008	Victoria	4.0
2009	Victoria	4.5
1999	Windsor	14.0
2000	Windsor	10.0
2002	Windsor	12.0
2003	Windsor	10.0
2004	Windsor	10.0
2005	Windsor	10.0
2006	Windsor	9.0
2007	Windsor	10.0
2008	Windsor	9.0
2009	Windsor	7.0
2010	Windsor	8.0
2011	Windsor	8.0
2012	Windsor	8.0
2013	Windsor	10.0
1997	Winnipeg	5.0
1998	Winnipeg	6.0
1999	Winnipeg	6.0
2000	Winnipeg	6.0
2001	Winnipeg	6.0
2002	Winnipeg	6.0
2003	Winnipeg	6.0
2004	Winnipeg	5.0
2005	Winnipeg	5.0
2006	Winnipeg	5.0
2007	Winnipeg	5.0
2008	Winnipeg	5.0
2009	Winnipeg	5.0
2010	Winnipeg	7.0
2011	Winnipeg	7.0
2012	Winnipeg	7.0
2013	Winnipeg	6.0

7.2 Results of Hausman's Test

The null is of Hausman's Specification test is that the two estimation methods are both good and that therefore they should yield coefficients that are "similar". The alternative hypothesis is that the fixed effects estimation is appropriate and the random effects estimation is not; if this is the case, then we would expect to see differences between the two sets of coefficients.

In Appendix B and C, since both of the computed chi^2 (with a degrees of freedom equals 9) are less than the critical chi^2 ((16.919) at 5% level, therefore, we cannot reject the null hypothesis, so both methods are good for both datasets.

Appendix B – Hausman's Test for Table 7 (CANSIM dataset)

	Coefficients (b) fixed	(B) .	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
lnpm25				
L2D.	-.0597194	-.046417	-.0133024	.0062699
lnpopulation				
L2D.	-.8940203	.7591435	-1.653164	.8614243
lnaverageт~e				
L2D.	-.005816	.0590773	-.0648933	.0222677
lnunemploy~e				
L2D.	-.025595	-.0334449	.0078499	.0126112
year				
2001	-.0043289	-.0051918	.0008629	.0056853
2002	.010974	.0041402	.0068339	.0059591
2003	.038031	.0267254	.0113056	.0074117
2004	.0447192	.0331822	.011537	.0077506
2005	.0365143	.0291213	.0073931	.0073584
2006	.0734077	.0641052	.0093025	.0074505
2007	.0477634	.0380323	.0097311	.0073872
2008	.0053262	-.0053985	.0107247	.007557
2009	-.0426353	-.0516793	.009044	.0075151
2010	.0006015	-.0051356	.0057371	.0073063
2011	.005863	-.000523	.006386	.0093924
2012	.0079414	-.0019205	.0098619	.0071631
2013	-.007782	-.0152803	.0074983	.0071041

```
                 b = consistent under Ho and Ha; obtained from xtreg
  B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test:  Ho:  difference in coefficients not systematic

              chi2(9) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =       16.05
            Prob>chi2 =      0.0659
            (V_b-V_B is not positive definite)
```

Appendix C - Hausman's Test for Table 8 (Teranet dataset)

	Coefficients (b) fixed	(B) .	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
lnpm25				
L2D.	-.0661696	-.0642191	-.0019504	.0064055
lnpopulation				
L2D.	-.4190338	-.240878	-.1781558	.9856455
lnaverageт~e				
L2D.	.0994848	.0943613	.0051236	.0215529
lnunemploy~e				
L2D.	-.0355158	-.0496023	.0140865	.0120567
year				
2001	.0147366	.0116975	.0030391	.0067376
2002	.0381237	.0346765	.0034472	.0068833
2003	.0652216	.062621	.0026006	.0085074
2004	.0738594	.0738265	.0000329	.0086679
2005	.0648845	.0644396	.0004449	.0077255
2006	.0932508	.0909135	.0023373	.0078536
2007	.0988957	.0954166	.003479	.0075512
2008	.0202266	.0168084	.0034182	.0079135
2009	-.0523374	-.0532836	.0009462	.0081548
2010	.0491801	.0466997	.0024804	.0076728
2011	.0326888	.0349244	-.0022356	.0098918
2012	.0235041	.0211878	.0023164	.0078309
2013	-.002508	-.0056101	.0031021	.007391

```
                          b = consistent under Ho and Ha; obtained from xtreg
           B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(9) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        4.91
                Prob>chi2 =      0.8418
                (V_b-V_B is not positive definite)
```

Furthermore, In Appendix D and E, both of the computed chi^2 (with a degrees of freedom equals 9) are less than the critical chi^2 (9.488) at 5% level, therefore, we cannot reject the null hypothesis, both methods are good for both datasets.

Appendix D - Hausman's Test for Table 9 (CANSIM dataset)

```
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed          .          Difference          S.E.
-------------+----------------------------------------------------------------
      lnpm25 |
        L2D. |   -.0429783     -.031394       -.0115843        .0064671
lnpopulation |
        L2D. |   -.5589795     .8292146       -1.388194        .9370079
lnaveraget~e |
        L2D. |   -.0551754     .0074116       -.0625869        .0224618
lnunemploy~e |
        L2D. |   -.0337198    -.0417854        .0080656        .0082194
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        8.78
                Prob>chi2 =      0.0669
```

Appendix E - Hausman's Test for Table 10 (Teranet dataset)

```
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed          .          Difference          S.E.
-------------+----------------------------------------------------------------
      lnpm25 |
        L2D. |   -.0449547    -.0400782       -.0048764        .0081201
lnpopulation |
        L2D. |   -.9310981    -.4556429       -.4754552        1.204861
lnaveraget~e |
        L2D. |   -.0665336    -.0582134       -.0083203        .0244515
lnunemploy~e |
        L2D. |   -.0419323    -.0556732        .0137409        .0115958
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        2.82
                Prob>chi2 =      0.5881
```