Inverse Probability Tilting with Spatial Data: Some Monte Carlo Evidence and an Application to Commercial Real Estate Prices

Jeffrey P. Cohen[*], Ke Yang*[†]

Abstract

Examining the impacts of develpment on one side of a city's border or the other in a cross-sectional data set is important for real estate investments. Proximity to more densely developed urban centers can lead to higher prices per square foot because of agglomeration economies. An important consideration is the identification strategy in the research design. While a study focusing on sales prices data spanning a long period of time should consider an event, together with a proximity variable, as part of an identification strategy for estimating the "Average Treatment Effect" (ATE) of a development location decision, there is a different strand of literature focused on ATE identification in a cross-sectional context. One set of approaches in more general settings is propensity score approaches. Within the context of propensity score approaches, there is an extensive body of literature on Inverse Probability Weighting (IPW), as in Imbens et al (2003), and more recently Inverse Probability Tilting (IPT), as in Graham et al (2012). In particular, the attractive features of IPT that rely on a relatively straightforward method of moments approach have prompted us to explore a more general version of IPT. We consider an additional adaptation to the IPT estimator as a part of our ATE identification strategy - specifically, re-weighting that allows for geographic heterogeneity in a cross sectional context, in addition to a propensity score approach. We present our innovation that incorporates geographic heterogeneity in the data and the adjustments to the weights that we make to allow for more geographically distant observations to be down-weighted relative to more close observations. We call this semi-parametric approach our "Inverse Probability Tilting-Locally Weighted " estimator (IPT-LW). We describe the computation process of the IPT-LW ATE, then provide some Monte Carlo simulation evidence to demonstrate our estimator performs well in small samples. An application of how a cross-section of commercial property prices are impacted by being sold in 2013 in the city limits of Vancouver, BC, Canada (opposed to commercial properties that sold in the Vancouver suburbs in 2013) demonstrates the implementation of the IPT-LW estimator in calculating the ATE of a decision of whether to develop real estate on one side or the

[*]Corresponding author, University of Connecticut, School of Business, 2100 Hillside Road, Unit 1041-RE, Storrs, CT 06269. Jeffrey.Cohen@business.uconn.edu

[†]University of Hartford, Barney School of Business, 200 Bloomfield Ave, West Hartford, CT 06117; Kyang@hartford.edu

other of the city limits border.

Keywords: commercial real estate prices; spatial dependence; non-parametric regression; kernel regression; inverse probability tilting.

JEL Codes: R3

1 Introduction

Examining the impacts of develpment on one side of a city's border or the other in a cross-sectional data set is important for real estate investments. Proximity to more densely developed urban centers can lead to higher prices per square foot because of agglomeration economies. An important consideration is the identification strategy in the research design. While data on sale prices spanning a long period of time often consider an "event", together with a proximity variable, as part of an identification strategy, there is a different strand of literature focused on identification in a cross-sectional context.

One set of approaches in more general settings is propensity score approaches. The finance literature has considered inverse Mills ratio adjustments, for instance. More generally, there is an extensive body of literature on Inverse Probability Weighting (IPW), as in Imbens et al (2003), and Inverse Probability Tilting (IPT), as in Graham et al (2012).

Our objective in this paper is to consider an additional adjustment as a part of the identification strategy – specifically, re-weighting that allows for geographic heterogeneity in a cross sectional context, in addition to a propensity score approach. In particular, the attractive features of IPT that rely on a method of moments approach instead of a Maximum Likelihood approach have prompted us to explore a more general version of IPT. This type of additional adjustment is important in the context of development decisions because the ATE from locating in the city limits can be different than the ATE from locating outside the city limits.

In the remainder of this paper, we first motivate our specific application, and then describe the literature on the theory of IPT and IPW. Next we explain our innovation that incorporates geographic heterogeneity and the adjustments to the weights that we make to allow for more distant observations to be down-weighted relative to more close observations. We call this an "IPT-LW" estimator (representing "Inverse Probability Tilting-Locally Weighted Regressions). We describe the computation process of the IPT-LW estimator, then provide some Monte Carlo evidence to demonstrate that our estimator performs well. Our application of how a cross-section of commercial property prices are impacted by location on one side of the border of the City of Vancouver, BC, Canada (opposed to on the other side, in the suburbs) demonstrates the implementation of the IPT-LW estimator. Finally, we discuss potential future extensions to our approach and summarize our findings.

2 Motivation

Consider the following problem. First, assume we are interested in analyzing a cross-sectional data set on real estate values in a metropolitan area where there is a major city, to determine the property value impact of location on one side or the other of the city limits. Since there is only one time-period, we can consider "treated" vs. "untreated" properties as those that are just on the "inside" of the city limits and those that are just on the "outside" of the city limits, respectively. Then, we can estimate the effect of being in the treatment sub-sample opposed to the non-treated sub-sample, assuming that the sample mean of the treated observations control variables equals the sample mean of the entire sample; and the sample mean of the untreated observations control variables equals the sample mean of the entire sample.

If these last assumptions do not hold in practice, then in order to obtain valid treatment effects, it is necessary to re-weight the data so that we equate the treated sample mean to the entire sample mean, after re-weighting. Similarly, we will need to equate the untreated sample mean to the entire sample mean, after re-weighting. There are several approaches to accomplishing this. One is an Inverse Probability Weighting (IPW) approach, which has received extensive attention in the literature (see, e.g., Imbens et al (2003), and others). IPW uses Maximum Likelihood estimation techniques to obtain the weighting parameters. An attractive alternative is the Inverse Probability Tilting (IPT) approach, as in Graham et al (2012), which is based on a relatively straightforward moments condition technique.

Another important consideration in this type of data is "neighborhood effects", as in Ioannides (2012). Sale prices of properties are typically dependent on nearby comparable properties, so it is more likely that a given property's sale price would move more closely with nearby properties than those further away. Ignoring these dependencies may impact the propensity scores, and it would be appropriate to adjust them accordingly.

The IPT and IPW approaches assume no geographic heterogeneity in the tilting parameters. If the geographic locations of observations are varied, this could be an important consideration in a particular application, and therefore it may be helpful to re-weight a second time, to consider the geographic distance between observations. This is common in the nonparametric literature, specifically, with an approach called Locally Weighted Regressions (LWR). McMillen and Redfearn (2010) describe LWR as well as present an application. However, no known work has incorporated LWR into an IPT framework.

3 Approach

3.1 Model

Suppose that there N units, indexed by i = 1, . . . ,N, viewed as drawn randomly from a large population. We postulate the existence for each unit of a pair of potential outcomes, $Y_i(0)$ for the outcome under the control treatment and $Y_i(1)$ for the outcome under the active treatment. In addition, each unit has a vector of characteristics, referred to as covariates, pretreatment variables or exogenous variables, and denoted by X_i. Each unit is exposed to a single treatment; $D_i = 0$ if unit i receives the control treatment and $D_i = 1$ if unit i receives the active treatment. We therefore observe for each unit the triple (D_i, Y_i, X_i), where Y_i is the realized outcome:

$$Y_i \equiv Y_i(D_i) = \begin{cases} Y_i(0) & if \quad D_i = 0, \\ Y_i(1) \quad | & if \quad D_i = 1. \end{cases}$$

Distributions of (D_i, Y_i, X_i) refer to the distribution induced by the random sampling from the population. The average treatment effect (ATE) is

$$\gamma_0^{ATE} = E[Y(1) - Y(0)].$$

In practice, however, one only observe

$$Y_i = (1 - D)Y_i(0) + DY_i(1)$$

, i.e. only $Y_i(1)$ for actively treated units and $Y_i(0)$ for control treatment units. One common practice in this case to adjust the two sub-samples based on their own distributions.

3.2 Inverse Probability Tilting Estimator

Several additional pieces of notation will be useful in the remainder of these notes. Let N_1 and N_0 denote the number of treated units and control units, respectively. The propensity score (Rosenbaum and Rubin, 1983) is defined as the conditional probability of receiving the treatment,

$$p(x) = Pr(D_i = 1|X_i = x) = E[D_i|X_i = x].$$

Imbens (2004) and Wooldridge (2007) propose the inverse probability weighting ATE estimator as

$$\hat{\gamma}_{IPW}^{ATE} = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{D_i}{G\left((t(x_i)'\hat{\delta}_{ML}\right)} - \frac{1-D_i}{1-G\left(t(x_i)'\hat{\delta}_{ML}\right)}\right\}Y_i \tag{1}$$

where $G(t(x)'\delta_0) = p(x)$ for all $x \in X$ and some δ_0, $t(x)$ is a $1+M$ column vector of known functions of X with a constant as its first element, $\hat{\delta}_{ML}$ is a vector of maximum likelihood estimates of δ_0.

Graham, Pinto and Engel (2012) propose an alternative method to (1) by replacing the

$\hat{\delta}_{ML}$with a particular method of moments estimator, $\hat{\delta}_{IPT}$, that is the solution to

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{D_i}{G\left((t(x_i)'\hat{\delta}_{IPT}^1\right)}-1\right\}t(x_i)=0. \tag{2}$$

Rearranging (2) we have

$$\sum_{i=N_0+1}^{N}\frac{1}{G\left((t(x_i)'\hat{\delta}_{IPT}^1\right)}t(x_i)=\frac{1}{N}\sum_{i=1}^{N}t(x_i).$$

Thus, the motivation of the IPT estimator is to choose the propensity score estimator to equalize the mean of $t(x_i)$ (Note that higher moments of x can be included in $t(x_i)$.) across treated units to that of the full sample.

3.3 Locally Weighted Regression (LWR)

The locally weighted regressions (LWR) approach is commonly used in spatial studies to account for the effects of missing variables that are correlated over space. The basic idea behind LWR is to assign higher weights to observations near the target point when calculating an point specific estimate. The measure of distance between observations has a natural geographic interpretation in spatial modeling. Given a simple linear regression function, $y_i = \beta' x_i + \mu_i$, for $i = 1, \cdots, n$. The LWR estimate for observation i is obtained simply by Weighted Least Squares (McMillen and Redfearn, 2010):

$$\hat{\beta}_i=\left(\sum_{j=1}^{n}w_{ij}x_jx_j'\right)^{-1}\left(\sum_{j=1}^{n}w_{ij}x_jy_j\right),$$

where $w_{ij} = K\left(\frac{d_{ij}}{b}\right)$ with $K(\cdot)$ being the Gaussian kernel (column normalized, as described below), b being the bandwidth parameter, and d_{ij} being the geographic distance between observations i and j.

The LWR approach is readily extended to Maximum-Likelihood Estimation (MLE) methods as well. While a typical MLE procedure chooses estimates to maximize the log-likelihood function, $\sum_{i=1}^{n} lnL_i$, the locally weighted version of MLE estimate a pseudo log-likihood function, $\sum_{i=1}^{n} w_{ij} lnL_{ij}$, where the log-likelihood function depends on the functional form of the regression model (See McMillen and McDonald, 2004, for more details).

3.4 Incorporating Locally Weighted (LW) estimation Into the IPT Framework

We incorporate LW estimation into the IPT estimator from Graham et al, in the following way. We modify eq (2) by incorporating kernel weights (in McMillen and McDonald, 2004) and a bandwidth parameter. Alternatively, we denote $\tau(w_{ij}x_i) = [1, w_{ij}x_i]$, a column vector where the weight $w_{ij} = K(d_{ij}/b)$ with $K(\cdot)$ being the Gaussian kernel (column normalized, as described below), b being the bandwidth parameter, and d_{ij} being the geographic distance between observations i and j. This setup amounts to a nonparametric specification of the tilting parameter, j , as defined below.

We describe the moment generating functions for the treated and non-treated samples, and then we discuss how one would compute the tilting parameters. Our IPT-LW discussion below closely parallels parts of the IPT approach of Graham, Pinto and Engel (2012).

Suppose G is the Logit functional form, that is, $G(v) = exp(v)/[1+exp(v)]$, and $\phi_1 = 1/G(v)$. First, $\hat{\delta}_j^{IPT,1}$ is a solution to:

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{D_i}{G\left(\tau(w_{ij}x_i)'\hat{\delta}_j^{IPT,1}\right)} - 1\right\}\tau(w_{ij}x_i) = 0.$$

Following the logic of Graham et al, the propensity score for the treated sample can be written as:

$$\hat{\pi}_{1i} = \frac{1}{N}\frac{1}{G\left(\tau(w_{ij}x_i)'\hat{\delta}_j^{IPT,1}\right)}.$$

These two equations imply:

$$\sum_{i=N_0+1}^{N_1}\hat{\pi}\tau(w_{ij}x_i) = \frac{1}{N}\sum_{i=1}^{N}\tau(w_{ij}x_i).$$

In words, this equation states that after twice reweighing the mean of x_i across treated units – once with the propensity score parameter and again with the geographic distance weights – this equals the (geographically weighted) mean of x_i over the entire sample. A similar set of equations imply an analogous relationship for the untreated sample and the entire sample. Note that higher order moments can be included in $\tau(\cdot)$, however this can complicate the computational procedure.

In terms of computation of $\hat{\delta}_j$, for each target observation j , we solve the following optimization problem, adapted from eq (A.22) of Graham et al (2012), to incorporate spatial heterogeneity:

Choose δ_j to $\max l(\delta_j) = (1/N)\sum_i D_i w_{ij} \phi(\tau(w_{ij}x_i)'\delta_j) – (1/N)\sum_i [1, w_{ij}x_i]'\delta_j$

Substituting $\tau(w_{ij}x_i) = [1, w_{ij}x_i]$ as defined above, this leads to the following revised optimization problem:

Choose δ_j to $\max l(\delta_j) = (1/N)\sum_i D_i \phi([1, w_{ij}x_i]'\delta_j) – (1/N)\sum_i w_{ij} t(x_i)'\delta_j$

The first order condition for this optimization problem is:

$\partial(l(\delta_j))/\partial\delta_j = (1/N)\sum_i D_i [1, w_{ij}x_i]'\phi'(\cdot) – (1/N)\sum_i [1, w_{ij}x_i]' = 0,$

and the second order condition is:

$\partial^2(l(\delta_j))/(\partial\delta_j)^2 = (1/N)\sum_i D_i [1, w_{ij}x_i]'[1, w_{ij}x_i]\phi''(\cdot)$

Graham et al show that $\phi'' < 0$ (see eq A.21), so that l is strictly concave.

It should be reasonably straightforward to solve the optimization problem above (analogous to Graham et al's equation A.22) for $\hat{\delta}_j$ for all j. A major difference between our IPT-LW approach and the IPT approach is that the IPT-LW estimator will lead to separate parameter estimates of $\hat{\delta}_j$, for each of the N observations in our sample. These $\hat{\delta}_j$ are what we would call our "IPT-LW" estimator. In contrast, the IPT estimator leads to one estimate of $\hat{\delta}_j$, for all j.

Then we can get the ATE for each observation. In footnote 21 of the Appendix of Graham et al foe, they describe the process for obtaining the overall ATE that is based on the single treatment effect for each observation. Our approach to obtaining the ATE for each observation is similar to the overall ATE generation process outlined by Graham et al, but we modify the moments condition using $\tau(w_{ij}x_i)$ instead of $t(x)$. With IPT-LW, we obtain a very representative estimate of the local treatment effect by generating an ATE for each target point, rather than generating one treatment effect for each target effect and using these to calculate one overall ATE. We would expect that generation of these multiple treatment effects for each observation would lead to a more precise estimate of the ATE at each location, and in turn, the overall average of the ATE may have lower bias.

We next perform Monte Carlo simulations to determine how well this IPT-LW estimator performs in small samples. Subsequently, we estimate an empirical application for the Vancouver, BC, Canada data set, including results from our IPT-LW estimation procedure, and for comparison purposes also estimate the application for OLS and the IPT estimator.

4 Monte Carlo Study

Often, estimating treatment effects is the key for evaluating treatments, but the interventions of interest are not usually assigned at random. In this Monte Carlo study we generate our response variables, y_{it}, from the following causal model and selection model:

$$y_i = \beta_0(g_i) + D_i \cdot \beta_1(g_i) + x \cdot \beta_2(g_i) + u_i, \tag{3}$$

$$D_i = \begin{cases} 1 \quad for \quad \sqrt{(g_i^1)^2 + (g_i^2)^2} < 0.5 \\ 0 \quad for \quad \sqrt{(g_i^1)^2 + (g_i^2)^2} \geq 0.5 \end{cases}, \ i = 1, \cdots, N \tag{4}$$

where (3) is the causal model that produces the response variable y_i and (4) is the selection model that produces the treatment status D_i (1 indicates that the unit is treated and 0 indicates being not-treated); $g_i = g_i^1, g_i^2$ is a two-dimentional location vector generated from a bi-variate standard normal distribution truncated between $[0, 2]$, u_i is i.i.d. following a standard normal distribution; x_i is a random variable generated from the normal distribution $N[0, 3]$, and v_i is i.i.d from the standard normal distribution; Additionally, we set $\beta_0(g_i)$ and $\beta_2(g_i)$ as constant, and $\beta_1(g_i)$, the main interest in the estimation, is a bivariate standard normal density function:

$$\beta_1(g) = \frac{1}{2\pi} \exp\left(\frac{g_i^1)^2+(g_i^2)^2}{2}\right).$$

We use $N = 600$ as the number of individuals in each period. This model is estimated with OLS, IPT and IPT-LW estimator defined in section 2. The average bias and average squared errors are reported in the following table (**??**). The IPT-LW estimates have a smaller bias as well as that of the IPT and the OLS estimates.

5 Application: Commercial Real Estate Prices in the Vancouver, BC Metro Area

The impacts on property values of proximity to the city center have long been studied in the urban economics literature (Alonso, 1964; Muth, 1969; among others). This "monocentric model" implies that property values rise as businesses and individuals move closer to the city center, and they tend to fall off as one moves to the outskirts of the metropolitan area. This is closely related to the concept of agglomeration economies, which can arise due to labor market pooling where having substantial numbers of skilled employees can facilitate recruitment of workers and in turn, lower the operating costs of businesses. Since agglomeration economies can be beneficial to the bottom line of firms, the desire for firms to locate in the city limits can lead to higher density as well as increased commercial property values per square foot. In empirical applications with a cross-sectional context, it is important to consider the identification of the ATE from locating on one side of the city border opposed to outside the city limits. In this "application", we examine commercial real estate that sold in the metro-Vancouver, BC area in the year 2013. The "treatment" sample of properties is those that sold within the city-limits, and the "control" sample is those properties that sold outside the city-limits. We are most interested in the set of properties that sold just outside the border of the City of Vancouver, to generate an ATE of developing right on the other side of the city limits relative to just outside the border. We use the IPT-LW estimation procedure to demonstrate the estimation of the ATE for properties that are within the City of Vancouver city-limits. Figure 1 shows the locations of our sample of 103 commercial properties (for which we have usable data) that sold (as arms-length transactions) in the entire metro-Vancouver area in the year 2013. In addition to in the City of Vancouver, some of these properties sold in Coquitlam, Port Coquitlam, Burnaby, Surrey, and Richmond, BC. While there are more than 103 commercial properties that sold in 2013 in these cities in BC, there are 103 properties for which our database contains information about building area square footage and "effective year" (i.e., the most recent year of major renovations, which indicates the quality of the property). These data are from the BC Assessment database, which we purchased from Landcor. Descriptive statistics are presented in Table 2. The average commercial property sold for approximately C$ 3.72 million, had 23,416 square feet (for an average price per square foot of C$ 135), was built in 1966 and last renovated in 1978. The lowest price property sold for C$ 585,000 and the highest sold for approximately C$ 44 million. Approximately 50% of the properties in this sample were sold in the City of Vancouver and the remaining 50% were sold in the other cities described above. We first estimate the following OLS model: $Y_i = b_0 + b_1 X_i + b_2 D_1 + e$, where Y_i is price per square foot for property i, X_i is the effective year (which indicates the most recent year of major renovations), and e is an i.i.d. error term with mean 0 and constant variance, and $E(e_i e_j) = 0$ for $i \neq j$. $D_i = 1$ for properties that sold in 2013 in the City of Vancouver, and $D_i = 0$ for properties that sold in 2013 outside of the City of Vancouver. The regression coefficient b2 is the "treatment effect" of locating in the City of Vancouver. The second model we estimate is the IPT model, where $t(x) = [1, X]$, and X is the effective year and Y is the sale price per square foot. In this context, we are reweighing the X's so that the sample mean of X in the treated sub-sample equals the entire sample mean of X. Once again, the data set we use is the 2013 sales in the Vancouver metro area; $D_1 = 1$ for properties that sold in the City of Vancouver, and $D_1 = 0$ otherwise. We then calculate the ATE based on the IPT estimator. Finally, we estimate an IPT-LW model, with Gaussian kernel weights given as $w_{ij} = exp(-0.5 * (d_{ij}/b)^2)$, where d_{ij} is the Euclidean distance between properties i and j, and b is a bandwidth parameter. Note that we have initially estimated the bandwidth to be slightly higher than the maximum of the standard deviations among all of the dij in the sample. In the IPT-LW model, we use $(w_{ij}X) = [1, w_{ij}X]$. In this context, we are reweighing the distance-weighted averages of X so that the reweighed distance-weighted mean of the X's for the treated sample equals the distance-weighted mean for the entire sample. We present the results of the OLS and the IPT estimations in Table 3. First, the treatment dummy, D_1, has a coefficient estimate of $b_2 = 493$, implying that the typical commercial property in the City of Vancouver sold for approximately C$ 493 more per square foot than the typical property outside of the city-limits. Also, b_2 is highly significant (P-value=0.000). With IPT, the ATE is estimated to be C$ 496 (with P-Value=0.000), and with a standard error that is slightly lower than the corresponding OLS standard error (C$ 98 opposed to C$ 102). This smaller standard error for the IPT estimator is in line with our expectations. We next estimate the ATE using the IPT-LW estimator that we have developed in this paper. Figure 1 shows a map of the metro-Vancouver area with the locations of the sample of commercial properties that sold in 2013. The range of the ATE for the 103 observations is C$ 220 to approximately C$ 15,000 , but the latter ATE has a very large standard error and is statistically insignificant (t-stat=0.34, d.f.=100, P-Value=0.735). Due to this implausibly large but statistically insignificant positive ATE outlier, we drop this ATE observation, and take the average of all of the remaining 102 ATE's (which we denote as the "AATE"). In this context, the AATE equals approximately C$ 472, while the average of the standard errors is C$ 125. In general, the properties with the highest ATE's are located in the downtown Vancouver area, while for the most part those with lower ATE's are further to the east and south. As another way of presenting the locations of the properties in the IPT-GWR results, we generate a "heat map" of the individual properties, as shown in Figure 2. It is apparent that there are bright red clusters of properties in the downtown areas of the City of Vancouver, implying the strongest potential benefits from nearby clusters of properties in those urban locations. Also, those properties tend to have the highest ATE's among all properties in the metro area. As one moves away from these centers, the intensity of the clustering diminishes. There are some pockets of somewhat strong clustering in other parts of the metro area, however, none of these are nearly as strong as in the central business district in the City of Vancouver.

In Figure 3 we present the locations of the higest, or those properties with ATE above C$ 470

, (in green), and the lowest, the ATE below C$ 220 (in orange). The points of greatest interest are those that are just outside of the border of the City of Vancouver. These include those locations that are from north to south along Route 1 and continuing south immediately below that, all the way down to the Fraser River, as well as those below the Fraser River from east to west in the City of Richmond. All of these properties are in the top half of ATE estimates, implying substantial benefits to developers from choosing to develop on the Vancouver side of the border.

6 Conclusion/Discussion

There are some benefits, as well as drawbacks, of our IPT-LW approach. One advantage of IPT-LW is that we are able to generate different ATE estimates across locations (opposed to one treatment effect estimate for each observation with IPT), and the IPT-LR ATE's are dependent on the geographic location of the observations relative to each target point. The average of the ATE's, or the AATE, is a way of summarizing this information over all observations. In our application, one may be particularly interested in the ATE estimates along the border of the City of Vancouver limits, which can help developers decide whether to develop on the Vancouver side of the border or on the suburbs side of the border (if the ATE on the suburbs side of the border is negative). Comparing the magnitudes of the ATE estimates across various border locations can also help developers choose inside of which part of the border would be the best to develop.

Another advantage of IPT-LW, as demonstrated by our Monte Carlo simulations, is that the bias of the IPT-LW estimator tends to be lower than the bias for the OLS and IPT estimators. However, at least in our preliminary simulations analysis the average squared error from IPT-LW tends to be higher than that of IPT. Also, IPT-LW is more computationally intensive and in some cases this can diminish its feasibility, especially in very large samples. Clearly, there are advantages and disadvantages of both the IPT and IPT-LW approaches to addressing the missing data problem in generating estimates of ATE's. In future work, we aim to further refine the Monte Carlo simulations by using alternative functional forms for the “true” ATE. It is possible that the lower bias for IPT-ATE bias and/or the lower average squared error results for IPT may depend on the relationship between the true ATE and the location of the individual observations. Also, we plan to run some simulations using multiple repetitions of the data generating process, to confirm that our Monte Carlo results are not merely a coincidence for the particular sample that we have generated.

References

[1] Abadie, Alberto, David Drukker, Jane Leber Herr and Guido W. Imbens. (2004). ìImplementing Matching Estimators for Average Treatment Effects in Stata. The Stata Journal 4 (3): 290 - 311.

[2] Alonso, W. (1964). Location and land use. Toward a general theory of land rent.

[3] Graham, B.S., C. Campos De Xavier Pinto, D. Egel. (2012). “Inverse Probability Tilting for Moment Condition Models with Missing Data,” Review of Economic Studies, 79: 1053-1079.

[4] Graham, Bryan S., Cristine Campos de Xavier Pinto, and Daniel Egel. (2011). "Inverse Probability Tilting Estimation of Average Treatment Effects in Stata." The Stata Journal, pp. 1-16.

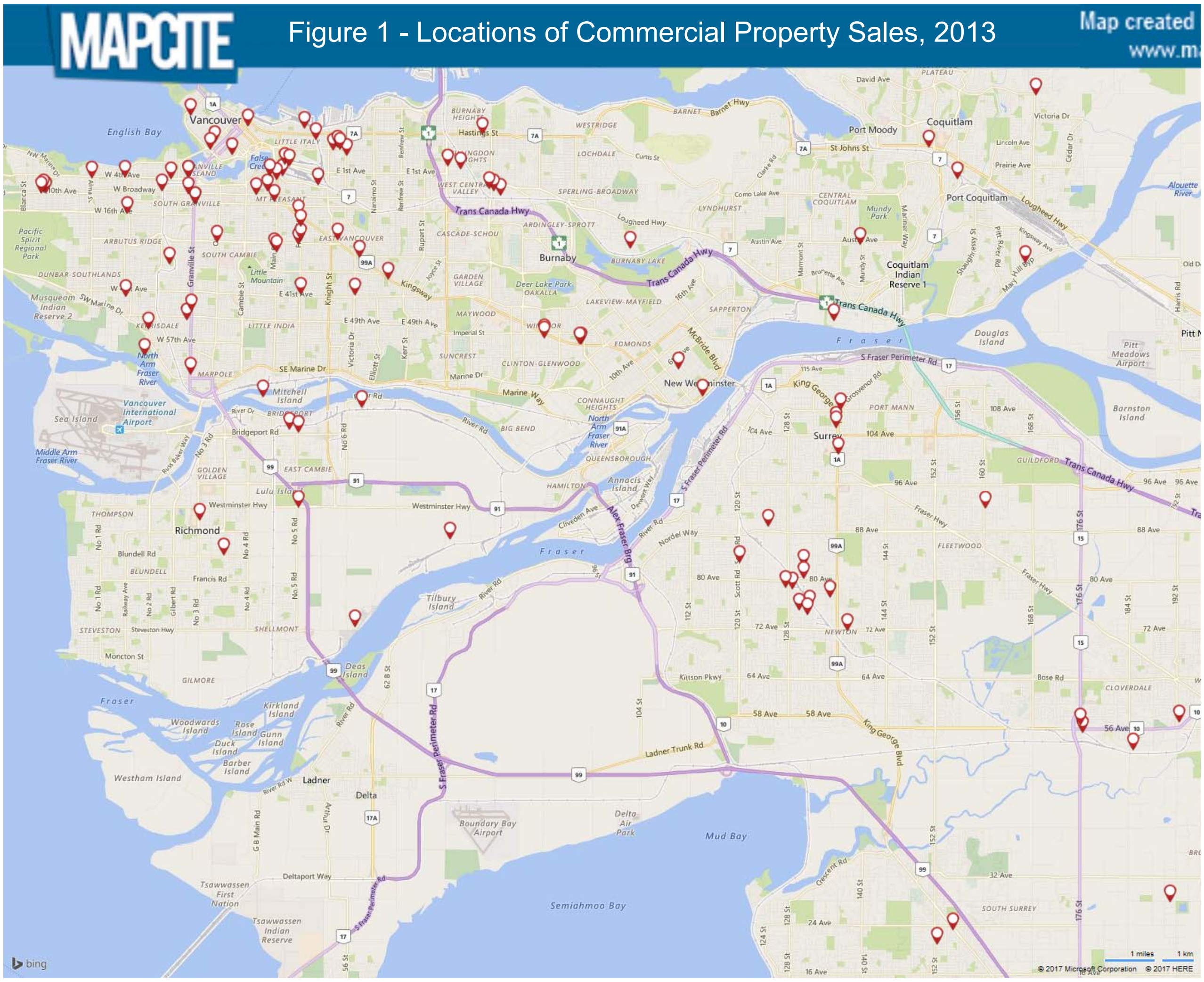
[5] Imbens, G.W. (2004). “Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review,” Review of Economics and Statistics, 86: 4-29.

[6] McMillen, D.P. and J. F. McDonald (2004). “Locally Weighted Maximum Likelihood Estimation: Monte Carlo Evidence and an Application,” in Luc Anselin, Raymond J.G.M. Florax, and Sergio J. Rey (eds.), Advances in Spatial Econometrics. New York: Springer, 225- 239.

[7] McMillen, D.P. and C. Redfearn (2010). “Estimation and Hypothesis Testing for Nonparametric Hedonic House Price Functions,” Journal of Regional Science 50, 712-733.

[8] Muth, R. F. (1969). Cities and Housing; The Spatial Pattern of Urban Residential Land Use.

A Figures


MAPCITE
Figure 1 - Locations of Commercial Property Sales, 2013
Map created
Vancouver
Burnaby
Coquitlam
Port Moody
Port Coquitlam
New Westminster
Surrey
Richmond
Ladner
Delta
Vancouver International Airport
Boundary Bay Airport
Mud Bay
Semiahmoo Bay
English Bay
Trans Canada Hwy
bing

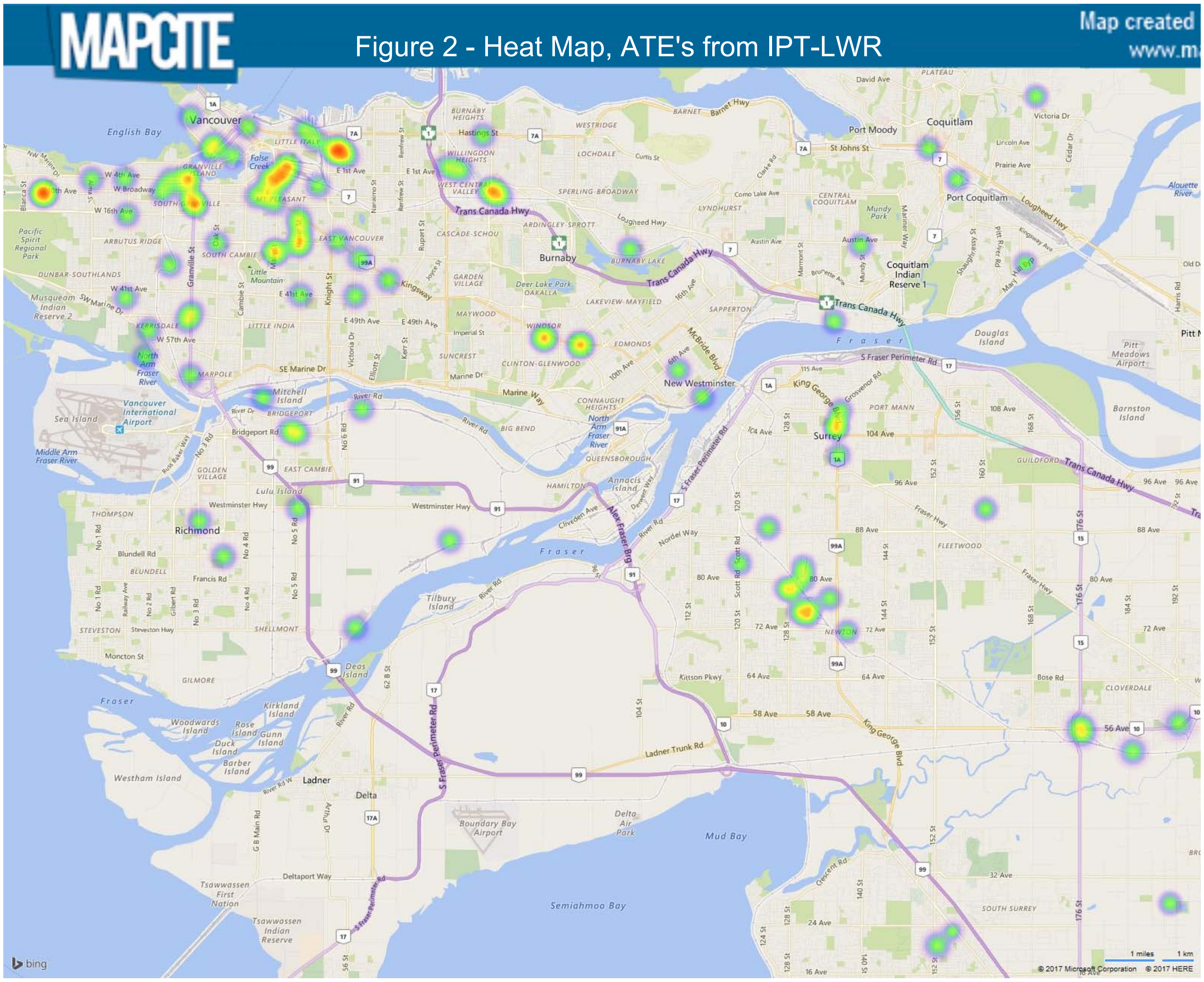


Figure 2 - Heat Map, ATE's from IPT-LWR

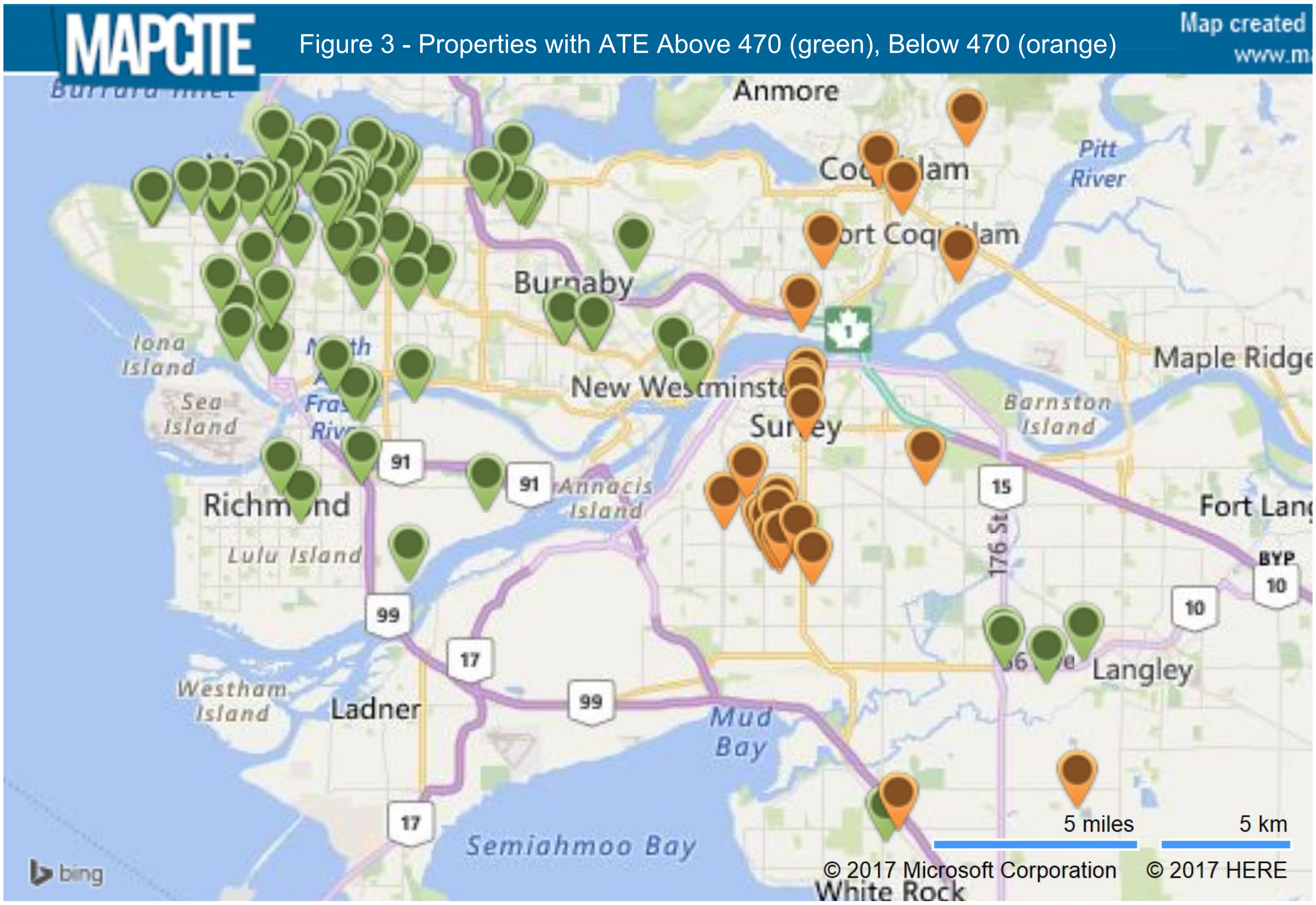


Figure 3 - Properties with ATE Above 470 (green), Below 470 (orange)

B Tables

Table 1: Simulations: IPT & IPT-LW Estimation on Average Treatment Effect

		OLS			IPT			IPT-LWR	
True ATE		**Bias**	**ASE**		**Bias**	**ASE**		**Bias**	**ASE**
$\beta_1(g) = \frac{1}{2\pi}\exp\left(\frac{(g_i^1)^2+(g_i^2)^2}{2}\right)$		-0.3001	0.1189		-0.3014	0.1196		-0.2778	0.1172

Table 2 - Descriptive Statistics
y is sale price per square foot of building area (C$)

Variable:	y	EFF_YEAR	DUM_TREAT2013	YEAR_BUILT
Mean	525.3584	1977.689	0.504854	1965.951
Median	366.0322	1978.000	1.000000	1971.000
Maximum	2980.448	2014.000	1.000000	2014.000
Minimum	43.27273	1925.000	0.000000	1901.000
Std. Dev.	544.6467	18.59318	0.502421	28.28371
Observations	103	103	103	103

Table 3 – Ordinary Least Squares (OLS) Regression Results
y is sale price per square foot of building area (C$)
gamma is the Average Treatment Effect (ATE) parameter

Source	SS	df	MS
Model	6033286.99	2	3016643.49
Residual	24223995.5	100	242239.955
Total	30257282.5	102	296640.024

Number of obs	=	103
F(2, 100)	=	12.45
Prob > F	=	0.0000
R-squared	=	0.1994
Adj R-squared	=	0.1834
Root MSE	=	492.18

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
eff_year	1.417224	2.749469	0.52	0.607	-4.037645	6.872092
gamma	497.3257	101.75	4.89	0.000	295.4566	699.1948
_cons	-2528.547	5453.549	-0.46	0.644	-13348.23	8291.139

Table 4 – Inverse Probability Tilting (IPT) Estimation Results
y is sale price per square foot of building area (C$)
gamma is the Average Treatment Effect (ATE) parameter

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
delta1						
eff_year	-.0323186	.0119886	-2.70	0.007	-.0558157	-.0088214
_cons	63.93792	23.70628	2.70	0.007	17.47448	110.4014
delta0						
eff_year	-.0439811	.0160761	-2.74	0.006	-.0754897	-.0124724
_cons	86.99598	31.77389	2.74	0.006	24.7203	149.2717
ate						
gamma	496.1368	97.78329	5.07	0.000	304.485	687.7885

Total number of observations: 103

Table 5 - Inverse Probability Tilting Locally Weighted Regressions (IPT-LWR) Results
y is sale price per square foot of building area (C$)
gamma is the Average Treatment Effect (ATE) parameter

	Mean	Std. Dev	Min	Max
ate				
gamma	472	38	220	596
std. error	125	34	88	329

Total number of observations: 102
Note: we dropped one outlier ate observation that was statistically insignificant