BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2018-18

Housing Price Network Effects from Public Transit Investment: Evidence from Vancouver

By Alex W. Chernoff and Andrea Craig

www.bank-banque-canada.ca

Bank of Canada Staff Working Paper 2018-18

April 2018

Housing Price Network Effects from Public Transit Investment: Evidence from Vancouver

by

Alex W. Chernoff[1] and Andrea Craig[2]

[1] International Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
achernoff@bankofcanada.ca
[2] Queen's University
craiga@econ.queensu.ca

ISSN 1701-9397

Acknowledgements

The views in this paper are those of the authors and do not necessarily reflect those of the Bank of Canada. The content of this paper has been vetted by Statistics Canada and passed the compliance requirements of the agency's confidentiality protection and disclosure.

Abstract

In this paper, we estimate the effect on housing prices of the expansion of the Vancouver SkyTrain rapid transit network during the period 2001–11. We extend the canonical residential sorting equilibrium framework to include commuting time in the household utility function. We estimate household preferences in the sorting model using confidential micro data and geographic information systems (GIS) data on the SkyTrain network. Using these preference estimates and observed data for 2001, we simulate the equilibrium effects of expanding the SkyTrain. In our counterfactual analysis, the SkyTrain expansion increases housing prices not only in neighborhoods where the expansion occurred, but also in those with access to pre-existing segments of the network. We show how these network housing price effects depend on household commuting patterns, and discuss the implications of our results for targeted taxation policies designed to capture the housing price appreciation stemming from a public transit investment.

Bank topics: Asset pricing; Economic models; Housing
JEL codes: H41, R21, R41

Résumé

Dans cette étude, nous analysons l’incidence qu’a eue sur les prix des logements l’extension de SkyTrain (le réseau de transport en commun rapide de Vancouver), qui a eu lieu entre 2001 et 2011. À cet effet, nous enrichissons le modèle standard d’analyse de la répartition résidentielle de façon à intégrer le temps de trajet dans la fonction d’utilité des ménages. Nous estimons ensuite les préférences des ménages dans le modèle de répartition à l’aide de microdonnées confidentielles et de données de systèmes d’information géographique sur le réseau SkyTrain. À partir de ces estimations et de données observées remontant à 2001, nous simulons enfin les effets d’équilibre produits par l’extension du réseau SkyTrain. Notre analyse contrefactuelle indique que l’extension de SkyTrain a engendré un renchérissement des logements non seulement dans les quartiers touchés par ce prolongement, mais également dans ceux desservis par des lignes préexistantes du réseau. Nous montrons que les effets de réseau générés sur les prix des logements dépendent des habitudes de déplacement des ménages et nous abordons les implications de nos résultats pour des politiques fiscales ciblées qui viseraient à tirer profit de l’appréciation du prix des logements découlant d’investissements dans les transports en commun.

Sujets : Évaluation des actifs; Modèles économiques; Logement
Codes JEL : H41, R21, R41

Non-Technical Summary

This paper estimates the equilibrium effects of a large, publicly funded investment in rapid transit on housing price differentials between neighborhoods within and outside the transit corridor, based on our study of the expansion of the Vancouver SkyTrain rapid transit network over the period 2001–11. This includes the `Millennium Line' and `Canada Line' expansions, public investments totaling $3.2 billion. We use micro data from Canadian censuses (2001 and 2006) and the National Household Survey (2011), which include information on households' place of work and residence and other socio-economic variables. We combine these data with geographic information systems (GIS) data on the SkyTrain expansion sourced from the South Coast British Columbia Transportation Authority.

Using these data, we estimate an equilibrium sorting model in which households choose the neighborhood that maximizes their utility conditional on the exogenous and endogenous characteristics of each neighborhood. We extend the canonical residential sorting equilibrium framework to include commuting time in the household utility function. This extension adds realism to our analysis and allows us to estimate the equilibrium effects of the SkyTrain expansion using counterfactual simulation.

Our counterfactual analysis uses observed data from 2001 and our estimates of household utility preferences to quantify the effect of expanding the SkyTrain over the period 2001–11. Our results indicate that the expansion increased the monthly housing cost differential between neighborhoods with and without access to the network by $3.95. Much of this appreciation is driven by price increases in neighborhoods where the expansion occurred, where there is an increase of $5.07 in the monthly housing cost differential relative to the rest of the region. However, neighborhoods with access to the pre-existing SkyTrain network are also affected, as the monthly housing cost differential between these neighborhoods and the rest of the Vancouver region increased by $1.42.

Our results have important implications for land value capture (LVC) taxation policy. LVC is a tax on the portion of land appreciation directly attributable to a public investment. Our results show that rapid transit expansion projects affect housing prices across the transit network. Importantly, these network effects include appreciation in parts of the network that are not in the area where the expansion occurs. Therefore, LVC tax policies should be structured to reflect the housing price appreciation across the entire transit network, and not focus solely on local taxation in the area where the expansion occurs. Second, our results suggest that tax policies whose structure is based on the direct effects of improved transit access may be successful in capturing the majority of housing price appreciation from public transit investment. In particular, we find that the pattern of housing price appreciation is largely unaffected by changes in the endogenous neighborhood characteristics. This is important because predicting how households will relocate in response to a transit investment is a difficult challenge facing policy makers.

1 Introduction

Rapid transportation lines are often multi-billion dollar public investments that create unearned land appreciation for certain landowners. A number of cities, including New York and Vancouver, are considering targeted tax policies designed to capture the private gains of land appreciation that occur from public transportation investment (Dwyer, 2017; Cheung, 2017). Land value capture (LVC) is a tax on the portion of land appreciation directly attributable to a public investment. Designing LVC tax policies requires knowledge of how neighborhood-specific public investments affect local housing prices, and of how the value of the investment propagates and affects housing prices in other neighborhoods throughout the region. This is especially true of public transit investments, since the transit network itself provides a direct mechanism for the value of public investment to propagate across a region. This paper estimates the equilibrium effects of a large rapid transit investment on housing price differentials between neighborhoods within and outside of the transit corridor.

We study the expansion of the Vancouver SkyTrain rapid transit network over the period 2001–11. This period includes the 'Millennium Line' and 'Canada Line' expansions, investments in public transit infrastructure totaling \$3.2 billion.[1] We use micro data from Canada censuses (2001 and 2006) and the National Household Survey (2011), which include information on households' place of work and residence, and socio-economic variables. We combine these data with GIS data on the SkyTrain expansion sourced from the South Coast British Columbia Transportation Authority.

Using these data, we estimate an equilibrium sorting model closely following the theoretical and empirical methodology of Bayer and McMillan (2012).[2] In the model, households choose the neighborhood that maximizes their utility conditional on the exogenous and endogenous characteristics of each neighborhood. We extend Bayer and McMillan's (2012) framework by incorporating commuting time in the households' utility function.[3] This specification adds realism to our equilibrium analysis and allows us to estimate the equilibrium effects of the expansion of the SkyTrain network using counterfactual simulation. Bayer and McMillan (2012) model household utility as a

[1] The capital cost of the Millennium Line was \$1.2 billion (Wales, 2008) and the capital cost of the Canada Line was \$2 billion (Partnerships British Columbia, 2010). The Canada Line is a public-private partnership.

[2] Our estimation strategy uses Berry et al.'s (1995) two-stage estimator, following the methodology developed by Bayer et al. (2004). Bayer et al.'s (2004) methodology for estimating sorting models has been applied in many papers in the urban economics literature.

[3] In related work, Craig (2018) simulates non-price impacts of proposed public transit infrastructure by modelling commute mode and residential location as a simultaneous choice.

function of commute distance rather than time. It is not possible to explicitly model the effects of changes to the public transit network using Bayer and McMillan's (2012) model because distance is fixed and does not depend on the transit infrastructure.

Our counterfactual analysis uses observed data from 2001 and our estimates of household utility preferences to quantify the effects of expanding the SkyTrain in the manner that occurred over the period 2001–11. Our results indicate that the SkyTrain expansion increased the monthly housing cost differential between neighborhoods with and without access to the network by an average of \$3.95.[4] Much of this appreciation was driven by price increases in neighborhoods where the expansion occurred; in these there was an increase of \$5.07 in the monthly housing cost differential relative to the rest of the region. However, neighborhoods with access to the pre-existing SkyTrain network were also affected, as the monthly housing cost differential between these neighborhoods and the rest of the Vancouver region increased by an average of \$1.42.

To evaluate the importance of household sorting in response to the SkyTrain expansion we consider two counterfactual equilibria: i) a short run equilibrium where the endogenous neighborhood characteristics are fixed; ii) a sorting equilibrium where both housing prices and endogenous neighborhood characteristics adjust. The majority of the price adjustment comes from the change in public transportation access, whereas the second order effects of changes in endogenous neighborhood characteristics induce only minor changes to the equilibrium vector of housing prices.

Our results have important implications for LVC taxation policy. Our results show that rapid transit expansion projects affect housing prices across the transit network. Importantly, these network effects include appreciation in parts of the network that are not in the area where the expansion occurs. Therefore, LVC tax policies should be structured to reflect the housing price appreciation across the entire transit network, and not focus solely on local taxation in the area where the expansion occurs. Second, our results suggest that tax policies whose structure is based on the direct effects of improved transit access may be successful in capturing the majority of housing price appreciation from public transit investment. In particular, we find that the pattern of housing price appreciation is largely unaffected by changes in the endogenous neighborhood characteristics. This is important because predicting how households will relocate in response to a transit investment is a difficult challenge facing policy makers.

[4] While this value is small relative to the average monthly cost of housing in Vancouver, it is important that we estimate the *differential* change in neighborhood housing prices because the level change in housing prices from the transit investment is not identified in our econometric analysis. Instead, our estimates show the housing price increase in neighborhoods with access to the network relative the the rest of the region.

Our focus on housing prices is related to the extensive literature analyzing the effects of public transit lines on real estate prices, such as McMillen and McDonald (2004) and Gibbons and Machin (2005).[5] In modeling commute times, our work is also related to Baum-Snow and Kahn (2000), who develop a measure of public transportation access and estimate the effect of improved access on dwelling prices. Another paper that is closely related to our research is Cohen and Brown (2017), who also study the Canada Line expansion of the SkyTrain and its effect on the value of different types of commercial property in Vancouver. These papers largely rely on reduced form regression analysis to identify the effect of transit investment on housing prices. We complement this literature by using a residential sorting equilibrium model to estimate the effects of a rapid transit line expansion on housing prices. To our knowledge, our counterfactuals provide the first equilibrium analysis of how the value of a large transit investment propagates and affects real estate prices across neighborhoods within a region. These network effects are important, as we show that housing prices in neighborhoods with pre-existing access to the SkyTrain network experience an appreciation relative to the rest of the Vancouver region.

The economic literature on LVC policy is quite limited. Salon and Shewmake (2011) review the literature on the impact of public transportation investment on property values and discuss the theoretical foundation for subsidizing public transportation through LVC. Despite the limited economic research on this topic, LVC policies are increasingly used to capture the real estate appreciation that results from rapid transit investments. For example, the Mass Transit Railway Corporation in Hong Kong owns properties adjacent to train stations, directly capturing value from the public transit infrastructure (Padukone, 2013). As noted above, other cities such as New York and Vancouver are currently contemplating LVC tax policies. The results of our analysis have important implications for LVC policies and therefore help to fill the void in the economics literature on this topic.

The remainder of the paper is organized as follows: section 2 describes the empirical setting and data used in this paper; section 3 describes the theoretical model, as well as estimation and identification; section 4 presents the results of the empirical application; and section 5 concludes.

[5] See Higgins and Kanaroglou (2016) for a review.

2 Empirical Setting and Data

Vancouver's first SkyTrain rapid transit line opened in 1986, and extensions to this line opened between 1989 and 1994. This first SkyTrain line is now called the Expo Line and is shown in Figure 1 in black. The Millennium Line opened in segments between 2002 and 2006, and is shown in Figure 1 in blue. The Canada Line, shown in red, opened in 2009 and connects downtown Vancouver to the suburb of Richmond and the airport.

Our analysis includes 415 census tracts (CTs), corresponding to neighborhoods in our model, in the Vancouver and Abbotsford-Mission census metropolitan areas (CMAs). CTs are small geographic areas with relatively stable boundaries and usually have a population of 2,500 to 8,000 persons (Statistics Canada, 2012). As population density increases, CTs are subdivided. Sixteen percent of the 2001 CTs in these CMAs were split into multiple tracts by 2011. We create a constant geography for our analysis by using the 2001 CT boundaries to reconstitute CTs that were subdivided in the 2006 and 2011 censuses.

Our household data come from confidential micro data from Statistics Canada's 2001 and 2006 long-form censuses and the 2011 National Household Survey (NHS). Most importantly, these data include each household's residential and workplace location at the CT level.[6] These data also include household income.[7] The NHS includes households' commute durations, but neither the 2001 nor the 2006 census includes these data; in section 3.2 we describe how we address this empirical challenge.

We calculate neighborhood characteristics from the Statistics Canada's micro data and Fraser Institute school ratings.[8] For each CT-year, we use the micro data to calculate characteristics of the representative dwelling, including the percentage of dwellings in disrepair, the mean age of dwellings, and the mean number of bedrooms. To calculate the mean monthly dwelling price, we use rent and annual property taxes from these micro data. Using municipal tax rates and the reported property taxes, we infer assessed property values. To have comparable dwelling prices across owner-occupiers and renters, we estimate the monthly user cost of housing for owner-occupiers following Bayer and McMillan (2012).[9] Using the Fraser Institute's school performance ratings, we calculate

[6] We refer to economic families and persons not in an economic family as households. We designate the household's work location as the primary maintainer's workplace location.

[7] All values are in constant 2001 dollars.

[8] The school ratings data were downloaded from the Fraser Institute's website: https://www.fraserinstitute.org/school-performance

[9] Specifically, we regress the natural logarithm of monthly rent or assessed dwelling price on an owner-occupier

primary school quality and secondary school quality based on the closest three respective schools to each CT centroid.

3 Theoretical Model and Estimation Methodology

3.1 Theoretical Model

The theoretical framework for our analysis closely follows the sorting model of Bayer and McMillan (2012). We consider an economy of N households that choose a neighborhood from a set of $\{1, ..., H\}$ alternatives, to solve the following utility maximization problem:

$$\max_{h \in H} V_h^i = \alpha_X^i X_h + \alpha_I^i Inc_h + \alpha_P^i Price_h + \alpha_c^i commute_{i,h} + \xi_h + \epsilon_h^i, \tag{1}$$

where V_h^i is household $i's$ indirect utility function; X_h is a matrix of neighborhood characteristics that are exogenous to households' neighborhood choice; Inc_h and $Price_h$ are the average levels of household income and housing prices in neighborhood h, which are endogenous to the equilibrium of the model; $commute_{i,h}$ is the commute time required for household i to travel from neighborhood h to their workplace; ξ_h is the unobserved quality of neighborhood h; ϵ_h^i is an unobservable taste shock that is specific to individual i and neighborhood h; and, α_j^i is a heterogeneous preference parameter of individual i for attribute j, $j \in \{X, I, P, c\}$. The preferences are defined as $\alpha_j^i = \alpha_{0,j} + \alpha_{1,j} inc_i$, such that each household's preferences vary with their heterogeneous income level.

The probability that household i chooses neighborhood h depends on the household's heterogeneous characteristics and the full vectors of neighborhood characteristics across all H neighborhoods, $P_h^i = f_h(inc_i, \boldsymbol{X}, \boldsymbol{Inc}, \boldsymbol{Price}, \boldsymbol{commute}, \boldsymbol{\xi})$. The functional form of this probability function depends on the distributional assumption over the taste shock, ϵ_h^i. We follow Bayer and McMillan (2012) in defining a sorting equilibrium as an $H \times 1$ vector, $\boldsymbol{Price}$, and an $N \times H$ matrix of choice probabilities, $\boldsymbol{P}$, such that: i) households choose the neighborhood that maximizes utility; ii) housing demand equals supply in each neighborhood, $\sum_{i=1}^N P_h^i = stock_h$, for all $h \in \{1, ..., H\}$; and, iii) the set of choice probabilities, P_h^i, $\forall i \in \{1, ..., N\}$ and $\forall h \in \{1, ..., H\}$, aggregates households' income up to the equilibrium vector, $\boldsymbol{Inc}$.

Our framework differs from that of Bayer and McMillan (2012) in that we model household

indicator and controls. We estimate this regression for each census subdivision; this allows appreciation expectations to differ across census subdivisions.

utility as a function of commute times, whereas they model household utility over commuting distances. In reality, households' disutility from commuting originates from the opportunity cost of their time. Therefore, specifying household utility as a function of commute time rather than distance brings the model closer to the actual economic factors that households evaluate when choosing their residential location.

A second advantage of modeling commute times is that it provides the basis for counterfactual exercises that more directly capture the effects of a public transit investment.[10] In section 3.2 we propose modeling commute time as a function of commuting distance, population density, and rapid transportation access. This model allows us to simulate the equilibrium effects of the expansion of Vancouver's rapid transit network over the period 2001–11.

3.2 Estimation

The methodology used to estimate the model presented in section 3.1 closely follows the two-step procedure in Bayer and McMillan (2012). The first step uses maximum likelihood estimation, with the likelihood function defined by the assumption that the unobservable taste shock, $\epsilon_h^{i,t}$, follows a Type I extreme value distribution. The superscript t is introduced to denote the year of observation corresponding to the three census years 2001, 2006, and 2011.[11] In the first step we estimate the heterogeneous components of households' preference parameters, $\alpha_{1,j}$, $j \in \{X, I, P, c\}$ and $\alpha_{0,c}$.[12] We also estimate the mean indirect utility associated with each neighborhood, $\delta_{h,t}$.[13] In particular,

[10] Bayer and McMillan's (2012) counterfactual commuting simulations consider the equilibrium effects of reducing household preferences for commuting distance. The authors note that these preference parameter estimates incorporate both the financial and psychic costs associated with commuting. As such, it is not possible to explicitly partial out the effects of a public transit investment using their methodology, which motivates our approach of modeling commute time as a function of public transit infrastructure.

[11] Klaiber and Phaneuf (2010) estimate Bayer et al.'s (2004) sorting model with data from multiple time periods.

[12] The first step of the estimation procedure exploits the variation across households within the data. The parameter $\alpha_{0,c}$ is estimated in this step because the variable $commute_{i,h,t}$ varies across households, based on households' idiosyncratic workplace locations. This contrasts with the other neighborhood-level variables (e.g. school quality), which are constant across households.

[13] We construct household-level characteristics to have mean zero so that the $\delta_{h,t}$ vector can be interpreted as the vector of mean indirect utilities associated with each neighborhood in each year. To estimate each $\delta_{h,t}$, we use the contraction mapping algorithm described in Bayer et al. (2007). This contraction mapping algorithm is defined by the equilibrium market clearing condition that housing demand equals the housing supply in each year, $\sum_{i=1}^{N_t} P_h^{i,t} = stock_{h,t}$. Accordingly, the contraction mapping is implemented separately for each of the three years, 2001, 2006, and 2011. This contrasts with maximum likelihood estimation of the preference parameters, in which we pool the three years of data. For the maximum likelihood estimation routine, we also make use of the independence of irrelevant alternatives assumption that is implicit in Bayer and McMillan (2012) and our model, which allows us to estimate the model with a randomly selected set of 10 percent of the alternatives not chosen by each household, rather than the full choice set.

$\delta_{h,t}$ is defined by the following equation:

$$\delta_{h,t} = \alpha_{0,X} X_{h,t} + \alpha_{0,I} Inc_{h,t} + \alpha_{0,P} Price_{h,t} + \xi_{h,t}. \tag{2}$$

In the second step, we estimate the preference parameters $\alpha_{0,j}$, $j \in \{X, I, P\}$ regressing the first-stage estimates of $\delta_{h,t}$ on the observable neighborhood-level characteristics on the right-hand side of equation (2). However, ordinary least squares (OLS) estimation of equation (2) will produce biased estimates due to the correlation between unobserved neighborhood quality, $\xi_{h,t}$, and the endogenous neighborhood-level characteristics, $Inc_{h,t}$ and $Price_{h,t}$. This is the familiar endogeneity problem that arises when estimating differentiated product demand systems.

To address this endogeneity problem we use an instrumental variables (IV) identification strategy. We construct instruments using data in the censuses and NHS that indicate the census subdivisions (CSDs) where the households resided five years earlier. The instruments are constructed in three steps. First, we define the sub-population of movers in each CT as households that reported living in a different CSD five years earlier. Second, we calculate the median income and value of housing for each CSD in Canada for each year of our data, 2001, 2006, and 2011. Third, for each CT we calculate the average income and housing prices in movers' former CSDs, where the averages are calculated using the median values calculated in step two.

The IV exclusion restriction requires that the unobserved neighborhood-level attributes in $\xi_{h,t}$ are not correlated with our instruments. This assumption is reasonable, given that the instruments are constructed solely from the attributes of CSDs that are different from where households reside. Our identification strategy relies on the assumption that the median characteristics of movers' previous CSDs will be strongly correlated with the characteristics of the neighborhoods they relocate to. This assumption is plausible, provided that households are motivated to choose neighborhoods with similar characteristics when moving.

While these arguments provide support for our identification strategy, the empirical results in section 4.2 raise concern regarding the possibility of weak identification. To address this issue, we construct additional instruments using an approach that is motivated by Bayer and Timmins (2007). The approach to constructing these additional instruments leverages the equilibrium structure of the sorting model, and involves five steps. First, we estimate equation (2) by IV using the original instruments (i.e. the instruments constructed from the data on movers' former CSDs). Second, using the first stage and second stage IV preference parameter estimates, we solve for the equilibrium vector of prices that clears the housing market.[14] Third, we calculate the average income for each CT in each year. Fourth, we re-estimate equation (2) by IV, adding the model-derived prices and average income levels for each CT as additional instruments. Fifth, we iterate on steps two to four until our IV estimates converge.

As previously noted, our model is distinct from that of Bayer and McMillan (2012) in that we incorporate commute times (rather than distances) in the household utility function. We observe commute times in 2011 from the NHS; however, commute times are not available in the census data for 2006 and 2001.[15] We therefore estimate the sorting model using *predicted* commute times, which are generated from the output of the following regression model:

$$ln(commute_{h,g,t}) = \beta_0 + \beta_1 ln(distance_{h,g}) + \beta_2 ln(distance_{h,g}) \times 1(SkyTrain_{h,g,t})$$

$$+\beta_3 ln(population\ density_{j,t}) + \beta_4 ln(population\ density_{k,t}) + \nu_h + \nu_g + u_{h,g,t}, \tag{3}$$

where t denotes the year of the observation; (h, g) are origin-destination CT pairs; (j, k) are origin-destination CSD pairs;[16] $commute_{h,g,t}$ is the median commute time between CT_h and CT_g in year t; $distance_{h,g}$ is the Euclidean distance between centroids of CT_h and CT_g;[17] $1(SkyTrain_{h,g,t})$ is a dummy variable that is equal to one if the distances to the nearest SkyTrain station from the centroid of CTs h and g are each less than 2 kilometers in year t; and ν_h and ν_g are origin and destination fixed effects, respectively.

We estimate regression equation (3) using the observed median commute times from the 2011 NHS. While observed commute times are only available for 2011, all of the explanatory variables in equation (3) are available for 2001, 2006, and 2011. This allows us to use the estimated coefficients and observed data to calculate predicted commute times for each CT origin-destination pair for each of the three census years. These predicted commute times are used in the first stage of estimating the sorting model. In our counterfactual analysis, we also use the parameter estimates from (3)

[14]We follow Bayer and Timmins (2007) in setting the preference parameters associated with average neighborhood income to zero, and by setting the vector of unobserved neighborhood characteristics $\xi_{h,t}$ to zero. In this respect, the equilibrium vector of prices reflects the exogenous characteristics of the CTs in households' choice set.

[15]In the 2011 NHS, we do not observe commute times for every CT pair. When estimating the sorting model, we therefore use predicted commute times for all CT pairs in 2011, as well as for the two earlier census years (2001 and 2006).

[16]CSDs are the next highest level of spatial aggregation above CTs. In our area of interest, there are 19 CSDs and 415 CTs.

[17]We set distance equal to one kilometer for commutes within a CT, $distance_{h,h} = 1$. This is necessary because the Euclidean distance between centroid CT_h and itself is zero, which implies $ln(distance_{h,h}) = -\infty$.

to simulate the effects of the 2001–11 expansions of the SkyTrain on commute times, by setting $1(SkyTrain_{h,g,2001}) = 1(SkyTrain_{h,g,2011})$ and holding all other variables in equation (3) at their observed values for 2001.

Our approach to modeling commute times relies on the assumption that there are no structural changes in model (3) over the observation period, 2001–11. We acknowledge that this assumption is unlikely to hold in reality and therefore we also estimate the sorting model using commute distance in place of commute time, as in Bayer and McMillan (2012). There are no qualitative differences in the preference estimates between the two specifications, and the quantitative differences between them vary only slightly. Given the similarities in the empirical results between our benchmark specification and the model using commute distances, we believe that the core value-added of our approach is in the realism it adds to the counterfactual analysis. In particular, our framework and data enable us to use the model to simulate the equilibrium effects of the SkyTrain expansion over the period 2001–11, by adjusting households' access to the rapid transportation network using the commute time model given by equation (3).

4 Estimation Results and Counterfactual Analysis

4.1 Commute Time Regression Results

Table 1 presents the OLS results for the commute time regression model given by equation (3). The dependent variable is the natural logarithm of the median reported household commute times between CT pairs in the 2011 NHS. Our results indicate that commute times increase with commute distance; however, this effect is mitigated by access to the SkyTrain. A 1 percent increase in distance is associated with a 0.456 percent increase in commute time for CT pairs without SkyTrain access, compared with a 0.452 percent increase for CT pairs with access to the SkyTrain.[18] The population density variables are included in the model to capture congestion effects at the more aggregated CSD level. The coefficients for the population density of the home CSD, j, and the population density of the destination CSD, k, are positive, as expected. However, the coefficient for population density in the destination CSD is much smaller and not statistically significant, indicating that congestion in the home CSD is the more important determinant of commute times. The commute time regression results are used to generate the predicted commute times that are used in estimating the sorting model and in the counterfactual analysis, as per the methodology described in section 3.2

[18]These differences are statistically significant at the 1 percent level.

4.2 Preference Estimates

Table 2 presents the sorting model preference estimates. The preference estimates indicate that households with the average level of income experience disutility from longer commutes; however, this effect is attenuated by having a higher income.[19] The positive coefficient on the interaction between commute time and income is a counterintuitive result, assuming that higher income households have a higher opportunity cost of time. This result may be explained by related research by Craig (2018) who models commute mode and residential location choice using a similar theoretical framework to this paper and data from the 2011 NHS.[20] In contrast with these results, she finds that households' disutility from commute time increases with income. The fact that modeling commute mode choice changes the sign of the interaction between commute time and income may be explained by the fact that higher income households have more flexibility in their mode of transportation. That is, higher income households may be more willing to tolerate a longer commute provided they can use private rather than public transportation.[21]

The two endogenous variables in the sorting model are average monthly dwelling price and average neighborhood income. The average household experiences disutility from neighborhoods with higher dwelling prices, although this effect is less severe for higher income households. This may arise if households with higher income have greater disposable income to spend on housing. For the average household, the preference estimate associated with average neighborhood income is positive, but not statistically significant. The positive coefficient on the interaction between neighborhood and household income is consistent with positive assortative matching based on income.

The next three variables capture the average characteristics of neighborhood dwellings: disrepair, age of housing, and the number of bedrooms. Not surprisingly, households dislike neighborhoods with dwellings in disrepair, and prefer neighborhoods where dwellings have more bedrooms. The magnitude of both effects is intensified with income. The interaction effects may be explained by the fact that higher income households are more likely to be homeowners, and thus are more

---

[19] For the remainder of this section, we refer to a "households with the average level of income" as the "average household."

[20] Data on commute duration are included in the 2011 NHS, but not the 2001 and 2006 censuses, which prevents us from studying commute mode choice in this analysis.

[21] A related explanation is that higher income households prefer suburban neighborhoods located in the outskirts of the city. Yet another explanation is that higher income households may have greater flexibility with regard to when they commute. For example, higher income families may be able leave early and/or come home late by availing themselves of child care services.

directly affected by the costs and benefits associated with dwelling characteristics. The positive coefficient on the age of housing may be explained by the average household having an affinity for "character" homes. The negative coefficient on the interaction with income may again be explained by the higher rate of home ownership amongst high income households, and the higher repair costs associated with older homes.

The last three rows of Table 2 report the preference estimates associated with population density and school quality. The average household prefers neighborhoods with higher population density, presumably capturing the benefits associated with the amenities of metropolitan areas. However the utility from population density decreases with income, perhaps reflecting higher income households' desire to live in residential neighborhoods with larger lot sizes. Households prefer neighborhoods with higher primary and secondary school quality, and their preferences for these characteristics increase with income. The large magnitude of the utility gains from secondary school quality is consistent with research by Ries and Somerville (2010), who show that increases in secondary school performance are associated with higher residential housing prices in Vancouver.

Table 3 reports the first and second stage IV estimates, which are generated in the second step of estimating the sorting model. For purposes of comparison, the first column presents the OLS estimates of equation (2). The next three columns present the IV estimates using the two instruments derived from the characteristics of movers' previous CSDs, as described in section 3.2. The F-Statistics for the excluded instruments are reported at the bottom of the first stage results. For this IV specification the F-Statistics are small enough to raise concern regarding weak identification.

Using the strategy described in section 3.2, we supplement the IV regression model with the two additional instruments derived from the sorting model. The results of this second IV specification are presented in the final three columns of Table 3. The IV diagnostic tests are much improved by adding the additional model-derived instruments. The F-Statistics are sufficiently large to soundly reject the null hypothesis of weak identification at any reasonable level of significance. An additional benefit of the second IV specification is that the additional instruments allow us to report the results of the test of overidentifying restrictions relating to IV exclusion restriction. The p-value for the Hansen J-Statistic is reported at the bottom of Table 3, and indicates that the null hypothesis that the instruments are valid cannot be rejected at the 10 percent level. The final column of Table 3 presents the second stage IV estimates, which are the benchmark preference estimates that are also reported Table 2.

4.3 Model Validation

Prior to performing counterfactual analysis, it is important to evaluate the ability of the estimated model to fit the observed data. Using the estimated preference estimates, we solve the sorting model for an equilibrium set of housing prices and average neighborhood income.[22] In our model validation exercises we regress the observed vector of housing prices on the model-derived vector of housing prices, which we pool over the census years 2001, 2006, and 2011. Table 4 includes two columns that correspond to OLS specifications in levels and in first-differences of this regression. The results indicate that observed housing prices are highly correlated with the modeled prices, as the R-squared statistic in the OLS regression is 0.98. The first difference of the model-derived prices are also highly correlated with the observed first differenced prices, as the R-squared is 0.87. In sum, the results in Table 4 indicate that the modeled housing prices largely capture the cross-sectional and dynamic variation in the observed prices.

4.4 Counterfactual Analysis

In this section, we use our sorting model and preference estimates to study the effects of the expansion of the SkyTrain over the period 2001–11. Each of the counterfactual exercises in this section proceeds in four steps. First, we solve the model using the preference estimates and data on the exogenous household and neighborhood variables for the year 2001. Second, we re-solve the model using counterfactual predicted commute times that incorporate the 2002–06 Millennium and 2009 Canada Line expansions of the SkyTrain. In this second step, all exogenous variables other than SkyTrain access are held constant at their 2001 levels. Third, we normalize both modeled and counterfactual housing price vectors to have the same mean as the observed vector of prices in 2001.[23] Fourth, we examine the housing price changes of different sub-groups in order to identify the differential impact of the transit expansion on neighborhoods on and off the SkyTrain network.

We repeat this four-step procedure in two separate counterfactual exercises. The first exercise is a short run equilibrium analysis in which we solve for the equilibrium vector of housing prices that clears the housing market, holding the average neighborhood income levels fixed. This short run equilibrium exercise thus examines the direct effect of the SkyTrain expansion on housing prices,

[22] As in Bayer and McMillan (2012), nothing guarantees an equilibrium is unique.

[23] An equilibrium is invariant to adding an arbitrary constant to the vector of housing prices, so it is not possible to quantify the mean change in housing prices from the expansion of the SkyTrain. Therefore, we focus on the differential appreciation across different sub-groups of neighborhoods, while normalizing the average change in neighborhood housing prices to zero.

independent of equilibrium effects arising from household sorting. In the second counterfactual exercise, we solve for the sorting equilibrium vectors of housing prices and average income by allowing households to optimally choose the neighborhood that maximizes their utility given the endogenous vector of prices that clears the housing market.

Figure 2 presents a 'heat-map' of the short run equilibrium change in modeled housing prices arising from the expansion of the SkyTrain. Recall that the mean prices before and after the expansion have been normalized to the mean of the observed prices in 2001. Therefore, modeled price changes that are greater than zero (orange-red) depict CTs that experience an appreciation relative to the mean price change from the expansion. Conversely, modeled price changes less than zero (blue-green) depict CTs that experience a relative depreciation relative to the mean price change from the expansion.

CTs with access to the Millennium and Canada Line expansions are clearly visible in Figure 2 in red, reflecting the fact that these CTs experience the largest appreciation. This is not surprising, since households commuting from these CTs enjoy a shorter commute as a result of the expansion.

It is interesting that most CTs on the portion of the SkyTrain that existed prior to the expansion (i.e. the Expo Line) also experience an appreciation in housing prices. Housing prices in these CTs are driven up by the reduced commute times of households that commute to work in CTs with access to expansion lines of the SkyTrain.

Table 5 quantifies the short run equilibrium differential price changes from the SkyTrain expansion across different sub-groups of neighborhoods. The first panel of the table indicates that the SkyTrain expansion results in an increase of \$3.76 to the monthly housing cost differential between neighborhoods with and without access to the rapid transit network. The second panel reports a \$4.61 increase in the monthly housing cost differential between CTs serviced by the expansion lines of the SkyTrain and that of other CTs in the Vancouver region. The final panel reports a \$1.55 increase in the monthly housing cost differential between CTs serviced by the Expo Line and that of other CTs in the region.

Figure 3 presents a heat-map of the sorting equilibrium changes in the modeled housing prices arising from the expansion of the SkyTrain. The heat-map is very similar to Figure 2, indicating that the price changes are mainly driven by changes in rapid transit access. Comparing Tables 5 and 6 confirms that there is little quantitative difference between the short run equilibrium and sorting equilibrium price changes. The similarity between the two equilibria suggest that changes in neighborhoods' average income have a small effect on the price changes.

To summarize, our counterfactual sorting equilibrium analysis indicates that the SkyTrain expansion increased the monthly housing cost differential between neighborhoods with and without access to the network by \$3.95. Much of this appreciation is driven by price increases in CTs with access to the expansion, where there is an increase of \$5.07 in the monthly housing cost differential relative to other CTs in the region. However, prices in CTs with access to the pre-existing Expo Line also appreciated, increasing the monthly housing cost differential between these CTs and the rest of the Vancouver region by \$1.42. Our results suggest that the mechanism driving appreciation is decreased commute times for households commuting to and from CTs along the SkyTrain expansion.

5 Conclusion

This paper studies the expansion of the Vancouver SkyTrain rapid transit network over the period 2001–11. Our results suggest that this public transit investment increased housing prices in neighborhoods in close proximity to where the expansion occurred, but also in neighborhoods with access to pre-existing segments of the network. We find that the relative appreciation of housing prices in neighborhoods with access to pre-existing lines of the SkyTrain was 28 percent as large as the relative appreciation of neighborhoods with access to the expansion lines. A policy implication of this result is that LVC tax policies should consider the potential housing price appreciation across the entire transit network, and not focus only on local taxation in the region where the expansion occurs.

A second policy implication of our analysis relates to the similarities in the counterfactual results between the short run and sorting equilibrium. We find that the pattern of housing price appreciation is largely determined by the change in rapid transit access. Conversely, the change in neighborhoods' average income level in response to the SkyTrain expansion has relatively little effect on housing price appreciation. This is important because it is difficult for policy makers to predict the relocation decisions of households in response to a transit expansion. Our results suggest that LVC tax policies whose design is based on the direct effects of improved transit access may be successful in capturing the majority of housing price appreciation from a public transit investment.

References

Baum-Snow, Nathaniel, and Matthew E. Kahn (2000) 'The effects of new public projects to expand urban rail transit.' *Journal of Public Economics* 77, 241263

Bayer, Patrick, and Christopher Timmins (2007) 'Estimating equilibrium models of sorting across locations.' *The Economic Journal* 117(March), 353–374

Bayer, Patrick, and Robert McMillan (2012) 'Tiebout sorting and neighborhood stratification.' *Journal of Public Economics* 96, 1129–1143

Bayer, Patrick, Fernando Ferreira, and Robert McMillan (2007) 'A unified framework for measuring preferences for schools and neighborhoods.' *Journal of Political Economy* 114(4), 588–638

Bayer, Patrick, Robert McMillan, and Kim Rueben (2004) 'An equilibrium model of sorting in an urban housing market.' *NBER Working Paper No. 10865*

Berry, Steven, James Levinsohn, and Ariel Pakes (1995) 'Automobile prices in market equilibrium.' *Econometrica* 63(4), 841–890

Cheung, Christopher (2017) 'Should developers pay to ride the benefits of transit?' *The Tyee*

Cohen, Jeffrey P., and Mike Brown (2017) 'Does a new rail rapid transit line announcement affect various commercial property prices differently?' *Regional Science and Urban Economics* 66, 74–90

Craig, Andrea (2018) 'Commute mode and residential location choice.' Job Market Paper

Dwyer, Jim (2017) 'The subways made them rich. Is it time for them to pay up?' *The New York Times*

Gibbons, Stephen, and Stephen Machin (2005) 'Valuing rail access using transport innovations.' *Journal of Urban Economics* 57, 148–169

Higgins, Christopher D., and Pavlos S. Kanaroglou (2016) 'Forty years of modelling rapid transits land value uplift in North America: moving beyond the tip of the iceberg.' *Transport Reviews* 36:5, 610–634

Klaiber, H. Allen, and Daniel J. Phaneuf (2010) 'Valuing open space in a residential sorting model of the Twin Cities.' *Journal of Environmental Economics and Management* 60, 57–77

McMillen, Daniel P., and John McDonald (2004) 'Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983-1999.' *Real Estate Economics* 32:3, 463–486

Padukone, Neil (2013) 'The unique genius of Hong Kong's public transportation system.' *The Atlantic*

Partnerships British Columbia (2010) 'Canada Line case study'

Ries, John, and Tsur Somerville (2010) 'School quality and residential property values: Evidence from Vancouver rezoning.' *The Review of Economics and Statistics* 92(4), 928944

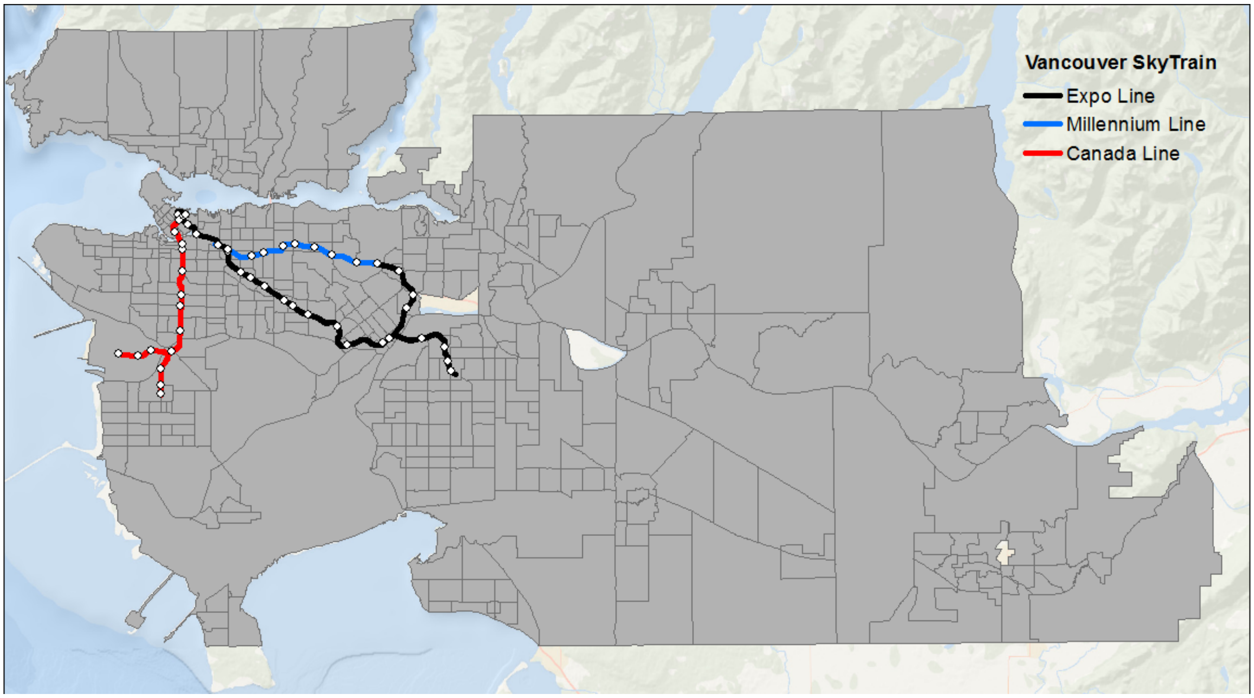
Salon, Deborah, and Sharon Shewmake (2011) 'Opportunities for value capture to fund public transport: A comprehensive review of the literature with a focus on East Asia'

Statistics Canada (2012) *Census Dictionary, 2011* (Ottawa, Ontario). Catalogue no. 98-301-X2011001

Wales, Trevor (2008) 'The road less travelled: Translinks improbable journey from 1999 to 2008.' TransLink

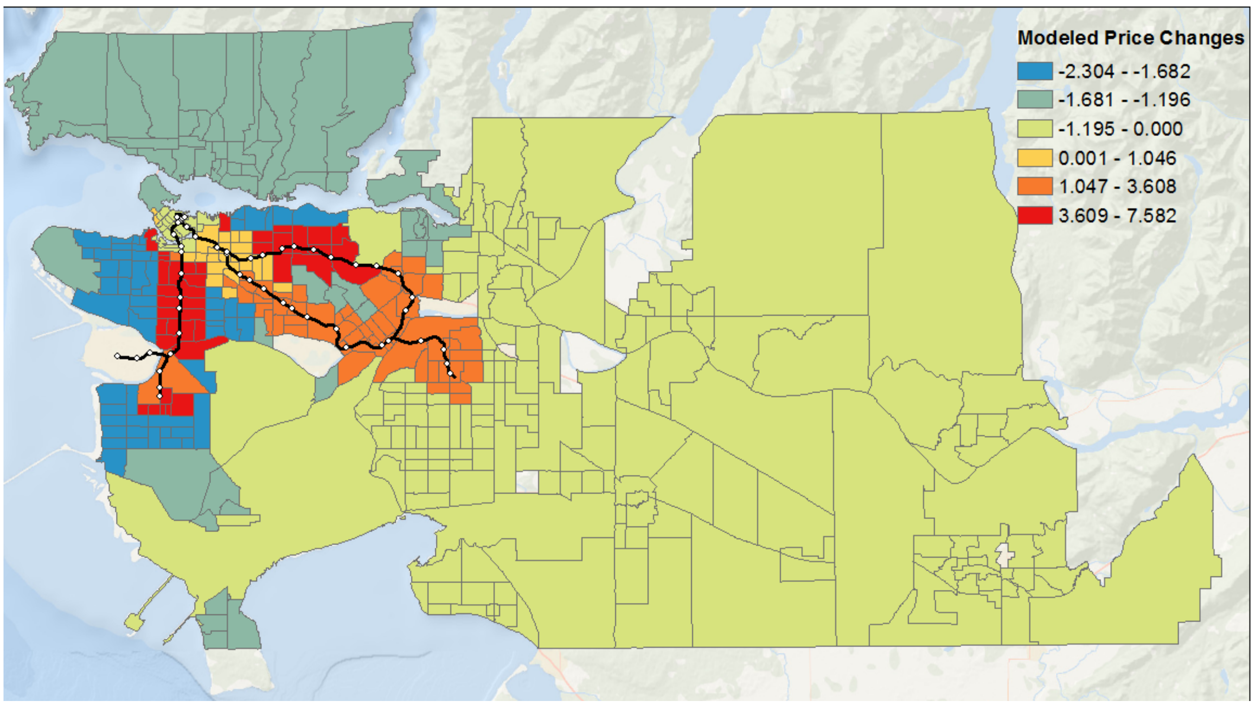
Appendix A - Figures

Figure 1: Vancouver Census Tracts and the SkyTrain Rapid Transit Network, 2011



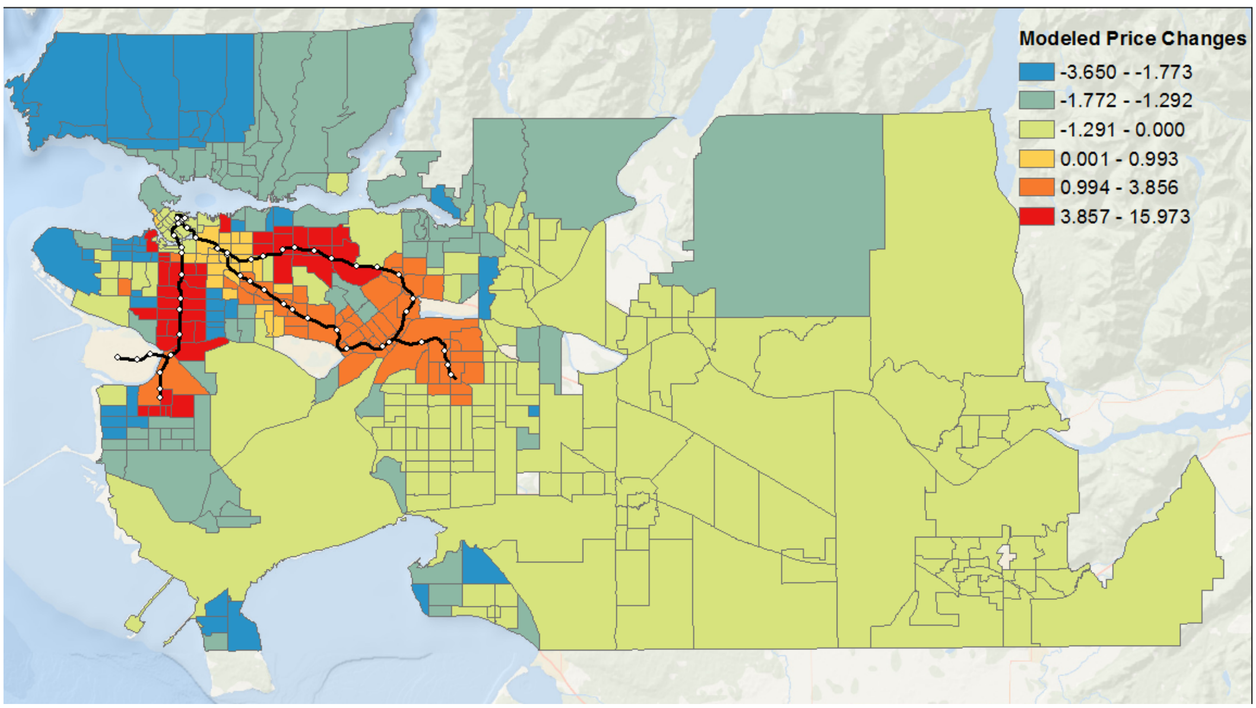
Vancouver and Abbotsford-Mission census tracts are displayed in gray, and have been re-defined to have a constant geography over the 2001, 2006, and 2011 censuses. Geo-spatial files for the census tracts are sourced from Statistics Canada. The three lines of the Vancouver SkyTrain rapid transit system in 2011 are displayed. The white dots define the location of SkyTrain stations. Geo-spatial files for the SkyTrain are sourced from the South Coast British Columbia Transportation Authority. The base map is sourced from ESRI, and uses spatial data from DeLorne, GEBCO, NOAA NGDC, and other contributors.

Figure 2: Short Run Equilibrium Modeled Price Changes



Modeled price changes are calculated by authors using output from the short run equilibrium simulations of the estimated model (see details in the text). The base map is sourced from ESRI, and uses spatial data from DeLorne, GEBCO, NOAA NGDC, and other contributors.

Figure 3: Sorting Equilibrium Modeled Price Changes



Modeled price changes are calculated by authors using output from general equilibrium simulations of the estimated model (see details in text). The base map is sourced from ESRI, and uses spatial data from DeLorne, GEBCO, NOAA NGDC, and other contributors.

Appendix B - Tables

Table 1: Commute Time Regression Results

	$ln(commute_{h,g})$
	$ln(commute_{h,g})$
$ln(distance_{h,g})$	0.456***
	(0.0029)
$ln(distance_{h,g}) \times 1(SkyTrain_{h,g})$	-0.00366***
	(0.0012)
$ln(population\ density_j)$	0.147**
	(0.061)
$ln(population\ density_k)$	0.0558
	(0.092)
Constant	-2.774***
	(0.71)
CT fixed effects (g and h)	Yes
Observations	32,045
R-Squared	0.54

Standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2: Neighborhood Sorting Model Estimation Results

	α_0		α_1
commute	-1.24***	*commute x inc*	0.192***
	(0.001)		(0.003)
Price	-6.919***	*Price x inc*	0.050***
	(0.45)		(0.012)
Inc	1.364	*Inc x inc*	5.113***
	(2.02)		(0.043)
Disrepair	-4.130**	*Disrepair x inc*	-1.075***
	(1.83)		(0.125)
Age of Housing	3.515***	*Age of Housing x inc*	-0.600***
	(0.86)		(0.042)
Number of Bedrooms	1.969***	*Number of Bedrooms x inc*	0.057***
	(0.33)		(0.009)
Population Density	0.499**	*Population Density x inc*	-0.230***
	(0.20)		(0.011)
Primary Schools Quality	5.426***	*Primary Schools Quality x inc*	0.887***
	(0.97)		(0.045)
Secondary Schools Quality	11.52***	*Secondary Schools Quality x inc*	0.442***
	(1.45)		(0.056)
Household Observations	249,985		
Census Tracts	415		

Second stage regression includes constant and year 2006 and 2011 indicator variables. The number of observations has been rounded to base 5, as per Statistics Canada's vetting regulations. Standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3: Neighborhood Sorting Model - Second Stage Estimation and IV Diagnostic Results

Estimator/Stage in 2SLS: Dependent Variable:	OLS δ	$2SLS/1^{st}$ $Price$	$2SLS/1^{st}$ Inc	$2SLS/2^{nd}$ δ	$2SLS/1^{st}$ $Price$	$2SLS/1^{st}$ Inc	$2SLS/2^{nd}$ δ
Price	-1.352***			-4.092***			-6.919***
	(0.079)			(0.86)			(0.45)
Inc	1.639***			4.530			1.364
	(0.27)			(3.58)			(2.02)
Disrepair	-1.913***	-0.295	-0.269***	-2.177	0.184	0.136	-4.130**
	(0.71)	(0.29)	(0.082)	(1.43)	(0.30)	(0.088)	(1.83)
Age of Housing	-3.164***	1.223***	0.120***	-0.246	0.197	0.307***	3.515***
	(0.27)	(0.11)	(0.030)	(1.22)	(0.14)	(0.042)	(0.86)
Number of Bedrooms	0.0930	0.338***	0.140***	0.587	0.0255	0.100***	1.969***
	(0.057)	(0.018)	(0.0052)	(0.62)	(0.029)	(0.0087)	(0.33)
Population Density	-0.227***	0.143***	-0.0107	0.159	0.0406	0.0634***	0.499**
	(0.079)	(0.032)	(0.0091)	(0.16)	(0.039)	(0.012)	(0.20)
Primary Schools Quality	-0.934***	1.117***	0.308***	1.290	-0.0422	0.0335	5.426***
	(0.27)	(0.10)	(0.029)	(1.58)	(0.15)	(0.043)	(0.97)
Secondary Schools Quality	1.915***	1.682***	0.477***	5.178**	0.0361	0.231***	11.52***
	(0.37)	(0.14)	(0.040)	(2.44)	(0.19)	(0.055)	(1.45)
$1(Year = 2006)$	-0.0236	-0.237***	-0.0294***	-0.503**	-0.0669**	-0.0196**	-1.111***
	(0.063)	(0.030)	(0.0085)	(0.21)	(0.030)	(0.0089)	(0.18)
$1(Year = 2011)$	0.126*	0.0908***	0.0473***	0.292	-0.0594*	0.0251**	0.872***
	(0.065)	(0.034)	(0.0098)	(0.29)	(0.034)	(0.010)	(0.21)
Constant	0.877***	-1.797***	-0.267***	-3.718*	0.0755	0.0760	-9.917***
	(0.29)	(0.11)	(0.032)	(1.99)	(0.19)	(0.056)	(1.14)
Movers Inc		-0.478***	0.0338		-0.163	0.0653*	
		(0.12)	(0.034)		(0.11)	(0.033)	
Movers Price		0.379***	0.0840**		0.202	0.0397	
		(0.13)	(0.038)		(0.13)	(0.037)	
Model Inc					0.809	1.713***	
					(0.55)	(0.16)	
Model Price					0.975***	-0.0251	
					(0.071)	(0.021)	
Observations	1,245	1,245	1,245	1,245	1,245	1,245	1,245
R-squared	0.42	0.55	0.68	-0.17	0.61	0.71	-2.60
Number of Instruments:	0	2	2	2	4	4	4
F-Statistic (Price)		8.336	8.336	8.336	54.88	54.88	54.88
F-Statistic (Inc)		5.910	5.910	5.910	31.18	31.18	31.18
p-value (J-Statistic)					0.143	0.143	0.143

Standard errors are in parentheses and are clustered at the census tract level; *** p<0.01, ** p<0.05, * p<0.1. The reported F-Statistics are for first stage F-tests of the excluded instruments, reported individually for each of the endogenous variables.

Table 4: Model Validation Regressions

	$Price\ Observed_h$	$\Delta\ Price\ Observed_h$
$Price\ Modeled_h$	0.927***	
	(0.0037)	
$\Delta Price\ Modeled_h$		0.921***
		(0.012)
Constant	93.21***	5.020*
	(4.97)	(2.57)
Observations	1,245	830
R-squared	0.98	0.87

Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 5: Differential Price Changes Across Neighborhoods - Short Run Equilibrium

	$1(SkyTrain) = 0$	$1(SkyTrain) = 1$	$\|Diff\|$
Mean	-1.24	2.52	3.76***
	(0.0231)	(0.1935)	(0.1949)
Neighborhoods	274	135	

	$1(SkyTrain) = 0\ OR$ $1(Expansion\ Line) = 0$	$1(SkyTrain) = 1\ AND$ $1(Expansion\ Line) = 1$	$\|Diff\|$
Mean	-0.72	3.89	4.61***
	(0.0612)	(0.3157)	(0.3216)
Neighborhoods	345	64	

	$1(SkyTrain) = 0\ OR$ $1(Expansion\ Line) = 1$	$1(SkyTrain) = 1\ AND$ $1(Expansion\ Line) = 0$	$\|Diff\|$
Mean	-0.27	1.28	1.55***
	(0.126)	(0.0965)	(0.1587)
Neighborhoods	338	71	

Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 6: Differential Price Changes Across Neighborhoods - Sorting Equilibrium

	$1(SkyTrain) = 0$	$1(SkyTrain) = 1$	$\|Diff\|$
Mean	-1.30	2.64	3.95***
	(0.035)	(0.236)	(0.2386)
Neighborhoods	274	135	

	$1(SkyTrain) = 0\ OR$ $1(Expansion\ Line) = 0$	$1(SkyTrain) = 1\ AND$ $1(Expansion\ Line) = 1$	$\|Diff\|$
Mean	-0.79	4.28	5.07***
	(0.0637)	(0.3985)	(0.4036)
Neighborhoods	345	64	

	$1(SkyTrain) = 0\ OR$ $1(Expansion\ Line) = 1$	$1(SkyTrain) = 1\ AND$ $1(Expansion\ Line) = 0$	$\|Diff\|$
Mean	-0.25	1.17	1.42***
	(0.1435)	(0.0936)	(0.1714)
Neighborhoods	338	71	

Standard errors are in parentheses; *** p<0.01, ** p<0.05, * p<0.1