INVERSE PROBABILITY TILTING WITH SPATIAL DATA: AN APPLICATION OF HOW A SEVERE STORM IMPACTS COMMERCIAL REAL ESTATE PRICES

JEFFREY P. COHEN AND KE YANG

Abstract. We develop a nonparametric, geographically-weighted Inverse Probability Tilting (GIPT) estimator, which allows for average treatment effect (ATE) heterogeneity across geographic space, and addresses "missing data" problems when data vary geographically. GIPT re-weights twice: using propensity scores that equate moments across treated (and untreated) sub-samples with the entire sample, as in Graham et al (2012); and also, down-weighting observations far from each target point. This allows for heterogeneous ATE estimates. Monte Carlo simulations validate the strong small sample performance of GIPT. Among many possible applications of this GIPT estimator, we demonstrate how a severe storm leading to an extended water-boil advisory, imposed much longer on some sub-sections of Metro-Vancouver Canada, impacted individual commercial property price ATEs differently.

Keywords: Inverse probability tilting; missing data; storm impacts on real estate; spatial dependence.

JEL Codes: C31, R3

Date: 4/17/2018.
Jeffrey P. Cohen is an associate professor at University of Connecticut, School of Business, Jeffrey.Cohen@business.uconn.edu; Ke Yang is an associate professor at University of Hartford, Barney School of Business, Kyang@hartford.edu. We acknowledge the helpful comments on earlier versions of this paper from seminar participants at SUNY-Albany, University of Missouri-Columbia, Rutgers University, and University of Cincinnati. We also appreciate feedback from participants at the 2018 Camp Econometrics. Any errors are our own.

1. Introduction

Two increasingly popular areas of focus in recent applied statistical research are average treatment effect heterogeneity, and missing data problems. We propose a methodology to address both of these issues in one framework, when there are data available on the geographic locations of observations.

One set of approaches to missing data problems in general settings is propensity score approaches. There is an extensive body of literature on Inverse Probability Weighting (IPW), as in Rosenbaum and Rubin (1983) and followed by Imbens (2004) and Wooldridge (2007). More recently, Graham et al (2012) developed a methods-of-moments based approach called Inverse Probability Tilting (IPT).

Some recent research has focused on specific types of missing data problems and some have addressed them with methods-of-moments approaches. For instance, Abrevaya and Donald (forthcoming) consider a situation where some observations on an explanatory variable are missing, and they develop a methods-of-moments estimator to handle this problem.

One objective of this paper is to consider a second adjustment for missing data problems as a part of the estimation strategy – specifically, re-weighting that allows for geographic heterogeneity in a cross sectional context, in addition to a propensity score approach for the missing data problem. The attractive features of IPT that we describe below have prompted us to explore a generalization of IPT. This type of additional spatial adjustment is important in the context of many treatment effect problems, because the ATE at one geographic location can be different than the ATEs at other locations.

In particular, the issue of ATE heterogeneity has received some recent attention. While one advantage of IPT is that it leads to a unique treatment effect for each observation, it may also be desirable to consider spatial heterogeneity in a nonparametric framework that could lead to different ATEs across each individual observation. Bitler et al (forthcoming) demonstrate that using constant mean-impacts in the treated versus untreated subgroups ignores much of the heterogeneity in these two subgroups. In such situations, an approach to deal with the missing data problem while preserving heterogeneity in ATEs across geographic locations is desirable. Thus, a second objective of our research is to demonstrate how a general version of the IPT approach that considers geographic variation in the data can address the missing data problem while at the same time allowing for heterogeneity in the ATEs across geographic space to be brought out.

We demonstrate the use of our estimator with one particular application of commercial property sales, where a treatment is imposed on some properties in a geographic region, but neither on others in the same region nor upon any properties in a neighboring region. With this particular missing data problem, the researcher knows what price a given property sold for at one location, but does not know how much the same property would have sold for if it had been in a different location. While this is the specific application that we consider in this paper, there are many other potential applications of GIPT, in contexts where there is geographical variation in the data and a treatment that is imposed on units in some locations after a random event, but not in others.

In the remainder of this paper, we first motivate one type of missing data problem (although our estimator can be applicable to a broad range of other missing data problems). Next we explain our innovations to the IPT estimator that incorporates geographic heterogeneity, and the adjustments to the propensity score weights we make to allow for more distant observations to be down-weighted relative to more close observations. We call this an GIPT estimator (representing "Geographically-weighted Inverse Probability Tilting"). We describe the computation process of the GIPT estimator, then provide some Monte Carlo evidence to demonstrate that our estimator performs well. We demonstrate the use of this GIPT estimator with one specific application of how commercial property prices in the metro-Vancouver, BC Canada region may be impacted differently, shortly before versus after a storm leading to an extended water-boil advisory that is imposed on some parts of the region for much longer than other areas. Finally, we discuss potential future extensions to our approach and summarize our findings.

2. Motivation

Consider the following general problem. First, suppose one is interested in analyzing a data set on units that are in various locations throughout a particular geographic region, to determine the ATEs at each location in the region shortly after versus shortly before a random "event" . If the treatment area is confined to the borders of a particular city in a metro area, for instance, we might consider focusing attention on "treated" observations as a set of those that are on the "inside" of the city limits after the "event" . The "untreated" observations are a set of units that are on the "outside" of the city limits, e.g., those observations in a neighboring city, before and after the "event" as well as those within the city limits before the "event" . Then, we can estimate the effect of being in the treatment sub-sample opposed to the non-treated sub-sample, assuming that there are no missing data.

But we know for many empirical applications that the treatment is observed contingent on the location of the observations. In other words, it is not known what the treatment outcome would have been if a particular unit had been located elsewhere. In these cases, in order to obtain valid treatment effects, one can re-weight the data with propensity scores. There are several approaches to accomplishing this. One is an IPW approach, which has received extensive attention in the literature (see, e.g., Rosenbaum and Rubin (1983), and Imbens (2004), among others). IPW uses Maximum Likelihood estimation techniques to obtain the propensity score weighting parameters. An attractive alternative is the IPT approach, as in Graham et al (2012), which is based on a relatively straightforward moments condition technique. An advantage of IPT is that it generates separate tilting parameters for the treated and untreated samples. There are alternative missing data approaches that have been proposed by others, such as random forests (Wager and Athey, 2017), and some methods closely related to IPT (e.g., Imai and Ratkovic, 2014, and Hainmueller, 2012), among others. A comparison of many of these approaches is presented in Frölich et al (2017), however there is no known analysis of GIPT in the literature. Incorporating the spatial variation into the missing data literature is one main contribution of our paper.

Specifically, the IPT and IPW approaches do not allow for geographic heterogeneity in the ATEs and the tilting parameters. If the geographic locations of observations are varied, this could be an important consideration in many particular applications. While it may be possible to include geographic coordinates directly as control variables in the IPT estimation, this would violate the "strong overlap" assumption of Graham et al (2012). Therefore, as an attractive alternative to including geographic coordinates directly as IPT control variables, it may be helpful to re-weight a second time, to consider the geographic distance between observations. This is common in the non-parametric estimation literature, specifically, with an approach called Locally Weighted Regressions (LWR), also commonly referred to as Geographically Weighted Regressions (GWR), as in Brunsdon et al (1996). McMillen and Redfearn (2010) describe LWR and present an application. However, no known work has incorporated geographically weighted estimation into an IPT framework.

3. Approach

3.1. **Model.**

Suppose that there are 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 covariates, pretreatment variables or exogenous variables, Z_i, and vector of covariates for its geographic location, L_i. Also let $X_i = \{Z_i, L_i\}$. Each unit is exposed to a single treatment; $D_i = 0$ if unit i is untreated 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:

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

Distributions of (D_i, Y_i, X_i) refer to the distribution induced by the random sampling from the population. We follow the potential outcomes of Neyman (1923) and Rubin (1974) , assuming the existence of potential outcomes, $Y(1)$ and $Y(0)$, corresponding respectively to the outcome the subject at a specific location would have experienced with or without treatment. Then we can define the average treatment effect (ATE) at l as

$$\gamma(l) = \mathbb{E}[Y(1) - Y(0)|L = l].$$

In practice, however, one only observes

$$Y_i = (1 - D_i)Y_i(0) + D_iY_i(1)$$

i.e., only $Y_i(1)$ for actively treated units or $Y_i(0)$ for untreated units are observed at any given location. One common assumption needed for estimation in this case is the following:

Assumption 1. *(Unconfoundedness)* $\{Y(1), Y(0)\} \perp D|X$.

This assumption effectively implies that we can treat nearby observations in each geographic location as having come from a randomized experiment. It follows immediately that the ATE at location l, $\gamma(l)$, can be identified as

$$\gamma(L = l) = E\left[E\left[Y|D = 1, X\right] - E\left[Y|D = 0, X\right]|L = l\right]$$

or equivalently

$$\gamma(L = l) = E\left[\frac{DY}{p(X)} - \frac{(1 - D)Y}{1 - p(X)}|L = l\right] \tag{3.1}$$

where $p(X) = P\left[D = 1|X = x\right] = E[D_i|X_i = x]$ is the propensity score function that prescribes the conditional probability of receiving treatment at x (which is a generalization of the setup in Rosenbaum and Rubin, 1983). As this propensity score function is generally unknown, many earlier methods on average treatment effect estimation differ in how they estimate $p(X)$ using, e.g., variants of maximum likelihood approaches, such as the Inverse Probability Weighting (IPW) estimator that we describe in the next section, and then plug $p(X)$ in equation (3.1) to calculate an average treatment effect.

3.2. **Geographically Weighted Inverse Probability Tilting Estimator (GIPT).** Rosenbaum and Rubin (1983) proposed the Inverse Probability Weighting ATE estimator by first replacing the $p(X)$ with a maximim likelihood estimator, then averaging over sample points. The Rosenbaum and Rubin (1983) setup implicitly assumes no variation in l across observations. Graham et al (2012) proposed an alternative method by estimating the propensity score function with a particular method of moments estimator consisting of two separate tilting parameters - two sets of propensity scores - one set for each observation in the treatment group and another for observations in the control group. Our method of estimating the geographically specific average treatment effects is based on a generalization of the IPT estimator proposed by Graham et al (2012) and it requires the following assumptions 2 through 8 below, in addition to assumption 1 above (the unconfoundedness assumption).

Assumption 2. *(Random Sampling). $\{D_i, X_i, Y_{1i}\}_{i=1}^N$ is an independently and identically distributed random sequence. We observe D, X, and $Y = DY_1$ for each sampled unit.*

Assumption 3. *(Identification) For some known $K \times 1$ vector of functions $\Phi(Y, X, \gamma)$,*

$$E(\Phi(Y, X, \gamma)) = 0$$

with (i) $E(\Phi(Y, X, \gamma)) \neq 0$ for all $\gamma \neq \gamma_0$, $\gamma \in \Theta \subset \mathbb{R}^K$, and Θ compact with $\gamma_0 \in int(\Theta)$,(ii) $|\Phi(Y, X, \gamma)| \leq c(Y, X)$ for all Y, X with $c(\cdot)$ a non-negative function and $\mathbb{E}(c(Y, X)) < \infty$, (iii) $\Phi(Y, X, \gamma)$ is continuous on Θ for each Y, X and continuously differentiable in a neighborhood of γ_0, (iv) $\mathbb{E}[\|\Phi(Y, X, \gamma)\|^2] < \infty$, and (v) $\mathbb{E}[sup_{\gamma \in \Theta} \|\nabla_\gamma \Phi(Y, X, \gamma)\|] < \infty$.

Assumption 4. *$p(X) = P[D = 1|X = x]$ is bounded away from 0 and 1 over $\aleph$, the support of X.*

Assumption 5. *There is a continuous function $\delta_0(\cdot)$ and compact, know vector $r(X)$ of linearly independent functions of X, and known function $G(\cdot)$such that (i) $G(\cdot)$ is strictly increasing, continuously differentiable, and maps into the unit interval with $\lim_{\nu \to -\infty} G(\nu) = 0$ and*

$\lim_{\nu\to\infty} G(\nu) = 1$, *(ii)* $p(x) = G\left(r(z)'\delta_0(l)\right)$ *for all* $x \in \aleph$, *and (iii)* $G(r(z)'\delta_0(l))$ *is bounded away from 0 and 1 for* $\delta_0(\cdot)$ *and* $x \in \aleph$.

The geographically weighted regressions (GWR) approach is a commonly used non-parametric estimation procedure in spatial studies to allow for geographic heterogeneity in regression parameters over space. In other words, this approach leads to the possibility of different marginal effects at each target point. The basic idea behind GWR is to assign higher weights to observations near the target point when calculating a point specific estimate. The measure of distance between observations has a natural geographic interpretation in spatial modeling. The GWR 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, the geographically weighted version of MLE estimates a pseudo log-likelihood function, where the log-likelihood function depends on the functional form of the regression model. See McMillen and McDonald (2004), for more details.

We incorporate Geographical weights into the IPT estimator from Graham et al (2012), in the following way. We modify equation (A.22) by incorporating kernel weights and a bandwidth parameter. If the researcher believes that the potential outcome function $G(\cdot)$ is a non-parametric function, then we could transform both $t(\cdot)$ and D_i with some kernel weights. More specifically, suppose one is interested in the first two moments (although one can include additional moments). Then, we denote $\tau(w_i(l)x_i) = [1, w_i(l)x_i, (w_i(l)x_i)^2, \cdots (w_i(l)x_i)^m,]'$, as a column vector where the weight $w_i(l) = \left[K\left(\frac{d_i(l)}{b}\right)\right]^{1/2}$, with $K(\cdot)$ being the Gaussian kernel, b being the bandwidth parameter, m is the number of moments included, and $d_i(l)$ being the geographic distance between observations i and location $L = l$. This setup amounts to a non-parametric specification of the tilting parameters, $\delta^0(l)$ and $\delta^1(l)$, as defined below.

Specifically, suppose, for computational simplicity, one allows G to take the Logit functional form, that is, $G(v) = exp(v)/[1 + exp(v)]$, and $\phi_v = 1/G(v)$. In terms of computation of $\tilde{\delta}^h(l)$, $h = 0, 1, ..., H$, for each target observation, where H is the number of treatment groups and 0 represents the control group, the GIPT estimator solves the following optimization problem, adapted from equation (A.22) of Graham et al (2012), to incorporate spatial heterogeneity across target points:

Choose $\delta^h(l)$ to $\max \mathtt{L}(\delta^h(l)) = (1/N)\sum_i D_i^h w(l)\phi^h(\tau(w(l)x_i)'\delta^h(l)) - (1/N)\sum_i \tau(w(l)x_i)'\delta^h(l)$

where D_i^h is the treatment dummy for group h and ϕ^h are specific to group h. If there is one treatment and a control group, then $h = 0, 1$, and the notation for these dummies would reduce to $(1 - D)$ and D, respectively..

The first order condition for this optimization problem is:

$$\partial(\mathtt{L}(\delta^h(l)))/\partial\delta^h(l) = (1/N)\textstyle\sum_i D_i^h w(l)\tau(w(l)x_i)'\phi_\delta^h(\cdot) - (1/N)\displaystyle\sum_i \tau(w(l)x_i)' = 0,$$

and the second order condition is:

$$\partial^2(\mathtt{L}(\delta^h(l)))/(\partial\delta^h(l))^2 = (1/N)\sum_i D_i^h w(l)\tau(w(l)x_i)''\phi_{\delta\delta}^h(\cdot)$$

Graham et al (2012) show for the special case where there is no geographic heterogeneity, that $\phi_{\delta\delta}^h(\cdot) < 0$ (see their equation A.21), so that $(\mathtt{L})$ is strictly concave.

When $h = 0, 1$, it is reasonably straightforward to solve the optimization problem above (analogous to equation A.22 in Graham et al, 2012) for $\tilde{\delta}^h(l)$ for all l. A major difference between our GIPT approach and the IPT approach is that the GIPT estimator will lead to separate parameter estimates of $\tilde{\delta}^h(l)$, $l = 1, \cdots, N$. These $\tilde{\delta}^h(l)$ are what we call our GIPT estimator. In contrast, the IPT estimator leads to one estimate of $\tilde{\delta}^h(l)$, for all l. When there is no geographic variation in cross sectional data, the estimates from IPT and GIPT should be identical, and therefore the additional computational burden from GIPT would not yield any of the benefits that may be present with geographic data. 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 GIPT discussion below closely parallels parts of the IPT approach of Graham et al (2012). When there is one treatment group and one control group, then let N_1 and N_0 denote the number of treated units and untreated units, respectively. First, for the unit at location $L = l$ in the treatment group, the locally weighted IPT estimator of δ, denoted by

$\tilde{\delta}^1$, is a solution to:

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{w_i(l)\cdot D_i}{G\left(\tau(w_i(l)\cdot x_i)'\tilde{\delta}^1(l)\right)}-1\right\}\tau(w_i(l)\cdot x_i)=0, \tag{3.2}$$

where $G\left(\tau(w_i(l)\cdot x_i)'\tilde{\delta}^1(l)\right)=p(x)$ for all $x\in\mathbb{X}$ and some δ_1, $\tau(w_i(l)\cdot x_i$ is a $1+M$ column vector of known functions of X with a constant as its first element, and $\tilde{\delta}^1$ is a vector of maximum likelihood estimates of δ_1.Following the logic of Graham et al (2012), the propensity score for the i^{th} unit in the treated sample can be written as:

$$\tilde{\pi}_i^1(l)=\frac{1}{N}\frac{w_i(l)}{G\left(\tau(w_i(l)\cdot x_i)'\tilde{\delta}^1(l)\right)},\quad i=N_0+1,N_0+2,\cdots,N. \tag{3.3}$$

These two equations imply:

$$\sum_{i=N_0+1}^{N_1}\tilde{\pi}_i^1(l)\cdot\tau(w_i(l)\cdot x_i)=\frac{1}{N}\sum_{i=1}^{N}\tau(w_i(l)\cdot x_i). \tag{3.4}$$

Second, for the unit at location $L=l$ in the untreated group, the GIPT estimator of δ^0, denoted as $\tilde{\delta}^0(l)$, is the solution to:

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{w_i(l)\cdot(1-D_i)}{1-G\left(\tau(w_i(l)\cdot x_i)\tilde{\delta}^0(l)\right)}-1\right\}\tau(w_i(l)\cdot x_i)=0,\quad i=1,\cdots,N_0. \tag{3.5}$$

Similarly, the propensity score for the i^{th} unit in the control sample can be written as:

$$\tilde{\pi}_i^0(l)=\frac{1}{N}\frac{w_i(l)}{1-G\left(\tau(w_i(l)\cdot x_i)'\tilde{\delta}^0(l)\right)}. \tag{3.6}$$

These two equations imply:

$$\sum_{i=1}^{N_0}\tilde{\pi}_i^0(l)\cdot\tau(w_i(l)\cdot x_i)=\frac{1}{N}\sum_{i=1}^{N}\tau(w_i(l)\cdot x_i). \tag{3.7}$$

In words, equation (3.4) states that after twice reweighting the moments of x_i across treated units – once with the propensity score parameter and again with the geographic distance weights

– this equals the (geographically weighted) moments of x_i over the entire sample. An analogous relationship for the untreated sample and the entire sample is in equation (3.7). Note that higher order moments can be included in $\tau(\cdot)$, however this can complicate the computational procedure.

The geographically weighted IPT average treatment effect estimate for the unit at location $L = l$ is given by

$$\tilde{\gamma}^{GIPT}(l) = \sum_{i=N_0+1}^{N} \tilde{\pi}_i^1(l) Y_i - \sum_{i=1}^{N_0} \tilde{\pi}_i^0(l) \cdot Y_i \tag{3.8}$$

where $\tilde{\pi}_i^1(l)$ and $\tilde{\pi}_i^0(l)$ are location dependent and defined by (3.3) and (3.6).

Some discussion of the IPT assumptions, 1 through 4, in the context of GIPT estimation, is worthy of some attention. The ATE identification strategy, Assumption 1, is still relevant at each target point. In terms of random sampling (Assumption 2), for a given target point j, the observations on the $w_i(l) \cdot D_i$, $w_i(l) \cdot x_i$, and Y_{1i} are independent over all i. For data missing at random, Assumption 3, the probability of being treated, conditional on $w_i(l) \cdot x_i$, is independent of the outcome of the treated sample. More formally, this assumption translates into $\mathbb{P}(w_i(l) \cdot D_i > 0 | w_i(l) \cdot x_i, Y_1) = \mathbb{P}(w_i \cdot D_i > 0 | w_i \cdot x_i)$. The strong overlap assumption, 4, implies that the probability of being treated, given any $w_i(l) \cdot x_i$, should be positive. Later we discuss how these assumptions are satisfied in our GIPT Monte Carlo Simulations and in our GIPT empirical application.

The geographically weighted IPT estimator is estimated with a kernel based moment condition. The following additional regularity assumptions are needed for the GIPT estimator to have desirable large sample properties. Assumptions 6 through 8 are analogous to assumptions made by Abrevaya and Donald (forthcoming).

Assumption 6. *(Distribution of X): the Support χ of the k-dimensional covariate X is a Cartesian product of compact intervals, and the density of X, $f(X)$ are $p-$times continuously differentiable over χ.*

Assumption 7. *(Kernels): $K(\cdot)$ is a kernel of order s, is symmetric around zero, is equal to zero outside $\prod_{i=1}^{k}[-1,1]$, integrate to 1 and is continuously differentiable.*

Assumption 8. *(Bandwidths): The bandwidths b satisfy the following conditions as $N \to \infty$:*

$b \to 0$ and $log(N)/(Nb^{k+s}) \to 0$.

With GIPT we estimate an ATE for each target observation. In footnote 21 of the Appendix of Graham et al (2012), 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 target observation is similar to the overall ATE generation process outlined by Graham et al (2012), but we modify the moments condition using $\tau(w_{ij}x_i)$ instead of $t(x)$ as shown in equations 6 and 8 above. With GIPT, we obtain a very representative estimate of the ATE by generating an ATE for each target point, rather than generating one treatment effect for each target point and using these to calculate one overall ATE. Assumptions 1 through 8 are satisfied in our Monte Carlo study below, and in many applications that consist of randomized treatments. We focus on such randomized treatments in our application later in this paper. In applications where one knows the exact locations of the observations, we would expect that generation of the separate ATEs for each observation would lead to a precise estimate of the ATE at each location, and in turn, the overall average of the ATEs may have lower bias.

We next perform Monte Carlo simulations to demonstrate that the GIPT estimator performs well in small samples.

4. Monte Carlo Study

We first denote the two-dimensional location vector, $l_i = [l_i^1, l_i^2]$ .In this Monte Carlo study we generate our response variables, y_i, from the following causal model and selection model:

$$y_i = \beta_0(l_i) + DT_i \cdot DS_i \cdot \beta_1(l_i) + x \cdot \beta_2(l_i) + u_i, \tag{4.1}$$

$$DS_i = \begin{cases} 1 & for \quad l_i^1 + 0.25 \times l_i^2 > 1.25 \\ 0 & for \quad l_i^1 + 0.25 \times l_i^2 < 1.25 \end{cases}, \; i = 1, \cdots, N \tag{4.2}$$

$$DT_i = \begin{cases} 1 & for \quad i > N/2 \\ 0 & for \quad i \leq N/2 \end{cases}, \; i = 1, \cdots, N \tag{4.3}$$

where (4.1) is the causal model that produces the response variable y_i, (4.2) and (4.3) is the selection model that produces the treatment group. If DS_i equals 1, this indicates that the unit is in the location where some observations are treated and 0 indicates being in the control group. Also, DT_i is a dummy such that a value of 1 indicates an observation is only possibly treated after an unexpected event. Therefore, the treated sample will be comprised of the observations for which $D_i = DT_i \times DS_i = 1$; in other words, the treated sample consists of those units for which both $DS_i = 1$ and $DT_i = 1$. The vector $l_i = [l_i^1, l_i^2]$ is a two-dimensional location vector generated from a bi-variate uniform distribution 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, for simplicity we set $\beta_0(l_i) = 0$ and $\beta_2(l_i) = 0.2$, and $\beta_1(l_i)$, our main interest in the estimation, is a bi-variate standard normal density function:

$$\beta_1(l_i) = \tfrac{1}{2\pi} \exp\left(\tfrac{(l_i^1)^2 + (l_i^2)^2}{2}\right).$$

Note that this data generating process - as given in (4.1) (4.2) and (4.3) - is designed to meet the identification strategy and assumptions discussed in Section 3. First, the distribution of the outcome, Y, is independent of the treatment status ("unconfoundedness"; Second, $\{Y_i, X_i, D_i\}_{i=1}^N$ are i.i.d. (the "random sampling" assumption). Third, $\mathbb{P}(D_i = 1|Y, X) = \mathbb{P}(D_i = 1|X)$ (The "missing at random" assumption). Finally, $\mathbb{P}(D_i = 1|X = x) = \mathbb{P}(D_i = 1) > 0$, as D_i and X are independent in these data generating processes (The "strong overlap" assumption).

We use two different sample sizes, $N = 300$ and $N = 600$, as the number of individuals. This model is estimated with difference-in-differences, IPT and GIPT as defined in section 2. For the GIPT estimator, the optimal bandwidth for each sample size is calculated through a grid search. For a grid of b values, the average squared error, $ASE(b) = \frac{1}{N}\sum_{i=1}^{N}\left\{\tilde{\gamma}_j^{GIPT} - \gamma_j^{GIPT}\right\}^2$, is computed for 100 replications and then averaged to estimate the mean ASE (MASE). The function $MASE(b)$ is then plotted over the grid values of b. The optimal bandwidth, b_{MASE}, is chosen to be the value of b that yields the minimum MASE value. One optimal bandwidth is obtained for each sample size for the GIPT estimator. Next, using the optimal bandwidth for each sample size, we perform 500 iterations for each sample size, and then compute the average bias and ASE for each. The average bias and average squared errors are reported in Table 1. In addition, in Figure 1 we also plot the distributions, with histogram and estimated density, of the ASE results from the 500 repetitions on each estimator with two different sample sizes.

Since some preliminary finite sample experimental evidence on the performance of the IPT estimator is already available (Graham et al, 2012), we are primarily interested in the performance of the GIPT relative to estimators that do not account for geographic variation. There are general regularities that are evident. As expected, increases in the sample size reduce the ASE for all estimators, suggesting that the estimators under study converge with sample size. Across both sample sizes, the IPT estimator performs at least as well as the difference-in-differences estimator, in both ASE and average bias. Improvement of GIPT, as measured by ASE, over IPT and difference-in-differences, ranges from 25% to 57%. The key implication of these results is that in situations where geographic variation is an important factor in the data, the proposed GIPT estimator provides a simple but effective way to account for it. The ASE distribution plots in Figure 1 indicate a similar pattern. For each of the three estimators, increases in the sample size from 300 to 600 generally shift the ASE distribution towards zero. When the three estimators are compared with each other for the same sample size, the ASE distribution of GIPT are much closer to zero than that of the other two estimators.

We also plot the GIPT estimated ATEs based on our simulations, in Figures 2a and 2c (separately for N=300 and N=600, respectively). The corresponding true ATEs for these samples are plotted in Figures 2b and 2d, respectively. In comparing the GIPT ATEs against the corresponding true ATEs, it is apparent that as the sample size increases from N=300 to N=600, the

GIPT ATEs more closely approximate the true ATEs. This implies that GIPT is a consistent estimator of the true ATEs as the sample size increases.

Meanwhile, the simulation results suggest that GIPT should only be used instead of the IPT estimator when the data include information about geographic location. One potential alternative to GIPT when there is geographic information in the data might be including the geographic location variables as control variables and using the IPT estimator. However, this would violate the strong overlap assumption (Assumption 4 above) because having some of the lower values of latitude, for instance, would place those observations in the control group and it also would preclude having any overlap with the treated group. Therefore, we do not consider including geographic location variables as control variables in the regular IPT estimator as a viable alternative. Moreover, we attempt this in the simulations, and find that the model is unable to solve, which is not surprising given that Graham et al (2012) indicate this outcome would be likely when the strong overlap assumption is violated.

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

The metro-Vancouver area was hit with a series of major storms in November, 2006, which led to severe mudslides that caused contaminated storm runoff to enter the water supply (Evans, 2007). Some parts of the metro area were required to boil water for an extended period of 10 days longer (i.e., 12 days total) than the rest of the metro area (CBC News, 2006). This impacted restaurants, coffee shops, and other water-dependent businesses (Dowd, 2006). The affected area included the City of Vancouver, while the adjacent City of Richmond (and many other parts of the metro area) had the advisory lifted on the second day. We examine how sale prices for properties that sold within several months after this advisory in a section of Vancouver (the treated sample) were affected differently from other properties sold in Vancouver several months before the advisory and properties that sold in nearby parts of Richmond before and after the advisory (the control sample). Thus, our identification strategy relies upon an unexpected event (the extended water boil advisory) that affects some geographic areas but not others. We have a missing data issue with this data set, because we know what properties in the control group sold for, but we do not know what these properties would have sold for if they had been in the treatment group. Thus, some sort of adjustment using a propensity score approach would be desirable. Meanwhile, there are clear differences in the geographic locations of properties in our sample. It is of interest to determine empirically how the effects of such a shock might be absorbed differently into property values across locations. Therefore, we consider three different approaches in this application in order to address the missing data problem, difference-in-differences, the IPT approach, and GIPT.

There is a literature that examines the effects of a storm on property values, including Bin et al (2013), Atreya and Czajkowski (2016), and others. None of this literature, however, considers the missing data problem in the same context or with the same approach as we are addressing it here. Also, most of the other studies in the literature focus on residential property values, while our study examines the commercial property value impacts. Finally, we are not directly interested in the impact of the storms in the entire metro Vancouver area in November 2006; rather, we study the effects of a water boil advisory that was imposed on some areas of the metro area, including the City of Vancouver, for much longer than others. Therefore, we can examine the differential impacts of the water boil advisory on treated versus control areas, shortly before versus shortly after the advisory. This is our identification strategy.

It is widely accepted in the real estate finance and investments literature (e.g., Ling and Archer, 2017), that a commercial property's value or sale price equals the ratio of its net operating income (NOI) to the capitalization rate (i.e., cap rate or discount rate). In theory, if there is an event that alters an investor's estimate of basic long term risk, then an increase in discount rate seems appropriate. In some cities, such as New Orleans, a major hurricane such as Katrina led to property destruction as well as major disruption in abilities of businesses to operate for an extended period of time. This increased risk likely led to a higher cap rate, due to the possibilities of repeat storm events in the future, which lowered the value of commercial properties. While people may have revised their estimate of New Orleans' vulnerability because of rising sea levels, eroded barrier marshes, etc, and difficulties in raising the levies to avoid future storms, the case of Vancouver was somewhat different. In other words, it is less likely that there was a risk adjustment because shortly after the 2006 storm, there were efforts completed to enhance the sewer systems to prevent future storms from causing similar problems that led to the water boil advisory in late 2006. Specifically, the resulting water boil advisory was not a permanent event, so therefore it did not alter estimates of basic long term risk. However, this 12 day water boil advisory in the city of Vancouver caused major disruption of some business operations, especially for those that were water-oriented such as supermarkets, restaurants, day care facilities, etc (Dowd, 2006; CBC News, 2006). Such a disruption can be expected to lead to lost revenues or additional costs, for instance, for certain businesses that are water dependent. These characteristics can be expected to impact their NOI, which translates into an effect on property values and in turn, the sale prices of many properties. But other commercial properties sale prices were not affected, perhaps because they may not have been as water dependent.

When we are estimating the ATE of the water boil advisory on the price per square foot of commercial properties, the lot size of the property is expected to be negatively correlated with the NOI (and the total sale price). This is due to the fact that a larger lot size requires higher expenses for lawn maintenance and snow removal, for instance. But the effect of lot size on the price per square foot may be either positive or negative. A larger lot size may or may not lead to economies of scale that are inherent in the maintenance of a commercial building. Greater economies of scale lead to higher NOI and therefore a higher price per square foot of the overall property. There also may be particularly strong price effects for older properties, or properties that have not been renovated recently. These older properties may be expected to rent for less, need more repairs, and require more to upkeep due to unanticipated issues resulting from the age of the property. This can also be expected to factor into the NOI for a property. In other words, an older property, or one that has not been renovated recently, should have a lower NOI than a similar, nearby property that has been renovated recently. Therefore, it is important to use the lot size and the effective age as a proxies for NOI. The effective age is the number of years between the year of most recent sale and the last major renovation of a property. Properties that were renovated in the year in which they were most recently sold have an effective age of 0. Similarly, properties that have never been renovated have an effective age equal to the actual age of the property. In our model specifications, we use as the control variable the interaction term of lot size (in thousand square feet) and the effective age of the property (in years). For reasons described above, these two variables are the two best proxies for NOI that we have available to us. Also, in the IPT and GIPT specifications, when we try to include two separate difference-in-differences for these two variables, using the first two moments of each, the model is unable to solve. We are interested in the treatment effect from the extended water boil advisory, and we desire to control for the lot size and effective age as proxies for NOI but are not directly interested in their marginal effects. Therefore, using the interaction term enables us to control for both of these factors as proxies for NOI. Finally, Graham et al (2011) and Anderson (1982) suggest interaction terms be included. So we use the first two moments of the interaction term in the IPT and GIPT specifications. Obviously, we use the interaction term in the difference-in-differences specification as well. In an attempt to avoid the potential problem of confounding factors that could be prevalent in other parts of the metro Vancouver area, we restrict our attention to a section of the metro area where some observations are in the City of Vancouver (which was subject to the water boil advisory for 12 days) and others in nearby parts of the neighboring City of Richmond (which had the water boil advisory lifted after one day). We avoid including properties in the central business district of Vancouver, where there are potentially many other confounding factors. We focus on a period of several months before, and several months after the 12 day water boil advisory which occurred for the City of Vancouver in November 2006. We end our sample in August 2007 because we want to avoid the effects of the recession that started in late-2007, and we begin in January 2006 because we want to avoid other events that might have impacted property values before 2006.

In our data set, there are 100 commercial sales observations in the selected neighborhoods between January 2006 and August 2007 for which there are also data on the building sale price, square footage, lot size and the effective age. For this reason, our sample focuses on these 100 observations. Figure 4 shows the locations of our sample of 100 commercial properties (for which we have usable data) that sold (as arms-length transactions) in parts of the City of Vancouver and City of Richmond between January 2006 and August 2007. These data are from the BC Assessment database, which were purchased from Landcor.

Descriptive statistics are presented in Table 2. The average commercial property sold for approximately C$ 211 per square foot, had a lot size of 36,000 square feet, had an effective age of 28.36 years (i.e., there were 28.36 years since the last major renovation), and 25 percent of the observations were in the treatment group (i.e., in the City of Vancouver - opposed to the City of Richmond - and sold after the extended water boil advisory was imposed on the City of Vancouver).

We first estimate the following difference-in-differences model: $Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + e$, where Y_i is price per square foot for property i, X_i is the product of the lot size and the effective age. We assume that 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 in our data set that sold between November 2006 and August 2007 in the City of Vancouver (i.e., after the water boil advisory), and $D_i = 0$ for properties that sold in the City of Richmond before and after the advisory, and those properties that sold in Vancouver before the advisory. The regression coefficient β_2 is the "treatment effect" of locating in the City of Vancouver after the storm.

The second model we estimate is the IPT model, where we consider the first 2 moments so that $t(x) = [1, X, X^2]$, and X is the product of the lot size and effective age, and Y is the sale price per square foot. We are reweighting the X's so that the sample mean and variance of X in the treated sub-sample (and separately, in the untreated sub-sample) equals the entire sample mean and variance of X. Once again, we utilize the same data set as we used for the difference-in-differences estimation. We then calculate the ATE based on the IPT estimator.

Finally, we estimate the GIPT model, with Gaussian kernel weights given as

$$w_i(l) = \left[exp(-0.5 * (d(l)/b)^2)\right]^{1/2}, \tag{5.1}$$

where $d_i(l)$ is the Euclidean distance between property i and location l, and b is a bandwidth parameter. We explain the bandwidth parameter determination in more detail below. In the GIPT model, we consider the first two moments and use $\tau(w_i(l)X) = [1, w_i(l)X, (w_i(l)X)^2]$ for each target point, l. In this context, we are re-weighting by including distance weights in the propensity score weighted averages of X so that the re-weighted mean and variance of X for the treated sample equals the re-weighted mean and variance for the entire sample.

We present the results of the difference-in-differences and the IPT estimations in Tables 3 and 4. First, we regress the sale price per square foot against a constant, the treatment dummy, and the product of the effective age and the lot size (for consistency with the IPT and GIPT estimations, we retain this interaction term here). The treatment dummy, D_i, has a coefficient estimate of $\beta_2 = -28.89$, implying that the typical commercial property in the treated sample sold for approximately C$ 28.89 less per square foot than the typical property in the control sample. Also, β_2 is highly insignificant (P-value=0.456). With IPT, the ATE is estimated to be C$ -35.55 (with P-Value=0.134), and with a standard error that is somewhat lower than the corresponding difference-in-differences standard error (C$ 23.70 opposed to C$ 38.55). This smaller estimated standard error for the IPT estimator is in line with our expectations based on our earlier Monte Carlo simulations results.

With the GIPT estimation approach, we first must determine which bandwidth to use in the estimation process. We first consider a "Rule of Thumb" bandwidth, as in Silverman (1986). However, this bandwidth selection criterion requires normality of the distances data in order for it to be applicable. Examining the Jarque-Bera statistics demonstrate that there is no supporting evidence for a symmetric distribution; in fact, nearly half of the distance vectors consist of observations that are skewed based on the Jarque-Bera normality test. Therefore, we estimate integer bandwidths in the range of 4 (km) through 12 (km), while integer bandwidths smaller than 4 cause difficulties in the GIPT estimations that preclude it from solving. We choose the smallest of these bandwidths, which was b=4. This allows for the maximum amount of variation in the parameter estimates. In fact, as we increase the bandwidth, the variation in the ATE estimates from GIPT across observations decreases dramatically, in general approaching the ATE estimate from IPT for the higher bandwidths. This result is expected, as with a higher bandwidth there are more observations receiving positive weight than with a lower bandwidth, so the GIPT ATE estimates with the higher bandwidths closely approximate the IPT ATE estimate.

In terms of the GIPT variants of the IPT assumptions that we describe in section 3.4 above, it is reasonable that our data set and application satisfy these assumptions. First, we rely on an identification strategy that considers properties that sold inside and outside of the water boil advisory zone, in a reasonably short time frame before versus after the advisory date. For unconfoundedness, we assume that we have random sampling for our data set, as the treatment does not depend on the price of the property. Specifically, one might argue that the kernel weights imply two properties, i and l, with a high $w_i(l)$ (i.e., properties close to each other) are necessarily treated. But this is not the case, as can be seen in Figure 4. Properties on the south side of the Frasier River are in Richmond (untreated), while those just to the north are in Vancouver. Also, any given pair of properties in Vancouver that are close to each other are not necessarily both treated, because some of the nearby properties sold before the advisory and were therefore untreated. For our control variables, the interaction of effective age and lot size, it is reasonable to assume that nearby observations have no impact on the value of these two variables at a particular target point. We have data missing at random, as we know what properties sold for at their location but not what they would have sold for at other locations. We also have strong overlap, since there are some older and younger properties, as well as some larger and smaller lot size properties, in both the treated and untreated samples. There is a unique propensity score estimated for each target point. We also assume assumptions 6 and 8 hold, and since we use the Gaussian kernel, the symmetric kernel distribution (assumption 7) is satisfied.

We next estimate the ATEs for all target points, l, using the GIPT estimator that we have developed in this paper. Figure 4 shows a map of the metro-Vancouver area with the locations of the sample of commercial properties that sold in the period of our sample, and the estimated

ATE for each property. The range of the ATEs for the 100 observations is C$ -8.45 to approximately C$ -55.05 , but the former ATE has a relatively large standard error and is statistically insignificant. We take the average of all of the 100 ATEs (which we denote as the "AATE"), in Table 5. In this context, the AATE equals approximately C$ -30.79, while the average of the standard errors is C$ 22.43. In general, the properties with the most negative and significant ATEs are located in the central and south areas of Richmond, while those with statistically insignificant ATEs are elsewhere. There are no properties with statistically significant ATEs in Vancouver.

The effects of this 12 day water boil advisory are expected to be relatively small. While the ATE from difference-in-differences and IPT are statistically insignificant, with GIPT we find that all 100 observations have negative ATEs, but 13 out of the 100 observations have statistically significant ATEs (with P-value$<$0.05). Thus, using GIPT enables us to unmask which specific locations would be significantly impacted by the water boil advisory. With GIPT, the ATEs are estimated based on the properties surrounding it or close by. The ATEs from GIPT are small in terms of the numbers of properties impacted, and we can determine where these properties are located. Interestingly, all of these properties with significantly negative ATEs are concentrated in 5 distinct neighborhoods of Richmond (which did not experience the 12-day extended water boil advisory). The GIPT approach tends to imply there will be similar ATEs for properties nearby each other, therefore we might expect that the ATE with IPT should be bigger than many of the ATEs from GIPT because IPT estimates the ATE based on the entire sample and does not down-weight for distant observations. This may also be an explanation for why the properties with significant ATEs are clustered together. Another explanation is the lot size. Many properties in the Vancouver part of our sample (some of which are in the treatment group) are located on very small lots, while those properties in Richmond (which consists of a subset of the control group) have larger lots. This has implications for re-weighting with the propensity scores, as one aspect of this is that the GIPT procedure re-weights based on the mean and variance of the geographically weighted product of the lot size and effective age of the properties. This heterogeneity in lot size across space can clearly impact the ATEs.

Finally, within each of these 5 neighborhoods of Richmond, at least one (and sometimes several) of the properties in our sample are in a water-intensive industry. For instance, in a neighborhood around Horseshoe Way in the southern part of Richmond, there is a company that manufactures liquid cleaning products and health/beauty products. Nearby there is a recycling center and a millworks production company. While we expect the ATE of the liquid product manufacturing company to be affected by an extended water boil advisory, the ATE of the other two companies in the same neighborhood are likely to be impacted by their proximity to the liquid product manufacturing company. About 0.5 km south of this neighborhood is another cluster of properties with significantly negative ATEs, including a company that processes fish products for use as fresh and preserved bait; nearby there is a produce market that undoubtedly relies on water to clean its produce; and an event planning company. In this situation, the fish products store and produce market may have a strong impact on the ATE of the event planning company due to its close proximity. Approximately 3 km north of this neighborhood, there is a daycare facility with a significantly negative ATE, which relies daily on clean water for the children and staff to wash hands, dishes, etc. There are no other commercial properties in our sample that sold near this daycare facility. Approximately 2 km to the northwest is a restaurant/bakery, and an office building. In this case, the restaurant/bakery clearly would be impacted by an extended water boil advisory, while the ATE of the office building may be impacted due to the proximity to the restaurant/bakery. Finally, approximately 0.5 km north of the restaurant/bakery there is a cluster of 4 other properties that have statistically significant (negative) ATEs. These include a large shopping plaza with restaurants, a coffee shop, doctor's offices, a drug store, and other offices. Very close to this shopping plaza is an automobile repair garage, a dermatology office, and an office building. It is likely that the water dependency of many of the businesses in the shopping plaza is one explanation for a significantly negative ATE for that property, while the significantly negative ATEs for the other nearby properties may be at least in part determined by proximity to the shopping plaza.

One might conjecture that some of the differences in ATEs in the treated area (in the City of Vancouver after the boil water advisory) versus the control area (in the City of Richmond before the boil water advisory, and both Richmond and Vancouver before the advisory) may be due to differences in property tax rates in the two cities in these two years. We analyzed the property tax rates in these two cities in 2006 and 2007, and found that the 2006 base rate in Richmond for class 6 properties (commercial) was C$ 22.38361 per thousand dollars of assessed values. There were some additional add-ons for sewer debt, which ranged between C$ 0.23300 and C$ 0.28300 in 2006, implying a total tax rate in the range of approximately C$ 22.64 per thousand dollars of assessed value. There is an additional parking tax for Richmond properties with parking, at a rate of C$ 0.78 per square meter of parking spaces. The 2007 tax rate in Vancouver for Class 6 properties (commercial) was C$ 24.87171. Therefore, there is a difference of approximately C$ 2.23 per thousand dollars of assessed value. Assuming this differential is expected to persist indefinitely into the future (i.e., an infinite time horizon), and a discount rate of 5%, this implies a difference of C$ 2.23*(1+0.05)/0.05 over the life of the property, or a total expected property tax differential of C$ 46.83 per thousand dollars of assessed value. We assume the sale price of a property is highly correlated with its assessed value. Then, if the ATE is C$ -45 for a property that sold in Richmond before the water boil advisory in 2006, for instance, then C$ 2.10 of this C$ -45, or less than 5% of the ATE, can be attributed to expected differences in property taxes in the two jurisdictions in the two years.

Finally, one might argue that a fuzzy regression discontinuity framework could be appropriate for this particular problem, as in Angrist and Pischke (2009). But this is not the case in our specific application. The propensity score,

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

does not necessarily jump at any particular value of x. There are both large and small lot sizes in our sample of properties in Richmond and Vancouver, and also there are both old and new properties in both cities as well (as required by the strong overlap assumption of IPT). Therefore, our X, the interaction term of lot size and effective age, does not have a natural jump point in the probability of treatment at any specific value of x. In future work, it may be of interest to explore how to address potential fuzzy regression discontinuity in the context of IPT and GIPT, for specific applications where at particular values of x there is a jump point in the propensity score.

6. Conclusion/Discussion

We develop a GIPT estimator that can be a useful technique to generate ATEs for each geographic location, and re-weight propensity score estimates when there is missing data, given information on the geographic locations of the observations. There are several benefits, as well as some potential limitations, of the GIPT approach. One advantage of GIPT is that we are able to generate heterogeneous ATE estimates for each target point across locations. We can also test for the statistical significance of each of the ATEs. The average of the ATE's, or the AATE, is one way of summarizing this information over all observations, if so desired. In our application, one may be particularly interested in the ATE estimates that are statistically significant, in order to determine where remediation should be undertaken to prevent similar damage to the water supply in the future. There are many other potential missing data problem applications of the GIPT estimator where it would be desirable to generate heterogeneous ATEs.

Another advantage of GIPT, as demonstrated by our Monte Carlo simulations, is that the bias and average squared errors of the GIPT estimator tends to be lower than the bias for the difference-in-differences and IPT estimators. However, this is only expected to hold if the data points exhibit geographic heterogeneity; otherwise, there is no clear advantage to using GIPT over IPT, and in fact, the additional computation time for GIPT would be a major drawback. Even when there is spatial variation in the data, GIPT is a more computationally intensive procedure and in some cases this may diminish its feasibility, especially in very large samples. However, one potential remedy to this curse of dimensionality is to consider some type of quantile regression approach in the context of GIPT, which is beyond the scope of this paper. We have also addressed the important issue of bandwidth selection, which is crucial for each specific context of a given empirical application when using the GIPT framework. As we have demonstrated in our application where there is information on the locations of the observations, the GIPT approach can extract important information about which individual observations have statistically significant ATEs, and it allows for heterogeneity in ATEs across space.

Clearly, there are advantages to both the IPT and GIPT approaches to addressing the missing data problem in generating heterogeneous estimates of ATE's. It is also clear that both IPT and GIPT are superior to difference-in-differences. IPT and GIPT perform better than difference-in-differences in our Monte Carlo simulations, and this is to be expected, in part because difference-in-differences ignores the missing data problem.

In future work, it would be of interest to consider modifying the GIPT framework to contexts where there is a balanced panel (space-time), to address a broader array of applied missing data problems. Such an extension could also contribute to the literature on treatment effect heterogeneity by allowing for the possibility that the ATE could vary over both geographic space and over a long period of time. This may first necessitate extension of the regular IPT framework to a balanced panel data setting, as well as generating Monte Carlo evidence to validate the performance of the approach.

References

[1] Abrevaya, J., & Donald, S. G. Forthcoming. A GMM approach for dealing with missing data on regressors and instruments. Review of Economics and Statistics.

[2] Anderson, J. A. (1982). 7 Logistic discrimination. Handbook of statistics, 2, 169-191.

[3] Angrist, J.D. & J.S. Pischke. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

[4] Atreya, A., & Czajkowski, J. (2016). Graduated flood risks and property prices in Galveston County. Real Estate Economics.

[5] Bin, O., & Landry, C. E. (2013). Changes in implicit flood risk premiums: Empirical evidence from the housing market. Journal of Environmental Economics and management, 65(3), 361-376.

[6] Bitler, M. P., Gelbach, J. B., & Hoynes, H. W. Forthcoming. Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment. Review of Economics and Statistics.

[7] Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical analysis, 28(4), 281-298.

[8] CBC News. 2006. Greater Vancouver boil-water advisory lifted. November 27. http://www.cbc.ca/news/canada/british-columbia/greater-vancouver-boil-water-advisory-lifted-1.584398 (accessed on 7/22/2017).

[9] Dowd, A. 2006. Water Warning Leaves Vancouver High and Dry. The Star Online, November 18. http://www.thestar.com.my/news/world/2006/11/18/water-warning-leaves-vancouver-high-and-dry_1/ (Accessed on 7/24/2017).

[10] Evans, G.M. 2007. Taking our water supply for granted. BC Medical Journal, 49(2): 62.

[11] Frölich, M., Huber, M., & Wiesenfarth, M. (2017). The finite sample performance of semi-and non-parametric estimators for treatment effects and policy evaluation. Computational Statistics & Data Analysis, 115, 91-102.

[12] 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.

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

[14] Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. Political Analysis, 20(1), 25-46.

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

[16] Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(1), 243-263.

[17] Ling, D., & Archer, W. (2017). Real estate principles: A value approach. 5th edition. McGraw-Hill Higher Education.

[18] 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.

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

[20] Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

[21] Silverman, B. W. (1986). Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability, London: Chapman and Hall, 1986.

[22] Wager, S., & Athey, S. (2017). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, (just-accepted).

[23] Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. Journal of Econometrics, 141(2), 1281-1301.

Appendix A. Figures

Figure 1: Simulation Results on Average Squared Errors (ASE) Distributions From DID, IPT and GIPT

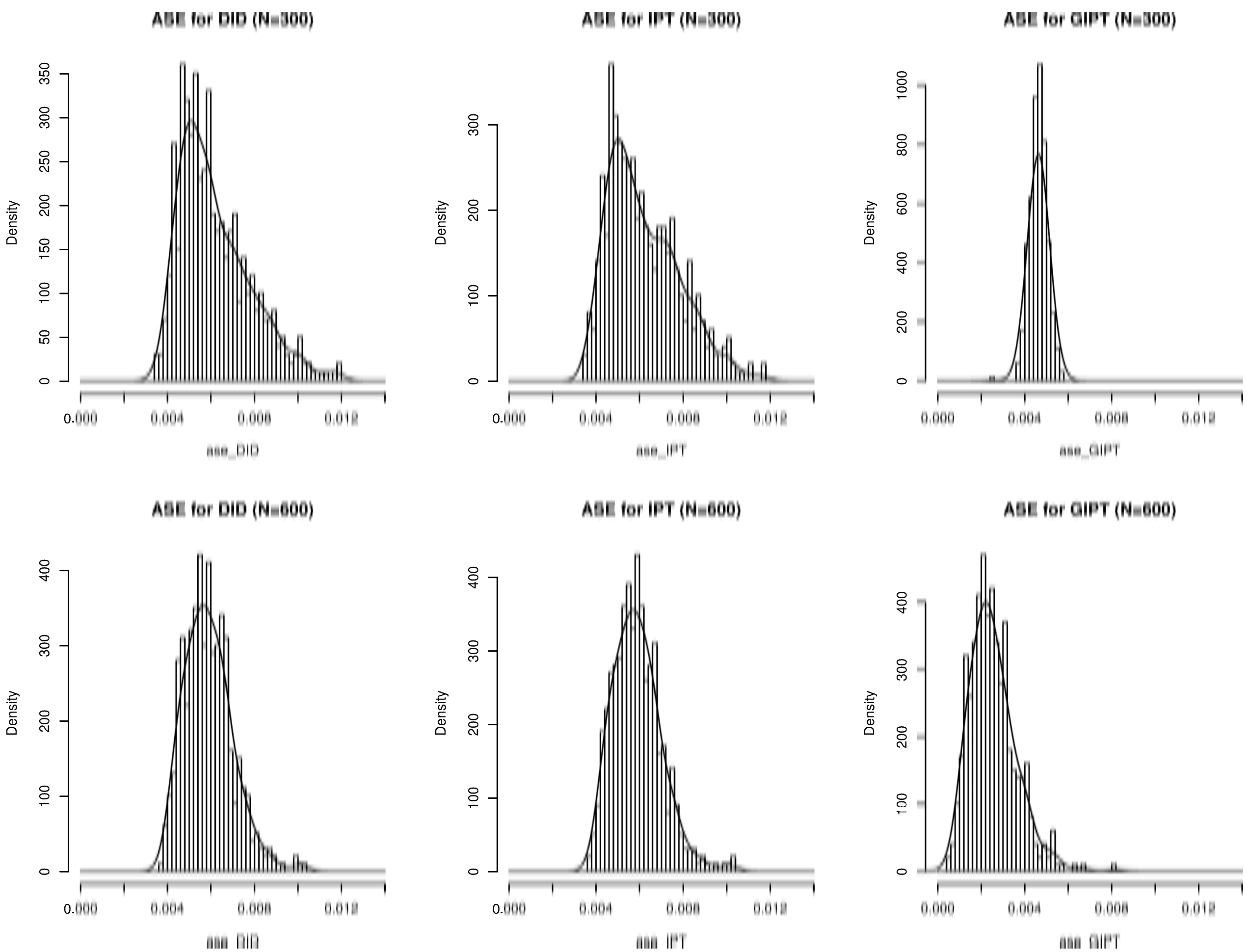


Figure 2: Simulations - The True ATEs and The GIPT Estimates

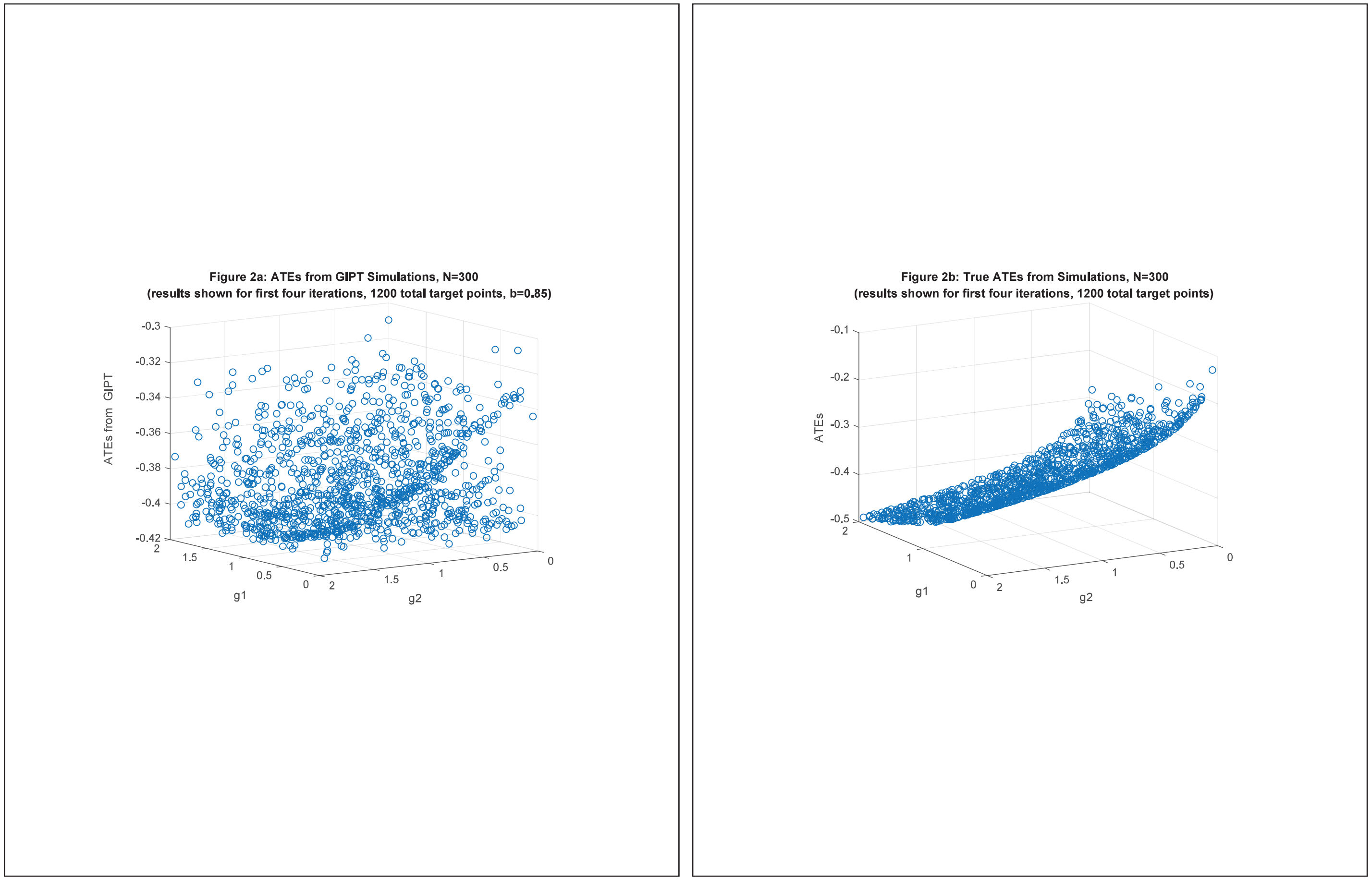


Figure 2c: ATEs from GIPT Simulations, N=600
(results shown for first two iterations, 1200 total target points, b=0.75)

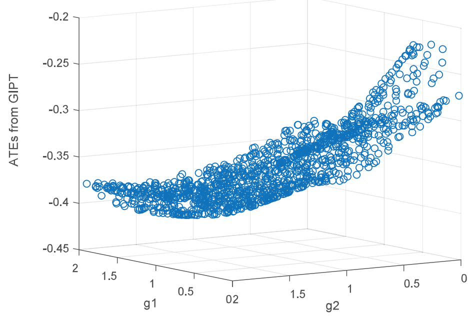


Figure 2d: True ATEs from Simulations, N=600
(results shown for first two iterations, 1200 total target points)

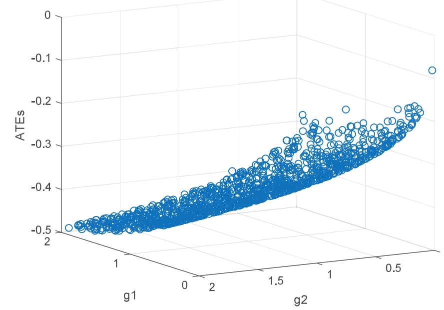


Figure 3: ATEs from GIPT, b=4, Vancouver/Richmond, BC Canada
Commercial Properties Sold between 1/2006-8/2007

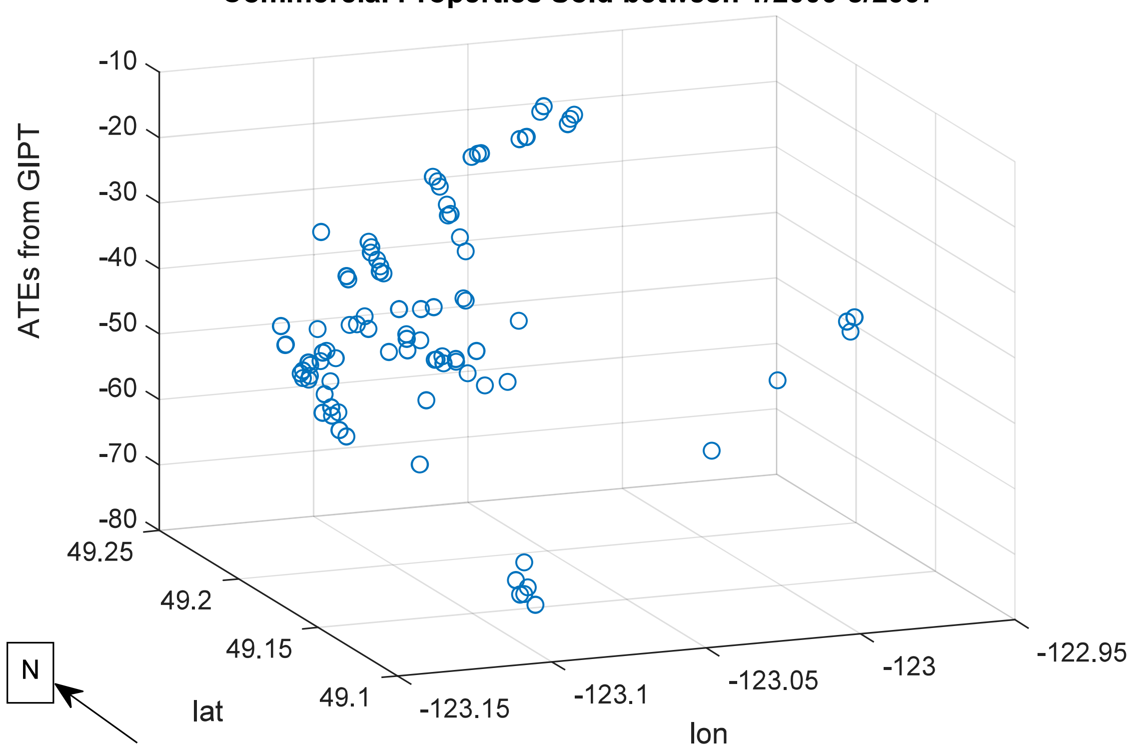


Figure 4: ATE Estimates from GIPT, 100 Commercial Properties, Vancouver and Richmond, BC, Sales from 1/2006-8/2007

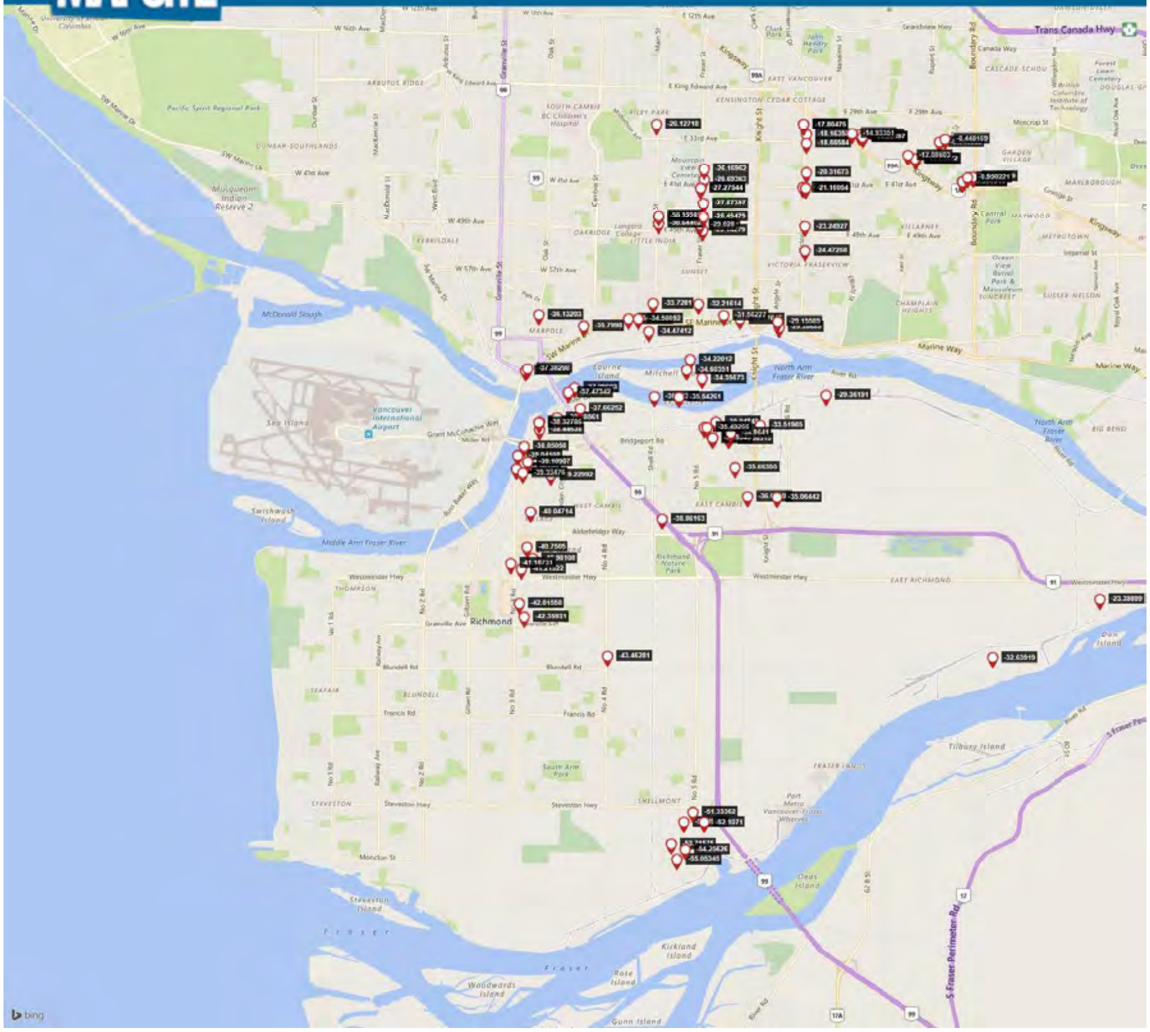
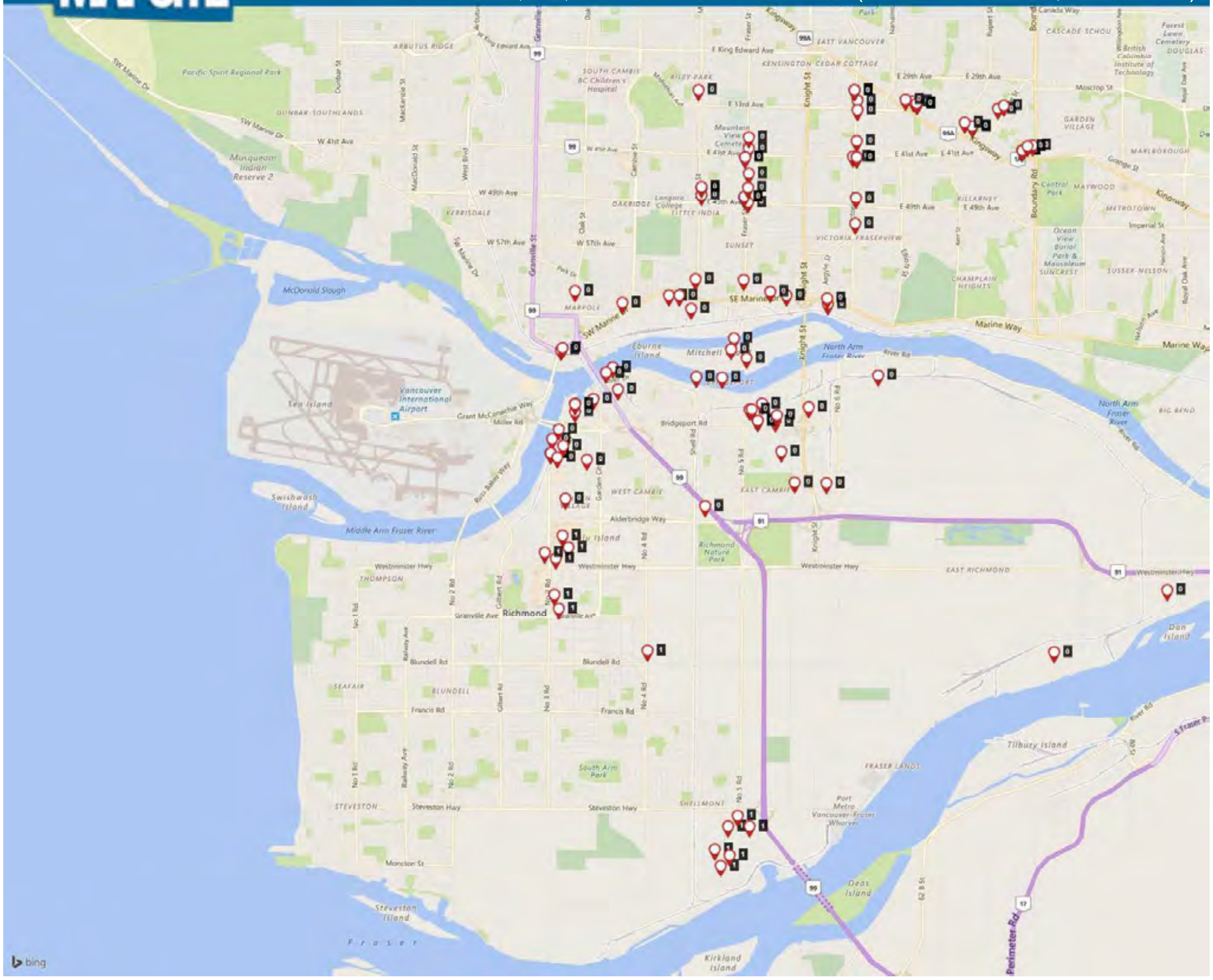


Figure 5: Statistically Significant ATE Estimates from GIPT, 100 Commercial Properties, Vancouver and Richmond, BC, Sales from 1/2006-8/2007 (=1 if P-value<0.05, 0 otherwise)



2. Tables

Table 1: Simulation Results - Small Sample Performances for GIPT, IPT and difference-in-differences[1]

	GIPT	IPT	difference-in-differences
		Sample Size = 600	
Bias	-0.007421	0.040871	0.0409438
ASE	0.002532	0.005879	0.005884
		Sample Size = 300	
Bias	0.004654	0.042592	0.042613
ASE	0.004632	0.006193	0.006189

[1]The bandwidth used for GIPT is 0.75 with N=600 and 0.85 with N=300. See section 4 for more details for bandwidth selection algorithm.

Table 2 - Descriptive Statistics

y is sale price per square foot of building area (C$); sales between 1/2006-8/2007

Variable:	y	DUMMY TREATMENT	EFFECTIVE AGE	LOT SIZE (THOUS SQ FT)	LATITUDE	LONGITUDE
Mean	211.0828	0.25	28.36	36.02742	49.2016	-123.0862
Median	161.6788	0	31	17.2062	49.20166	-123.0905
Maximum	1128.099	1	60	246.88	49.24108	-122.9651
Minimum	20.60159	0	0	2.76459	49.12316	-123.1376
Std. Dev.	167.7304	0.435194	12.7045	48.44602	0.030151	0.038866
Observations	100	100	100	100	100	100

Table 3 – Difference-In-Difference (DID) Regression Results

y is sale price per square foot of building area (C$)

"dummy treatment" is the Average Treatment Effect (ATE) parameter

commercial property sales between 1/2006-8/2007

Variable:	Coefficient	Std. Error	t-stat	P-value
dummy treatment	-28.88535	38.55443	-0.75	0.456
(lot size)×(effective age)	-0.0220654	0.012104	-1.82	0.071
constant	236.5202	21.11487	11.2	0
R-squared	0.042			
Observations	100			

Table 4 - Inverse Probability Tilting (IPT) Estimation Results

y is sale price per square foot of building area (C$);
ATE is the Average Treatment Effect parameter;
commercial property sales between 1/2006-8/2007

	Coefficient	Std. Error	z-stat	P-value
Treated Sample				
δ^1				
(lot size)×(effective age)	-0.0007207	0.000492	-1.46	0.143
$[(\text{lot size})\times(\text{effective age})]^2$	1.24E-07	6.88E-08	1.81	0.07
constant	-0.8611632	0.31655	-2.72	0.007
Control Sample				
δ^0				
(lot size)×(effective age)	-0.0020906	0.001804	-1.16	0.246
$[(\text{lot size})\times(\text{effective age})]^2$	3.23E-07	2.55E-07	1.27	0.204
constant	-0.4014677	0.528823	-0.76	0.448
ATE	-35.55546	23.69936	-1.5	0.134
Observations	100			

Table 5 - Geographically-weighted Inverse Probability Tilting (GIPT) Descriptive Statistics

y is sale price per square foot of building area (C$)
GIPT is the estimates of Average Treatment Effect parameter based on geographically-weighted inverse probability tilting method; there are 100 separate GIPT estimates
commercial property sales between 1/2006-8/2007

	Mean	Std. Deviation	Min	Max
GIPT	-30.788	11.193	-55.054	-8.448
Std Error of GIPTs	22.434	1.765	18.739	24.928
Observations	100			