Regression and matching in hedonic analysis: Empirical guidance for estimator

Klaus Moeltner

Professor, Department of Agricultural and Applied Economics, Virginia Tech, moeltner@vt.edu

Roshan Puri

Ph.D. Candidate, Department of Agricultural and Applied Economics, Virginia Tech, puriroshan@vt.edu

Robert J. Johnston

Professor, George Perkins Marsh Institute, Clark University, rjohnston@clarku.edu

Selected Paper prepared for presentation at the 2023 Agricultural & Applied Economics Association Annual Meeting, Washington DC; July 23-25, 2023

Regression and matching in hedonic analysis: Empirical guidance for estimator

Abstract

We illustrate how estimation results from hedonic regression, basic matching, and regression-adjusted matching can be combined to provide guidance on model choice for housing market studies with focus on binary treatment effects. We show that single-control matching is no longer optimal when spatial and temporal fixed effects are present, and exact matching in space and time is practically infeasible. A larger number of matched controls can, in fact, trigger a beneficial "cancellation effect," a to date unexplored route towards unbiasedness. This, in turn, reveals an empirical prescription for model search by choosing the number of matched controls that drive the cancellation effect to zero. In our application to flood zone discounts for property values in Massachusetts, we illustrate that a close-to-zero cancellation effect is empirically feasible, and that the resulting preferred model produces reasonable estimates and has desirable efficiency properties.

keywords: Special Flood Hazard Areas, Nearest-Neighbor matching, interior flood risk, Bayesian estimation

Introduction

Hedonic regression models remain popular for the analysis of environmental impacts on housing markets (e.g. Bishop et al. 2020; Mei et al. 2021; Manning and Ando 2022; Wolf and Takeuchi 2022; Wu et al. 2023; Brolinson et al. 2023), but rely heavily on the correct specification of the price function (Imbens and Wooldridge 2009; Kuminoff et al. 2010). In contexts where the main analytical focus is on a binary treatment indicator, non- or semi-parametric matching estimators can alleviate mis-specification risk. Recent examples in the environmentally motivated housing market studies include Abbott and Klaiber (2013),Muehlenbachs et al. (2015), Meldrum (2016), Moeltner et al. (2017), and Johnston and Moeltner (2019). A few of these studies directly compare matching to traditional hedonic regression, and arrive at mixed conclusions regarding the relative quality of the respective treatment estimates (e.g. Abbott and Klaiber 2013; Muehlenbachs et al. 2015; Johnston and Moeltner 2019).

However, to date the *combined inferential power* of traditional regression and different types of matching estimators has not yet been harnessed in the hedonic literature. More broadly, the closely related question of the optimal number of matched controls, especially when matching is combined with regression, remains largely unanswered in the empirical economic literature at large (e.g. Imbens 2004).

The main objective of this paper is to develop prescriptive recommendations for housing market analysts and practitioners for choosing between regression and matching estimators when the objective is to estimate an unbiased binary treatment effect, while properly controlling for three sets of *observed confounders*: (i) standard home and neighborhood characteristics, (ii) generic spatial effects (modeled via fixed effects of spatial affiliation), and (iii) generic temporal effects (modeled via fixed effects of temporal affiliation). The second objective is to examine how the finite sample properties of these estimators behave under an increasing number of matched control observations in matching approaches and in regression with pre-matched ("balanced") data. Third, we show how the combined estimation output from regression and matching can be used to assess risk of bias and relative efficiency of these estimators, which can further guide model selection.

We first illustrate that for the type of non-experimental home sales data typically used in hedonic analysis, treatment effects estimated via linear regression models are at high risk of being biased. This is because the distribution of observables for treated and controls are generally not in perfect alignment, as can be determined via a *balance analysis*. As a result, the treatment effect will be correlated with the remaining data, making a correct specification of the regression function an absolute necessity to avoid bias. We further show that pre-matching at an increased number of matched controls does not fully alleviate this problem.

We then turn our attention to nonparametric and mixed matching estimators. We first consider a fully nonparametric matching estimator based solely on observed price differentials between treated and matched pairs. In regression contexts where spatial and temporal fixed effects do not play a role, this estimator has generally found to be biased with high probability since treated and matched controls generally differ in observable dimensions other than treatment status, despite best matching efforts. This can be confirmed via a multi-attribute *distance analysis*. Increasing the number of matches per treated further deteriorates distance, and is unlikely to help in that respect. However, we show that when spatial and fixed effects are brought to the problem, a second route towards unbiasedness opens even for the basic case. This occurs when the portion of the price difference between treated and matched controls related to observable characteristics cancels out with the portion associated with fixed effects, for the sample at large. We label this "difference in differences" a "cancellation effect." We show that, empirically, this effect can indeed be close to zero, and that it is a function of the number of matches per treated.

Moving on to the adjusted matching estimator proposed by Abadie and Imbens (2011), a hybrid between regression and matching, we first show that the efficiency properties of this estimator improve at increasing numbers of matched controls, as more observations are available to estimate the regression portion. We then illustrate that the bias of the adjusted estimator decreases with increasing strength of the cancellation effect. This suggest that searching for the number of matches that produces an estimated cancellation effect close to zero for the adjusted estimator is a productive avenue towards lowering the risk of a biased treatment estimate.

For our application to Special Hazard Flood Area (SFHA) effects in selected counties of Massachusetts, our preferred model predicts home value losses that are tightly estimated, of reasonable magnitude, and in alignment with those found in Johnston and Moeltner (2019) for a similar application and under similar modeling. However, in contrast to that study, which exclusively uses one-to-one matching and does not consider any cancellation effects, our estimates based on a much larger number of matches are twice as efficient as judged by posterior confidence bounds.

Modeling

We start by assuming the true population models for a treated and control home i with sale prices of y_{1i} (if treated) and y_{0i} (if untreated), respectively, are given as follows:

$$
\begin{aligned}
(1) \qquad & y_{1i} = \mu_1(\mathbf{x}_i, \boldsymbol{\beta}) + \nu_i + \tau_i + \epsilon_i \\
& y_{0i} = \mu_0(\mathbf{x}_i, \boldsymbol{\beta}) + \nu_i + \tau_i + \epsilon_i,
\end{aligned}
$$

where $\mu(.)$ is some function of observed home and neighborhood characteristics $\mathbf{x}_i$ and corresponding parameters $\boldsymbol{\beta}$, ν_i is a spatial fixed effect (e.g. school zone), τ_i is a temporal fixed effect (e.g. sale year), and ϵ_i is an error term capturing additional unobservables that affect home prices. Without loss of generality we assume a typical normal distribution with zero mean and common variance for this error, i.e. $\epsilon_i \sim n(0, \sigma^2)$. We further assume that there exists a finite set of S spatial zones and T temporal windows, such that $\nu_i \in \{\nu_s\}_{s=1}^{S}$ and $\tau_i \in \{\tau_t\}_{t=1}^{T}$, and with both S and T substantially smaller than total sample size n to assure identification.

In addition, we make the following assumptions throughout: (i) The idiosyncratic error ϵ_i is uncorrelated with the remaining model components. In other words, there are no confounding unobservables that could bias parameter or treatment effect estimates. This "unconfoundedness" assumption is standard in the general matching-on-observables literature (Imbens 2004; Abadie and Imbens 2006; Ho et al. 2007; Abadie and Imbens 2011) (ii)

Model parameters, such as $\boldsymbol{\beta}$, remain invariant over space and time. If this is questionable, the sample could be adjusted by narrowing the time frame and / or extent of the market. (iii) There are no unobserved effects that systematically vary over time and space. This could potentially be relaxed with a very large / rich data set of home sales that contains a sufficient number of controls per treated within each time-space unit (e.g. school zone in a specific sale year). In that case, the additive spatial and temporal effects in (1) could be replaced with a space-plus-time fixed effect without loss of generality.

Linear regression with treatment indicator

We first consider a generic regression model with a simple additive treatment effect, that is $\mu_1(.) = \mu 0(.) + \gamma_R$, where "R" stands for "Regression". The sought treatment effect γ_R can then be interpreted as the population expectation (over stochastic component ϵ) of a home with characteristics $\mathbf{x}_i, \nu_i, \tau_i$ in treated versus untreated state, i.e.:

$$(2) \qquad E\left(y_{1i} - y_{0i} | \mathbf{x}_i, \boldsymbol{\beta}, \nu_i, \tau_i\right) = \mu_0\left(\mathbf{x}_i, \boldsymbol{\beta}\right) + \gamma_R + \nu_i + \tau_i - \left(\mu_0\left(\mathbf{x}_i, \boldsymbol{\beta}\right) - \nu_i + \tau_i\right) = \gamma_R$$

In practice, this model would be empirically implemented as a regression with an indicator variable that flags treated observations. Most commonly, the analyst assumes a linear function for $\mu(.)$, resulting in the following specification at the sample level with n observations:

$$(3) \qquad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \gamma_R \mathbf{d} + \mathbf{N}\boldsymbol{\nu} + \mathbf{T}\boldsymbol{\tau} + \boldsymbol{\epsilon}$$

where $\mathbf{y}, \boldsymbol{\nu}, \boldsymbol{\tau}$ and $\boldsymbol{\epsilon}$ are n by 1 vector-equivalents of their counterparts in (1), $\mathbf{X}$ is the full n by k matrix of home and neighborhood attributes for all observations, $\mathbf{d}$ is a vector of binary treatment indicators taking the value of 1 for treated and 0 for untreated ("control") observations, and $\mathbf{N}$ and $\mathbf{T}$ are n by $S-1$ and n by $T-1$ indicator matrices flagging, respectively, spatial and temporal membership for each observation. Letting $\mathbf{Z} = \begin{bmatrix} \mathbf{X} & \mathbf{d} & \mathbf{N} & \mathbf{T} \end{bmatrix}$, and $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\beta}' & \gamma_R & \boldsymbol{\nu}' & \boldsymbol{\tau}' \end{bmatrix}'$, the OLS estimator for all model coefficients can be compactly expressed as:

$$(4) \qquad \hat{\boldsymbol{\theta}} = \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{y}$$

The estimated treatment effect $\hat{\gamma}_R$ is the $k+1$ element in $\hat{\boldsymbol{\theta}}$, and will typically be a function of the entire data. This immediately implies that any mis-specification of the regression function, either via falsely assuming linearity, or a false specification of $\mathbf{X}$ (e.g. linear vs. log transformation, squared terms, interactions, etc.) will lead to a biased estimate, as also noted in Imbens (2004). At the other extreme, the specification of $\mu\left(\mathbf{X}, \boldsymbol{\beta}\right)$ becomes irrelevant for causal estimation of the treatment effect if treatment status is completely independent, or orthogonal, to all other included regressors. In that case, one could simply run a regression with an intercept and the treatment vector, and the resulting estimate would be identical to the mean sale price of the treated minus the mean sale price of the controls, that is:

$$(5) \qquad \hat{\gamma}_R = \bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_0,$$

where $\mathbf{y}_1$ is the n_1 by 1 vector of home prices for all treated observations, $\mathbf{y}_0$ is its n_0 by 1 counterpart for the untreated, and the "bar" notation indicates sample averages. This would be akin to a fully experimental setup in other applications (say, labor market training), but is unlikely to hold for home sales data.

Full independence of $\mathbf{d}$ from $\mathbf{X}$ in (3) occurs when treated and controls share identical empirical distributions for all variables in $\mathbf{X}$, a condition described as "perfect balance" (e.g. Ho et al. 2007). Intuitively, this implies that nothing sets treated and controls apart - at least in observable dimensions - other than the actual treatment. An assessment of balance, for example by comparing distributional percentiles for a given covariate between treated and controls can thus be used to gauge the degree of independence of the treatment indicator from the remaining data, or, equivalently, the degree to which the specification of the regression function $\mu\left(.\right)$ matters (see e.g. Johnston and Moeltner 2019).

Balance can be improved with pre-matching, a process of selecting a subset of controls to feed into the regression models such that the empirical distributions of treated and controls approach alignment. By definition, this will enhance independence (Ho et al. 2007). We will compare unmatched and pre-matched regression-based treatment effects in our empirical application. However, until perfect balance / independence is achieved - a tall order for housing data even with careful pre-matching - the choice of the regression function matters for the unbiased estimation of γ_R. This places the standard regression approach at a distinct disadvantage compared to the non-and semi-parameteric methods that follow.

Formally, the estimation bias for γ_R at the observation level can be expressed as:

$$
\begin{aligned}
(6) \quad & B\left(\hat{\gamma}_R\right) = \\
& \mu_0\left(\mathbf{x}_i, \boldsymbol{\beta}\right) + \gamma_R + \nu_i + \tau_i - \left(\mu_0\left(\mathbf{x}_i, \boldsymbol{\beta}\right) + \nu_i + \tau_i\right) - \\
& E\left(\hat{\mu}_0\left(\mathbf{x}_i, \hat{\boldsymbol{\beta}}\right) + \hat{\gamma}_R + \hat{\nu}_i + \hat{\tau}_i - \left(\hat{\mu}_0\left(\mathbf{x}_i, \hat{\boldsymbol{\beta}}\right) + \hat{\nu}_i + \hat{\tau}_i\right)\right) = \\
& \gamma_R - E\left(\hat{\gamma}_R\right),
\end{aligned}
$$

where $\hat{\mu}\left(.\right)$ denotes the assumed form of the regression function in empirical implementation. While this bias expression does not directly include the presumed regression function $\hat{\mu}_0\left(\mathbf{x}_i, \hat{\boldsymbol{\beta}}\right)$, it does so implicitly via the dependence of the estimator $\hat{\gamma}_R$ on the functional form of $\hat{\mu}_0\left(.\right)$, as well as the entire $\mathbf{X}$ matrix, as discussed above.

Non-parametric matching estimator

While pre-matching the data is optional for regression analysis, it is a mandatory first step for the remaining estimators considered in this paper. Matching is the process of pairing each treated observation with one or more controls based on a measure of closeness or "distance" between $\mathbf{x}_i$ and $\mathbf{x}_j$, the Euclidean sense. In this study we consider nearest-neighbor (NN) matching, where this distance is based on a set of observable variables, such as those that figure on the right hand side of the regression model discussed above. Normally, this is the full set of continuous regressors, potentially with additional interactions and squared terms. For simplicity, we will continue to use the $\mathbf{x}$ notation for this set, with the understanding that it could differ slightly from the full set of covariates in the regression model.[1] The distance metric in NN matching can be computed as:

$$d_{i,j} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)' \mathbf{V}^{-1} (\mathbf{x}_i - \mathbf{x}_j)}, \tag{7}$$

where $\mathbf{V}$ is the sample variance-covariance matrix for $\mathbf{x}$ (for the *full* sample), or, alternatively, its diagonal. In the first case, the distance measure in (7) is referred to as *Mahalanobis* metric (e.g. Abadie and Imbens 2011). Abadie et al. (2004) and Abbott and Klaiber (2013) use the diagonal version instead.

Let M be the desired *minimum number of matched observations*, also referred to as the desired "number of neighbors." If there are ties, the actual set of matched observations might be larger than M.[2] Furthermore, let the set of M nearest neighbors corresponding to treated observation i be labeled as $\mathcal{J}_{i,M}$. That is, the elements j in $\mathcal{J}_{i,M}$ have a value of $d_{i,j}$ that is no larger than the M^{th} closest distance from i. Thus, if we let the number of elements in $\mathcal{J}_{i,M}$ be denoted as $L_{i,M}$, we have $L_{i,M} \geq M$.

As in Abadie and Imbens (2011) let $\mathcal{K}_j$ be the number of times each j was used as a match for some i, weighted, in each case, by the total number of matches for i, that is:

$$\mathcal{K}_j = \sum_{i=1}^{n} \left(\frac{I(j \in \mathcal{J}_{i,M})}{L_{i,M}} \right), \tag{8}$$

where $I(.)$ is an indicator function, which takes the value of 1 if the condition it represents holds, and a value of 0 otherwise. This metric is needed for the computation of classical standard errors, as well as for matching estimators with regression adjustment, considered in the next section.

As shown in Abadie and Imbens (2011) a generic Average Treatment Effect on the Treated (ATT) can be obtained by paring each treated observation with one or more matched controls

[1] An alternative to NN matching would be matching on the propensity score, i.e. the estimated probability of falling into the treated group. We prefer NN matching as it is not reliant on the correct specification of the propensity score equation, usually a probit or logit model.

[2] Abadie and Imbens (2011) rule out ties in their exposition, arguing that ties are impossible with continuous regressors in $\mathbf{x}$. However, in our case regressors may be "discrete enough" to allow for ties, so we consider the more general case followed in Abadie et al. (2004)

(depending on M) and obtaining the difference in sale prices between the treated and the (mean of) the price(s) of the matched control(s). Individual differences are then averaged to yield a sample-level estimate for the sought treatment effect γ_B ("B for "Basic"). Formally:

$$\text{(9)} \qquad \hat{\gamma}_B = \frac{1}{n_1}\sum_{i=1}^{n_1}(y_{1i} - \hat{y}_{i0}) = \frac{1}{n_1}\sum_{i=1}^{n_1}\left(y_{1i} - \frac{1}{L_{i,M}}\sum_{j\in\mathcal{J}_{i,M}} y_{j0}\right)$$

where the outside summation is over all treated observations, and $\hat{y}_{i0}$ is the estimated counterfactual sale price for treated observation i.

The finite sample bias for this estimator can be written as:

$$\text{(10)} \quad B(\hat{\gamma}_B) = E\left(\frac{1}{n_1}\sum_{i=1}^{n_1}((y_{1i} - \hat{y}_{i0}) - (y_{1i} - y_{i0}))\right) = E\left(\frac{1}{n_1}\sum_{i=1}^{n_1}(y_{0i} - \hat{y}_{i0})\right) =$$
$$\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\mu_0(\mathbf{x}_i, \boldsymbol{\beta}) + \nu_i + \tau_i - \frac{1}{L_{i,M}}\sum_{j\in\mathcal{J}_{i,M}}(\mu_0(\mathbf{x}_j, \boldsymbol{\beta}) + \nu_j + \tau_j)\right) =$$
$$\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{1}{L_{i,M}}\sum_{j\in\mathcal{J}_{i,M}}(\mu_0(\mathbf{x}_i, \boldsymbol{\beta}) - \mu_0(\mathbf{x}_j, \boldsymbol{\beta}))\right) -$$
$$\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{1}{L_{i,M}}\sum_{j\in\mathcal{J}_{i,M}}((\nu_i + \tau_i) - (\nu_j + \tau_j))\right)$$

where y_{i0} is the true (unobserved) counterfactual price of treated observation i. As has been pointed out numerous times in the literature (Imbens 2004; Abadie and Imbens 2006, 2011), and as is evident from the last line in (10), in absence of any fixed effects this bias goes to zero if every matched pair perfectly aligns on observables, that is $\mathbf{x}_i = \mathbf{x}_j, \forall\, i, j$. In the presence of fixed effects, the zero-bias condition becomes more stringent and adds the requirement that matched observations be located in the same spatial and temporal unit, i.e. $\nu_i = \nu_j, \;\; \tau_i = \tau_j, \forall\, i, j$.[3] In other words, perfect matching on observable characteristics is no longer sufficient to mitigate bias. Furthermore, it is unlikely to be schieved in practice for housing market applications, since most matched homes will differ at least to a residual

[3]We rule out the unlikely case of different spatial or temporal units sharing the exact same (true) effect on home prices.

extent in observable characteristics. In theory, it is possible in the matching process to force matched observations to reside in the same spatial and / or temporal location, which would eliminate the ν and τ terms in (10). However, as shown in Abbott and Klaiber (2013), forcing matches to the same spatial unit alone can substantially decrease the pool of candidate controls for a given treated, and thereby grossly inflate the Euclidean distance between $\mathbf{x}_i$ and the matched $\mathbf{x}_j$'s, thus aggravating bias concerns along the $\mathbf{x}$ dimension. This problem becomes even more pronounced if exact matching in time is imposed in addition to exact matching in space.

It tuns out that practical inability to force matches to common spatial and / or temporal units can become a virtue as it opens the door to a second route towards unbiasedness that has to date not been exploited in the matching literature. Specifically, bias vanishes if differences between matched observations along observables balance out with differences in spatial and temporal effects, i.e., if the difference between the last two lines in (10) goes to zero. We deem this route towards unbiasedness the *true cancellation effect*.

In concordance with the literature (Abadie et al. 2002; Imbens 2004; Abadie and Imbens 2011), we would ex ante expect bias problems along the lines of perfectly matched observables, i.e. the portion $(\mu_0(\mathbf{x}_i, \boldsymbol{\beta}) - \mu_0(\mathbf{x}_j, \boldsymbol{\beta}))$ in (10), to increase with a larger number of matched neighbors, as each additional match increases the matching distance in (7) by definition, and thus the gap between $\mathbf{x}_i$ and $\mathbf{x}_j$.[4] However, in the presence of fixed effects perfect matches along observable characteristics alone is no longer sufficient to eliminate bias. Thus, the optimal choice of M becomes an empirical question as increasing M may enhance the probability of reducing bias via the cancellation effect. There is no avenue for this basic estimators to empirically examine this conjecture since the elements in the last line of (10) are purely theoretical constructs. However, a more sophisticated regression-adjusted matching estimator can provide some guidance in that respect, as discussed in the next section.

[4]Abadie et al. (2002) show that this bias does not vanish asymptotically, i.e. with increasing sample size, even for $M = 1$. In this study, we will focus on finite sample results, given the practical limitations of considering a housing market of infinite size, and our Bayesian estimation framework for all estimators that include a regression component.

Non-parametric matching estimator with regression adjustment

Abadie and Imbens (2011) develop a nearest-neighbor matching approach with regression adjustment to control for residual differences in observables between treated and controls after matching. The adjusted ("A") estimator, also referred to as "Bias-Corrected Matching Estimator" (BCME), can be written as:

$$
\begin{aligned}
(11)\quad \hat{\gamma}_A &= \frac{1}{n_1}\sum_{i=1}^{n_1}\left(y_{1i}-\hat{y}_{i0}\right) = \\
&\frac{1}{n_1}\sum_{i=1}^{n_1}\left(y_{1i}-\frac{1}{L_{i,M}}\sum_{j\in\mathcal{J}_{i,M}} y_{j0}+\hat{\mu}_0\left(\mathbf{x}_i,\boldsymbol{\beta}\right)+\hat{\nu}_i+\hat{\tau}_i-\left(\hat{\mu}_0\left(\mathbf{x}_j,\boldsymbol{\beta}\right)+\hat{\nu}_j+\hat{\tau}_j\right)\right)
\end{aligned}
$$

where $\hat{\mu}_0\left(.\right)$ is a pre-specified regression function that is estimated using only matched controls, with each observation weighted according to (8). The adjustment thus consists of the difference in predicted prices flowing from this regression between the treated and a corresponding matched observation. The finite-sample bias of γ_A can be expressed as:

$$
\begin{aligned}
(12)\quad B\left(\hat{\gamma}_A\right) &= E\left(\frac{1}{n_1}\sum_{i=1}^{n_1}\left(y_{0i}-\hat{y}_{i0}\right)\right) = \\
&\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{1}{L_{i,M}}\sum_{j\in\mathcal{J}_{i,M}}\left(\left(\mu_0\left(\mathbf{x}_i,\boldsymbol{\beta}\right)-\mu_0\left(\mathbf{x}_j,\boldsymbol{\beta}\right)\right)+\left(\left(\nu_i+\tau_i\right)-\left(\nu_j+\tau_j\right)\right)\right)\right) - \\
&E\left(\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{1}{L_{i,M}}\sum_{j\in\mathcal{J}_{i,M}}\left(\left(\hat{\mu}_0\left(\mathbf{x}_i,\boldsymbol{\beta}\right)-\hat{\mu}_0\left(\mathbf{x}_j,\boldsymbol{\beta}\right)\right)+\left(\left(\hat{\nu}_i+\hat{\tau}_i\right)-\left(\hat{\nu}_j+\hat{\tau}_j\right)\right)\right)\right)\right)
\end{aligned}
$$

As for the generic ATT, the bias goes to zero if all observations match perfectly in observables, spatial zones, and time periods, i.e. $\mathbf{x}_i = \mathbf{x}_j$, $\nu_i = \nu_j$, and $\tau_i = \tau_j$, regardless of the assumed regression function $\hat{\mu}\left(.\right)$. As for the basic matching estimator γ_B, this can, in theory, be helped via forced matching on spatial zones and / or time periods, which annihilates the ν and / or τ terms from (12). However, as mentioned previously, this generally comes at the cost of poorer matches on $\mathbf{x}$, thus placing, again, a higher burden on the correct specification of $\hat{\mu}_0\left(.\right)$.

Second, the bias vanishes if the auxiliary regression model produces unbiased expectations for all components, i.e. if $E\left(\hat{\mu}_0\left(\mathbf{x}_r,\boldsymbol{\beta}\right)+\hat{\nu}_r+\hat{\tau}_r\right)=\mu_0\left(\mathbf{x}_r,\boldsymbol{\beta}\right)+\nu_r+\tau_r,\ r=i,j$. This second route towards unbiasedness has earned the adjusted estimator the label of "double robust" (Imbens 2004).

A third route towards bias reduction presents itself via the "cancellation" effect mentioned above between regression predictions based on observable housing and neighborhood characteristics, and spatial / temporal fixed effects. Specifically, unbiasedness is achieved if the *true cancellation effect* (line two in (12)) equals the expectation of the *empirical cancellation effect* (line three in (12)), or, alternatively, if both cancellation effects are individually zero. As before, this only needs to hold for the sample at large, and will be sensitive to the choice of M as we show below.

For their case without fixed effects, Abadie and Imbens (2011) argue that even if the auxiliary regression function is mis-specified, i.e. $E\left(\hat{\mu}_0\left(\mathbf{x}_r,\boldsymbol{\beta}\right)\neq\mu_0\left(\mathbf{x}_r,\boldsymbol{\beta}\right)\right)\ r=i,j$, the bias of the BCME will be small assuming $\mathbf{x}_i$ and $\mathbf{x}_j$ are reasonably close for most matched pairs, since both the true difference $\mu_0\left(\mathbf{x}_i,\boldsymbol{\beta}\right)-\mu_0\left(\mathbf{x}_j,\boldsymbol{\beta}\right)$ and the expectation of the estimated (empirical) difference $\hat{\mu}_0\left(\mathbf{x}_i,\boldsymbol{\beta}\right)-\hat{\mu}_0\left(\mathbf{x}_j,\boldsymbol{\beta}\right)$ will each be small, de facto reducing the bias to a difference of minor differences.

A similar argument holds with respect to the cancellation effect: Mis-specification of $\hat{\mu}\left(.\right)$, and thus presumably biased estimates of $\hat{\nu}$ and $\hat{\tau}$ do not preempt the empirical cancellation effect and thus its expectation from being (close to) zero (last line of (12)). If $E\left(\hat{\mu}\left(.\right)\right)$ is close to the true regression function $\mu\left(.\right)$, a near-zero empirical cancellation effect will imply the same for the true cancellation effect (line two in (12)). Even under strong mis-specification, there is still the theoretical chance that true and empirical cancellation effect are of similar magnitude, reducing bias along that route. In essence, *the analyst can do no better* in adjusted matching with fixed effects than find the setting of M that produces the empirical cancellation effect of smallest absolute magnitude.

We note that the true cancellation effect in line two of (12) is identical to the cancellation effect discussed above for the simple matching estimator (see equation 10), and the estimated cancellation effect is identical to the adjustment term that separates the simple from the adjusted version (see equation (11)). This implies that driving the empirical cancellation effect towards zero via choice of M de facto eliminates the adjustment term and leads to the convergence of the simple and adjusted estimator. In that sense, the main purpose of the regression adjustment is to identify the setting of M that minimizes the empirical cancellation effect, in absolute terms. The relationship between different settings of M and a possible cancellation effect is largely an empirical question, as we illustrate in our estimation below.

Bayesian estimation

We follow Johnston and Moeltner (2019) and estimate all regression models in a Bayesian framework. Compared to classical estimation, this brings the triple benefit of (i) non-reliance on asymptotic theory for inference, and (ii) directly obtaining a full finite sample distribution for all model parameters, and (iii) ease of generating full distributions for predictive constructs. Furthermore, since we are de facto dealing with a complete population of (relevant) sales, the classical notion of "increasing sample size" to invoke asymptotic properties and results is not particularly meaningful in our case. Instead, and in line with Bayesian paradigms, we focus exclusively on finite sample results and properties of estimators in our application.

Econometric details for Bayesian estimation of linear regression models (via Gibbs Sampling) are given in the online appendix of Johnston and Moeltner (2019).

Empirical application

Data and Descriptives

We obtained sales data from a commercial vendor for all single-family residential properties in Massachusetts (MA) sold between 1991 and 2014. We follow the exact same cleaning and data preparation steps as described in Johnston and Moeltner (2019) for their Connecticut sample (drop commercial properties, rentals, distressed and flip-sales, outliers, sales during hurricanes or major storms, etc.). To showcase our methodology, we focus on the adjacent Counties of Bristol and Plymouth in Southern MA. This geographic extent produces a sample size that is sufficiently large to allow for the estimation of rich regression models and offer an ample set of control homes for matching, while at the same time exhibiting relatively small sales counts for some spatial sub-zones. The latter makes it challenging to impose exact spatial matching, and thus highlights the practical relevance of the cancellation effect. Furthermore, this multi-county area corresponds well to typical spatial extents considered in existing housing market studies with focus on flood zone effects (Bin and Polasky 2004; Bin et al. 2008a,b; Kousky 2010; Atreya et al. 2013; Bin and Landry 2013; Zhang 2016; Johnston and Moeltner 2019).

As in Johnston and Moeltner (2019) our treatment effect of interest is the Special Flood Hazard Area (SFHA) status assigned by the U.S. Federal Emergency Management Agency (FEMA). To avoid difficult to control confoundedness of this effect with coastal amenities (see e.g. Troy and Romm 2004; Bin and Kruse 2006; Bin et al. 2008b; Johnston and Moeltner 2019), we restrict our sample to include only properties that are located at least one mile from the coastline. In that sense, our study can be interpreted as focusing on flood zone risks associated with interior flooding.

The final data set includes 113,358 sale observations corresponding to 81,157 properties. These units are located in 35 different school zones (SZ's), and sold between 1991 and 2014. Thus, our data comprises 24 different time periods (sale years, SY's). In total, 1,640 (1.45%) of sales involving 1,165 (1.44%) properties are located in an SFHA.

Table1 gives sample counts for treated and initially available controls for each SZ and SY, respectively. Primarily, this table indicates the upper bounds on the number of matched neighbors M when matching is forced to occur in the same SZ or SY. Specifically, the lowest number of controls for SZs is in area 9, with 125 observations. This caps M at that value for the whole sample under forced SZ matching.[5] In contrast, the lowest unit count for SY equals 977 for time period 24, which affords a much higher ceiling on M for forced matching over time units.

[5] In practice, forced matching is implemented via a large penalty term in variance matrix $\mathbf{V}$ in (7) for the variable at hand, here SZ or SY indicators. If a given treated observation "runs out" of controls to match on for a given SZ, this penalty will artificially inflate the distance statistic and muddle its interpretation as "distance along observable housing characteristics."

Table 2 shows basic property and neighborhood statistics (mean and standard deviation), grouped by SFHA status, and prior to any matching. The last two columns give statistics for the sample at large. As is evident from the table, control observations fetch, on average, a substantially higher sale price than treated, but the two groups are otherwise quite similar along many variable dimensions. The most pronounced differences in attributes are found for age, with treated being older, and, not surprisingly, elevation, with treated showing distinctly lower average elevation than controls. Controls are also slightly larger in terms of structural square footage and the number of bedrooms and bathrooms. In contrast, SFHA sales are associated with slightly larger overall lot size.

The second-to-last row of the table shows proportions of sales that occurred in communities that participated in FEMA's Community Rating System (CRS), which rewards participants with flood insurance premiums in exchange for taking voluntary measures to reduce flood risk beyond mandatory National Flood Insurance Program (NFIP) requirements.[6] The last row in the table captures the proportion of sales corresponding to a property that is located within 0.05 miles (264 feet) of a river, lake, or other body of water. This binary indicator is designed to flag properties that may have private water access, or views of water areas, in the spirit of Bin et al. (2008a). Not surprisingly, a substantially larger proportion of SFHA sales fall within this "water front" perimeter compared to controls (0.56 vs. 0.09).

Our matching variables ($\mathbf{x}$ in equation (7)) comprise all cardinal / continuous variables from Table 2, that is all variables except for the binary indicators for CRS membership and water front location. This holds for all models considered in this study, including the pre-balanced linear regression. The auxiliary regression model for the adjusted matching estimator includes all variables in Table 2, plus a constant term and a full set of binary indicators for school zones and sale years (minus a respective baseline category). This also holds for the generic linear regression model in (3), which, of course, also adds a treatment indicator.

[6]Further details in this program are given in Kousky (2017).

Matching statistics

We subject all three estimation frameworks to an increasing number of matches (= "nearest neighbors"), denoted as M above, ranging from one to 200 in increasing increments. Typically, the matching literature has used one to four matches, in an attempt to reduce bias due to an increasing mis-alignment between treated and matched controls along observable dimensions (e.g. Abbott and Klaiber 2013; Muehlenbachs et al. 2015; Johnston and Moeltner 2019).[7] As discussed above, in the presence of fixed effects increasing M beyond standard levels may be beneficial, as it can activate bias-reducing cancellation effects. As an added bonus, a larger M increases the sample size for the regression models we consider (primary or auxiliary), and can thus be expected to bring efficiency gains for estimated treatment effects.

Table 3 shows sample sizes for treated, original controls ("oc"), and unique controls after matching ("muc"), that is matched controls without counting repeated use of the same control as a match for multiple treated, as is allowed in our matching procedure described above. The table also shows the proportion of matched pairs that end up in the same school zone and / or the same sale year. These proportions, in turn, are captured for three variants of each estimation framework: unforced matching (controls can come from any school zone or sale year), forced matching on school zones (controls have to be from the same school zone, but can come from any sale year), and forced matching on sale year (controls have to be from the same sale year, but can come from any school zone). As is evident from the first four columns of the table, our data originally includes 1,640 treated observations (sales of homes that are located in an SFHA), and 111,718 control observations. At one-to-one matching ($M = 1$) we only retain a very small subset of these controls, that is 970 unique observations. This number grows rapidly as we increase M, up to 52,012 at $M = 200$.

The second triplet of columns gives the proportion of matched pairs that share the same

[7] Abadie and Imbens (2011) consider between 1 and 64 matches, and observe that bias generally increases for the BCME, but efficiency as measured via Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) improves over this range. However, their application is based on a much smaller sample of labor market participants, and their "true" treatment effect benchmark of training impact stems from a parallel estimation of a controlled experimental version of the same application. Most importantly the variables used in matching and in their auxiliary regression are identical, thus preempting any potential bias reduction via a cancellation effect.

SZ, SY, or both, in absence of any forced matching along these dimensions. As is evident from the "SZ" column, the vast majority of matches is SZ-consistent at $M = 1$. In contrast, very few pairs share the same sale year, as shown in the "SY" column. Naturally, even fewer matches pair up on both SY and SZ, as can be seen from the "both" column. Thus, under unforced matching, we can already conclude that unbiasedness of the basic ATT estimator under unforced matching ($\hat{\gamma}_B$) will have to rest on a true cancellation effect, as temporal effects differ for most matched pairs, and will thus not cancel out of equation (10).

As is evident from the next triplet of columns in the table, all matched pairs share the same SZ when we enforce this rule, with essentially no changes on the SY side compared to the unrestricted case. In contrast, the proportion of matched pairs that share the same SZ decreases noticeably compared to the unrestricted case when we force matches to occur in the same sale year, as can be seen from the last triplet of columns. This tradeoff effect of decreased shared proportions of SZs is exacerbated at higher M. As discussed above, forced matching on SZ's is capped at 125, preempting meaningful results for the last two rows for the "same SZ" columns.

In sum, the basic ATT $\hat{\gamma}_B$ will be biased by definition in all three unforced/forced cases, in absence of any cancellation effects. For the regression-adjusted estimator $\hat{\gamma}_A$ we note that in the unrestricted case correct estimation of SY effects at lower M will be more important than correct estimation of SZ effects, as those are already shared for most matches and thus (largely) drop out of the bias expression in (12). However, at larger M this proportion drops to 70-80%, placing more burden on correct estimation of SZ effects as well. Furthermore, forcing matched SZs brings relatively little gain in terms of bias relief, while forcing matched SYs flips the burden towards a need for higher accuracy in correctly estimating SZ effects. However, all these considerations ignore the possibility of a bias-reducing cancellation effect. Perfect equality of spatial or temporal effects between treated and matched controls, and/or correct specification of the true regression function may not be needed if the equality of true and empirical cancellation effects holds to a reasonable degree, or both are individually close to zero, as discussed above.

Balance and distance statistics

Table 4 gives balance statistics at increasing M, again separated by unforced matching, matching with forced shared SZ's and SY's respectively. For each case, we first compare distribution percentiles between treated and controls for a given variable that is used in the matching process, and determine the mean and maximum percentile deviation. The mean over all matching variables of these individual means is then entered in the respective "mean" column in the table. We proceed in analogous fashion for the "max" columns with respect to variable-specific and overall maximum percentile deviations. The closer these values are to zero, the more aligned are the distributions of treated and controls across matching variables (our $\mathbf{x}$ above), producing a higher degree of balance.

We first note that in absence of matching (first set of columns), mean percentile deviation amounts to 16.66 and maximum deviation equals over 238. This vastly improves with matching for all three "forcing" scenarios. Specifically, mean balance decreases to a range of 2.7 to 7.9 over increasing M for the unforced case, with maxima in the 26 to 42 range. This remains essentially unchanged when we force matching by SZ, as was expected given that most matched pairs already share the same SZ in the unforced scenario. Forcing matches to within the same SY, in contrast, leads to a clear deterioration in balance at any M, as is evident from the last set of columns. Overall, as expected, the sample becomes slightly less balanced as M increases, since an ever increasing cohort of controls are recruited into the model, with increasing disparity in the $\mathbf{x}$ dimension compared to their respective treated targets.

For the regression approach with treatment indicator, the statistics in Table 4 imply that a correct specification of the regression function $\mu_0(.)$ will be critical since balance is never close to zero for any of the matching scenarios or M values, thus correlating the treatment effect with other covariates. In other words, unbiased estimation of $\hat{\gamma}_R$ will *not* be robust to a mis-specification of the regression function.

Distance statistics, as computed via (7), are captured in Table 5. For each matching scenario, the table gives the minimum, mean, and maximum distance over all matched pairs. Essentially these results mirror those for balance, in that distance increases with M for all cases, remain largely unchanged for the "same SZ" scenario compared to the free matching case at any M, and are categorically higher for the "same SY" approach at a given M compared to the other two scenarios. Given that these distances are different from zero for all observations, and can take rather large values for some, we know that $\mathbf{x}_i \neq \mathbf{x}_j$ for every matched pair. This adds further bias to the basic matching estimator (in absence of any cancellation effects), and invalidates the first avenue towards unbiasedness for the BCME. Instead, unbiasedness of $\hat{\gamma}_A$ now needs to rest on unbiasedness the regression function $\hat{\mu}_0(.)$ in (12), in addition to unbiased estimation of the spatial and (especially) temporal fixed effects as discussed above, or on a potential cancellation effect.

Estimation results: Regression approach

We first illustrate results for a simple linear regression approach, akin to typical hedonic price regressions in the literature. Our regression model includes a constant term, all attribute variables listed in Table 1, plus a full set of binary indicators for school zones and sale years (minus a respective baseline category), as modeled in (3). In terms of Bayesian settings, we use extremely vague priors for all model coefficients (mean = 0, variance = 10^9), and a (typical) flat inverse gamma prior for the error variance (shape = scale = 0.5). This places the preponderance of inferential weight on the actual data, as is desired in absence of any specific prior information for model parameters. We discard the first 10,000 draws of the Bayesian simulator (Gibbs Sampler) as "burn-ins" to diminish any lingering effects of starting values, and retain the following 10,000 draws for inference.

Linear regression results are given in Table (6). Columns two through four repeat the sample sizes for treated, original controls ("oc") and unique matched controls ("muc") from Table 3. Overall, therefore, the available total sample size for the regression models *with pre-matching* increases from 2,610 (1,640 treated plus 970 controls) at $M = 1$ to 53,652 (1,640 treated plus 52,012 controls) for $M = 200$. Naturally, the unmatched regression still builds on the full data of 113,358 observations (1,640 treated plus 111,718 controls).

The next two quadruplets of columns show, respectively, Bayesian estimation results for the treatment effect γ_R for the unmatched, or "generic" regression and regression with pre-matching. In each case, the four columns capture the lower ("low") and upper ("high") bound of a 95% Highest Posterior Density Interval (HPDI), as well as the posterior mean and the HPDI range (upper minus lower). The HPDI gives the narrowest possible set of bounds that contain 95% of the posterior distribution.[8]

As is evident from the table, the generic regression estimates a counterintuitively positive average SFHA effect (in \$1000's), with a wide HPDI that includes zero. The mean treatment effect is also positive for the pre-matched case with 1:1 matching, but turns negative with increasing M. However, while HPDIs become generally more narrow with larger M, as one would expect under increasing sample size, the upper bound remains in positive territory throughout. In other words, the posterior distribution, at any M, assigns non-trivial probability of a positive flood zone effect, which is counter to theory and intuition. This casts doubt on the correct specification of the regression model, which, as shown above, is required for unbiasedness since perfect balance on $\mathbf{x}$, and thus independence of the treatment effect γ_R is never achieved, at any M. In contrast, it is likely that the increasing sample size at higher M helps with more accurate estimation of the SZ and SY effects, which may be the reason the posterior distribution moves into more negative (and thus more plausible) territory with a higher number of matches. We stress again that cancellation effects are absent in the generic regression framework, as its treatment estimator does not explicitly build on observable differences between matched pairs.

Estimation results: Basic Matching estimator

Table 7 shows results for the basic matching estimator given in (9). Column one repeats the number of matched controls per treated (M), while columns two through six show the average sale price for treated, original controls ("oc"), and unique matched controls ("muc"). The latter prices are given for all three matching scenarios, unforced, forced matching on school zones ("same SZ"), and forced matching on sale year ("same SY"). We note from

[8] As noted in Moeltner et al. (2017), while conceptually and computationally different, the 95% HPDI can be interpreted akin to a classical confidence interval, in the sense that there is 95% chance that the (presumably fixed but unknown) treatment effect lies between these bounds.

the "muc" columns that the average price of a control home increases with M in all three scenarios. This suggests that an increasing number of relatively higher control sales are recruited into the model as we boost the number of required matches per treated.

The following three columns give the sample-averaged counterfactual price

$$\bar{\hat{y}}_0 = \frac{1}{n_1} \sum_{i=1}^{n_1} \left(\frac{1}{L_{i,M}} \sum_{j \in \mathcal{J}_{i,M}} y_{j0} \right)$$

from equation (9). For all three scenarios this value increases at a comparable pace with higher M, though the counterfactual price for the "same SY" case starts at a much higher level than for the other two scenarios.

As a result, the treatment effect (or basic ATT), which corresponds to the price of treated minus the respective counterfactual price for the sample, decreases over increasing M for all three matching scenarios, as is evident from the last three columns of the table. However, it remains in (counterintuitive) positive domain for the unforced case for lower values of M and hardly drops below zero at any M for the "same SZ" implementation. In stark contrast, the basic ATT estimate emerges as clearly negative even at $M = 1$ (loss of \$7,891) when matches are forced to occur in the same sale year. The estimated SFHA loss further increases up to close to \$20,000 at $M = 200$.

Recall that the basic matching approach incorporates no means of capturing spatial and temporal effects in any other fashion than via forced matching. Since matched sales are naturally inclined to share the same SZ (see Table 3), forcing matches on SY becomes especially important for invoking unbiasedness along the avenue of perfect matches. However, as is evident from Tables 4 and 5, forcing matching in any fashion leads to a more unbalance sample and larger matching distances compared to the unforced case. This is especially true for the "same SY" scenario, and under increasing M. In absence of any further information, therefore, an analyst might conclude that the most plausible / least biased version of the basic matching estimator in Table 7 is the "same SY" result at $M = 1$. However, as we show below, results for the adjusted matching estimator reveal that the unforced scenario at $M = 120$ actually produces the basic estimator with the lowest risk of bias due to the presence of cancellation effects, which are highlighted in the regression-adjusted case.

Estimation results: Adjusted Matching estimator

Results for the adjusted matching estimator $\hat{\gamma}_A$ in (11) are shown in Table 8. As before, we present the posterior mean, 95% HPDI bounds, and the HPDI range for all three matching scenarios. We first note that HPDI ranges for all three implementations are of comparable magnitude and decrease notably with higher M, thus indicating clear efficiency gains at higher sample sizes. At $M = 20$ and beyond the full posterior distribution of $\hat{\gamma}_A$ is unambiguously located in the negative domain for both the unforced and "same SZ" scenario, as would be dictated by theory and intuition. In contrast, posterior distributions are fully negative at any M for the "same SY" case.

To examine if cancellation effects may be the drivers of these generally more reasonable results compared to the other two estimators, we break the empirical cancellation effect in the last line of (12) into its individual components:

$$
\begin{aligned}
(13)\quad & \frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{1}{L_{i,M}}\sum_{j\in\mathcal{J}_{i,M}}\left(\left(\hat{\mu}_0\left(\mathbf{x}_i,\boldsymbol{\beta}\right)-\hat{\mu}_0\left(\mathbf{x}_j,\boldsymbol{\beta}\right)\right)+\left(\left(\hat{\nu}_i+\hat{\tau}_i\right)-\left(\hat{\nu}_j+\hat{\tau}_j\right)\right)\right)\right)= \\
& \frac{1}{n_1}\sum_{i=1}^{n_1}\hat{\mu}_0\left(\mathbf{x}_i,\boldsymbol{\beta}\right)-\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{1}{L_{i,M}}\sum_{j\in\mathcal{J}_{i,M}}\hat{\mu}_0\left(\mathbf{x}_j,\boldsymbol{\beta}\right)\right)+ \\
& \frac{1}{n_1}\sum_{i=1}^{n_1}\left(\hat{\nu}_i+\hat{\tau}_i\right)-\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{1}{L_{i,M}}\sum_{j\in\mathcal{J}_{i,M}}\left(\hat{\nu}_j+\hat{\tau}_j\right)\right)
\end{aligned}
$$

Table 9 provides an overview of these adjustment terms for unforced matching (first block of rows), forced matching on SZ's (second block of rows), and forced matching on SY's (third block of rows). Each block, in turn, is divided into three sets of columns. The first set gives the portion of predicted price solely related to observable characteristics for the treated ("T"), i.e. the first term in the second line of (13), the analogous portion for matched controls ("C"), i.e. the second term in the second line of (13), and their difference. The second set provides corresponding results for the fixed effects, as captured in the last line of (13). The last column in each block gives the total adjustment, i.e. the full empirical cancellation effect (first line of (13)).

As is evident from the table the adjustment components based on observables $\mathbf{x}$ are by an order of magnitude larger than the adjustment portions corresponding to the spatial and temporal fixed effects, for all matching scenarios and settings of M. That said, the corresponding respective *adjustment differences* are of comparable magnitude, generally lying in the single-digit range (in terms of \$ 1000's). It is also clear from the table that $\mathbf{x}$-based adjustment and fixed effects adjustment can go in opposite directions, and all but cancel each other. The lowest total adjustment (0.002, i.e. \$2 in nominal terms) is observed for unforced matching at $M = 120$. We thus provide evidence that empirical cancellation effects close to zero can indeed exist.

In summary, our combined analytical results allow for the following conclusions:

1. Our balance analysis confirms that the treatment effect will be correlated with observables characteristics, under any matching scenario. This implies that the regression function $\mu_0(.)$ be properly specified to assure unbiased estimation in a standard hedonic framework.

2. Given our generic regression results, we can be fairy certain that $\hat{\mu_0}(.)$ is mis-specified. This rules out typical hedonic regression as a promising estimation framework. It also preempts bias reduction for the adjusted matching estimator via $E(\hat{\mu_0}(.)) = \mu_0(.)$ in (12).

3. The second avenue to bias reduction for both basic and adjusted matching estimator, via equality of observable characteristics plus common spatial and temporal effects for matched pairs, is equally unattainable, given our findings on balance and distance, and the fact that not all pairs share the same school zone, and hardly any share the same sales year.

4. Forced matching along one or the other dimension exacerbates discrepancies along observables based on our distance analysis. It also limits the scope for a possible cancellation effect. This is obvious from the last column of Table 9, which shows that the total

adjustment (i.e, the cancellation effect) is generally further from zero for the restricted cases, especially at higher M.

5. This leaves the cancellation route as the only viable option to reduce bias, and thus to select among competing specifications. We show that empirical cancellation effects can indeed be very small for certain choices of M, thus flagging models with the lowest risk of bias.

In our case, we would thus choose the unforced matching result at $M = 120$, and the corresponding treatment effect of -$4,532 (see Table 8). Since M is large, this brings the added bonus of a very tight posterior interval, with a range of less than $2,600. Since the empirical cancellation effect is identical to the regression adjustment, as discussed above, the adjusted and unadjusted matching estimators (-$4,530 in Table 7) converge, stressing again the reduced dependence on regression function specification when the cancellation effects are minimized through a targeted choice of M.[9]

Our preferred estimate of an average interior flood zone discount of $4,530, or 1.45% off the price of the average home across the entire sample, corresponds well to findings reported in Johnston and Moeltner (2019), who use a similar data set for Connecticut. Specifically, they estimate a 1.8% loss for interior SFHA properties located at least one mile from the coast. Our mean flood zone estimate is also somewhat lower than those reported in Bin and Polasky (2004), Kousky (2010), Bin and Landry (2013), and Zhang (2016), who estimate interior SFHA effects in the 3-6% range. However, none of the latter set of studies employ any form of matching in their analysis.

Conclusion

We illustrate how results from hedonic regression analysis and non-/semi-parametric matching, along with related data statistics such as balance and distance, can be combined to help identify models with comparatively low risk of estimation bias for binary treatment

[9] Since the total adjustment switches sign for the "same SY" case between $M = 2$ and $M = 4$, we also estimated a same-SY model for $M = 3$. The adjustment for this setting amounts to -2.689, and thus clearly exceeds, in absolute terms, the corresponding adjustment value for our preferred unforced model at $M = 120$.

effects. Specifically, we show that in the presence of spatial and / or temporal fixed effects a beneficial cancellation effect can result when forced matching along all unobserved dimensions is counter-productive. We further demonstrate that this cancellation effect is sensitive to the choice of the number of matched controls in a nearest-neighbor context. This, in turn, maps out an empirical route for the identification of low-bias models, via inspection of this cancellation effect at varying settings M, and selecting specifications that minimize this effect in absolute terms.

Our maintained assumption of unobservable spatial and temporal effects are, arguably, the norm in hedonic applications. In the same vein, forcing exact matching on space and time will generally be infeasible in typically-sized data sets. Even if this is possible based on sample counts, forced matching always implies a deterioration in matching quality on observables. Thus, pursuing the route of estimating cancellation effects and using them as guidance for model selection will be an attractive proposition in most housing market applications.

While the analysis presented in this study uses a simple cross-sectional hedonic model as example, our arguments translate in straightforward fashion to more sophisticated hedonic regression models, such as those building on repeat sales (e.g. Netusil et al. 2019), difference-in-differences (e.g. Brolinson et al. 2023), triple-differences (e.g. Muehlenbachs et al. 2015), or panel data structures (e.g. Kuminoff et al. 2010). As long as the final specification includes observed housing and landscape characteristics, differenced or otherwise, and fixed effects to control for spatial and / or temporal unobservables, our general framework applies.

We leave it to further study to determine the real-world underpinnings that can lead to a near-zero cancellation effect. In our case, for example, it might be that larger homes with higher-priced structural and landscape characteristics might be located in lower-performing school zones, or school zones with higher traffic volume or crime rates, for the sample at large. While a deeper understanding of these cancellation effects is certainly worth exploring, it is not necessary to implement our analytical framework and decision rule for model choice.

References

Abadie, A., Drucker, D., Herr, J., Imbens, G., 2002. Simple and bias-corrected matching estimators for average treatment effects. NBER Technical Working paper 283.

Abadie, A., Drucker, D., Herr, J., Imbens, G., 2004. Implementing matching estimators for average treatment effects in stata. Journal of Business & Economic Statistics 4, 290–311.

Abadie, A., Imbens, G., 2006. Large-sample properties of matching estimators for average treatment effects. Econometrica 74, 235–267.

Abadie, A., Imbens, G., 2011. Bias-corrected matching estimators for average treatment effects. Journal of Business & Economic Statistics 29, 1–11.

Abbott, J., Klaiber, H., 2013. The value of water as an urban club good: A matching approach to community-provided lakes. Journal of Environmental Economics and Management 65, 208–224.

Atreya, A., Ferreira, S., Kriesel, W., 2013. Forgetting the flood? An analysis of the flood risk discount over time. Land Economics 89, 577–596.

Bin, O., Brown Kruse, J., Landry, C., 2008a. Flood hazards, insurance rates, and amenities: Evidence from the coastal housing market. Journal of Risk and Insurance 75, 63–82.

Bin, O., Crawford, T., Kruse, J., Landry, C., 2008b. Viewscapes and flood hazards: Coastal housing market response to amenities and risk. Land Economics 84, 434–448.

Bin, O., Kruse, J., 2006. Real estate market response to coastal flood hazards. Natural Hazards Review 7, 137–144.

Bin, O., Landry, C., 2013. Changes in implicit flood risk premiums: Empirical evidence from the housing market. Journal of Environmental Economics and Management 65, 361–376.

Bin, O., Polasky, S., 2004. Effects of flood hazards on property values: Evidence before and after hurricane Floyd. Land Economics 80, 490–500.

Bishop, K.C., Kuminoff, N.V., Banzhaf, H.S., Boyle, K.J., von Gravenitz, K., Pope, J.C., Smith, V.K., Timmins, C.D., 2020. Best practices for using hedonic property value models to measure willingness to pay for environmental quality. Review of Environmental Economics and Policy .

Brolinson, B., Palmer, K., Wals, M., 2023. Does energy star certification reduce energy use in commercial buildings? Journal of the Association of Environmental and Resource Economists 10, 55–93.

Ho, D., Kosuke, I., King, G., Stuart, E., 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Policitcal Analysis 15, 199–236.

Imbens, G., 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. The Review of Economics and Statistics 86, 4–29.

Imbens, G., Wooldridge, J., 2009. Recent developments in the econometrics of program evaluation. Journal of Economic Literature 47, 5–86.

Johnston, R., Moeltner, K., 2019. Special flood hazard effects on coastal and interior home values: One size does not fit all. Environmental and Resource Economics 74, 181–210.

Kousky, C., 2010. Learning from extreme events: Risk perceptions after the flood. Land Economics 86, 395–422.

Kousky, C., 2017. Financing flood losses. Resources for the Future Discussion Paper #17-03, February, 2017.

Kuminoff, N., Parmeter, C., Pope, J., 2010. Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities? Journal of Environmental Economics and Management 60, 145–60.

Manning, D., Ando, A., 2022. Ecosystem services and land rental markets: Producer costs of bat population crashes. Journal of the Association of Environmental Economics and Management 9, 1235–1277.

Mei, Y., Gao, L., Zhang, W., Yang, F., 2021. Do homeowners benefit when coal-fired power plants switch to natural gas? Evidence from Beijing, China. Journal of Environmental Economics and Management 110, 102566.

Meldrum, J., 2016. Floodplain price impacts by property type in boulder county, colorado: Condominiums versus standalone properties. Environmental and Resource Economics 64, 725–750.

Moeltner, K., Blinn, C., Holmes, T., 2017. Forest pests and home values: The importance of accuracy in damage assessment and geocoding of properties. Forest Economics 41, 89–109.

Muehlenbachs, L., Spiller, E., Timmins, C., 2015. The housing market impacts of shale gas developments. American Economic Review 105, 3633–59.

Netusil, N., Jarrad, M., Moeltner, K., 2019. Floodplain designation and property sale price in an urban watershed. Land Use Policy 88, 104112.

Troy, A., Romm, J., 2004. Assessing the price effects of flood hazard disclosure under the California natural hazard disclosure law. Journal of Environmental Planning and Management 47, 137–162.

Wolf, D., Takeuchi, K., 2022. Holding back the storm: Dam capitalization in residential and commercial property values. Journal of Environmental Economics and Management 116, 102737.

Wu, J., Yu, J., Oueslati, W., 2023. Open space in US urban areas: Where might there be t much or too little of a good thing? Journal of the Association of Environmental and Resource Economists 10, 315–352.

Zhang, L., 2016. Flood hazards impact on neighborhood housing prices: Aspatial quantile regression analysis. Regional Science and Urbam Economics 60, 12–19.

Table 1: Sample counts for school zones and sale years

SZ	treated	controls	SY	treated	controls
1	9	3004	1	16	1613
2	170	8286	2	69	4961
3	52	1502	3	68	5412
4	49	3808	4	99	5704
5	1	2167	5	66	4965
6	155	3532	6	66	5511
7	110	3746	7	80	5767
8	47	2107	8	83	6670
9	3	125	9	98	6418
10	90	4842	10	111	6374
11	4	4593	11	96	6118
12	38	2981	12	101	6112
13	19	3306	13	89	5912
14	35	2503	14	114	6113
15	24	2188	15	74	5064
16	21	5299	16	63	4234
17	34	144	17	60	3858
18	44	3207	18	50	3103
19	1	154	19	42	3325
20	16	4457	20	43	3072
21	22	6097	21	41	2947
22	82	6002	22	39	3423
23	34	1791	23	53	4065
24	8	2733	24	19	977
25	7	2061			
26	16	6300	Total	1640	111718
27	10	1151			
28	5	3221			
29	98	3615			
30	75	2061			
31	78	8017			
32	236	2357			
33	43	1116			
34	1	421			
35	3	2824			
Total	1640	111718			

SZ = school zone
SY = sale year
controls = originally available controls before matching

Table 2: Basic sample statistics

	non-SFHA		SFHA		all	
variable	mean	sd	mean	sd	mean	sd
price($1000s)	306.68	172.22	274.52	153.84	306.21	172.01
age	34.38	39.36	45.27	47.28	34.54	39.51
square footage (00s)	33.86	16.21	28.58	15.01	33.79	16.20
lot size, sqft (000s)	40.46	79.04	44.06	107.45	40.52	79.52
bedrooms	3.19	0.72	2.96	0.83	3.19	0.72
bathrooms	1.64	0.69	1.56	0.67	1.64	0.69
elevation (m)	34.22	15.34	18.57	16.18	33.99	15.47
distance (miles) to nearest:						
beach	7.34	4.44	6.14	4.90	7.33	4.45
coast or estuary	8.61	5.52	7.77	5.72	8.60	5.52
reservoir or lake	0.27	0.18	0.20	0.16	0.27	0.18
river	0.23	0.16	0.08	0.10	0.23	0.16
interstate	5.00	4.31	4.16	4.16	4.99	4.31
principal artery	1.22	1.00	1.49	1.06	1.22	1.00
high-density development	1.24	1.10	1.19	1.06	1.24	1.10
land cover w/in 1 km radius:						
agricultural land, acres	43.34	51.58	38.05	41.14	43.26	51.44
industrial land, acres	35.47	66.52	28.56	40.59	35.37	66.22
open land, acres	439.29	164.34	415.19	153.47	438.94	164.21
CRS membership, time of sale	0.07	0.26	0.03	0.17	0.07	0.26

SFHA = Special Flood Hazard Area
sd = standard deviation

Table 3: Matching results under increasing number of matches

				no forced matching proportion of matches in same:			forced same SZ proportion of matches in same:			forced same SY proportion of matches in same:		
M	treated	oc	muc	SY	SZ	both	SY	SZ	both	SY	SZ	both
1	1640	111718	970	0.074	0.909	0.068	0.070	1.000	0.070	1.000	0.777	0.777
2	1640	111718	1787	0.069	0.898	0.064	0.066	1.000	0.066	1.000	0.735	0.735
4	1640	111718	3141	0.066	0.883	0.060	0.065	1.000	0.065	1.000	0.690	0.690
6	1640	111718	4367	0.065	0.871	0.058	0.064	1.000	0.064	1.000	0.660	0.660
8	1640	111718	5518	0.063	0.862	0.056	0.062	1.000	0.062	1.000	0.636	0.636
10	1640	111718	6633	0.060	0.851	0.053	0.061	1.000	0.061	1.000	0.618	0.618
20	1640	111718	11459	0.057	0.811	0.048	0.057	1.000	0.057	1.000	0.554	0.554
40	1640	111718	19308	0.055	0.761	0.043	0.055	1.000	0.055	1.000	0.484	0.484
60	1640	111718	25640	0.052	0.728	0.040	0.053	1.000	0.053	1.000	0.438	0.438
80	1640	111718	30986	0.052	0.703	0.038	0.052	1.000	0.052	1.000	0.405	0.405
100	1640	111718	35646	0.051	0.684	0.036	0.051	1.000	0.051	1.000	0.380	0.380
120	1640	111718	39686	0.051	0.668	0.035	0.051	1.000	0.051	1.000	0.359	0.359
150	1640	111718	44902	0.051	0.649	0.034	-	-	-	1.000	0.332	0.332
200	1640	111718	52012	0.050	0.623	0.032	-	-	-	1.000	0.299	0.299

M = number of matches ("nearest neighbors")
oc = number of original controls / muc = number of unique matched controls
SZ = school zone / SY = sale year

Table 4: Balance statistics under increasing number of matches

	before matching		after, unforced		after, same SZ		after, same SY	
M	mean	max	mean	max	mean	max	mean	max
1	16.660	238.692	2.708	26.122	2.795	27.883	8.078	129.124
2	16.660	238.692	3.043	25.147	3.039	23.741	8.410	127.714
4	16.660	238.692	3.391	25.234	3.444	24.122	8.956	128.789
6	16.660	238.692	3.913	25.369	3.835	24.673	9.114	127.447
8	16.660	238.692	4.198	25.749	4.058	25.058	9.522	131.017
10	16.660	238.692	4.479	25.959	4.298	24.907	9.780	132.497
20	16.660	238.692	5.636	29.274	5.516	30.921	10.332	133.264
40	16.660	238.692	6.597	35.231	6.436	35.754	10.951	136.089
60	16.660	238.692	7.046	36.062	6.951	37.813	11.710	147.198
80	16.660	238.692	7.363	37.696	7.244	38.639	11.955	148.410
100	16.660	238.692	7.601	39.115	7.530	39.670	12.031	148.465
120	16.660	238.692	7.697	39.226	7.677	40.240	12.094	148.674
150	16.660	238.692	7.854	41.381	-	-	12.209	148.865
200	16.660	238.692	7.900	41.667	-	-	12.250	149.126

mean = mean over all matching variables of: mean percentile deviation between treated and controls for each variable
max = maximum over all matching variables of: maximum percentile deviation between treated and controls for each variable
SZ = school zone / SY = sale year

Table 5: Distance statistics under increasing number of matches

	unforced			same SZ			same SY		
M	min	mean	max	min	mean	max	min	mean	max
1	0.080	1.213	12.832	0.080	1.264	12.865	0.080	1.955	27.658
2	0.098	1.295	12.846	0.098	1.350	12.891	0.163	2.119	30.008
4	0.255	1.415	12.859	0.257	1.479	13.012	0.456	2.326	31.743
6	0.277	1.501	12.943	0.280	1.572	14.097	0.530	2.463	32.576
8	0.294	1.568	13.431	0.297	1.646	15.904	0.559	2.566	33.346
10	0.306	1.624	14.248	0.310	1.706	17.245	0.582	2.650	33.999
20	0.365	1.818	17.566	0.369	1.920	21.220	0.764	2.929	35.892
40	0.451	2.037	21.646	0.455	2.171	27.046	1.089	3.239	37.713
60	0.545	2.175	24.161	0.551	2.334	30.140	1.301	3.437	38.817
80	0.607	2.276	25.964	0.614	2.456	31.992	1.461	3.585	39.590
100	0.654	2.356	27.418	0.661	2.556	33.325	1.593	3.703	40.195
120	0.696	2.422	28.539	0.704	2.641	34.376	1.711	3.804	40.708
150	0.756	2.506	29.937	-	-	-	1.862	3.931	41.314
200	0.862	2.615	31.572	-	-	-	2.086	4.103	42.021

min = minimum Euclidean distance over all matched pairs
mean = mean Euclidean distance over all matched pairs
max = maximum Euclidean distance over all matched pairs
SZ = school zone / SY = sale year

Table 6: Linear regression results for treatment effect ($1000's)

	sample size			unmatched				matched			
M	treated	oc	muc	low	mean	high	range	low	mean	high	range
unmatched	1640	111718	-	-3.737	1.926	7.314	11.051	-	-	-	-
1	1640	111718	970	-	-	-	-	-4.247	4.660	13.222	17.469
2	1640	111718	1787	-	-	-	-	-1.825	5.744	13.307	15.132
4	1640	111718	3141	-	-	-	-	-7.473	-0.462	6.221	13.694
6	1640	111718	4367	-	-	-	-	-6.281	0.435	6.812	13.093
8	1640	111718	5518	-	-	-	-	-7.051	-0.429	5.873	12.924
10	1640	111718	6633	-	-	-	-	-8.195	-1.444	4.797	12.992
20	1640	111718	11459	-	-	-	-	-9.382	-3.199	2.994	12.376
40	1640	111718	19308	-	-	-	-	-9.142	-3.079	2.994	12.136
60	1640	111718	25640	-	-	-	-	-8.520	-2.656	3.257	11.777
80	1640	111718	30986	-	-	-	-	-7.802	-2.269	3.821	11.623
100	1640	111718	35646	-	-	-	-	-7.954	-2.115	3.835	11.789
120	1640	111718	39686	-	-	-	-	-9.120	-2.815	2.660	11.780
150	1640	111718	44902	-	-	-	-	-8.129	-2.354	3.525	11.654
200	1640	111718	52012	-	-	-	-	-7.802	-1.659	3.988	11.790

mean = posterior mean
low [high] = lower [upper] bound of 95% highest posterior density interval

Table 7: Estimation results for basic matching estimator (1000$'s)

	price statistics (sample means)					counterfactual price (sample mean)			treatment effect		
			muc								
M	treated	oc	unforced	same SZ	same SY	unforced	same SZ	same SY	unforced	same SZ	same SY
1	274.515	306.677	277.026	272.506	286.163	261.848	256.884	282.406	12.668	17.631	-7.891
2	274.515	306.677	281.701	279.428	287.130	266.575	262.101	280.633	7.940	12.414	-6.117
4	274.515	306.677	290.932	288.124	293.052	272.461	268.224	283.633	2.055	6.291	-9.117
6	274.515	306.677	293.084	288.974	295.304	271.370	267.985	284.854	3.145	6.531	-10.670
8	274.515	306.677	293.971	290.368	296.705	270.641	267.581	284.961	3.874	6.934	-10.446
10	274.515	306.677	296.572	292.744	297.686	271.054	268.522	285.255	3.462	5.993	-10.740
20	274.515	306.677	303.864	301.046	300.778	275.177	272.105	286.580	-0.662	2.410	-12.065
40	274.515	306.677	308.144	304.910	303.648	277.336	274.014	288.982	-2.821	0.501	-14.467
60	274.515	306.677	310.626	307.289	304.639	278.048	274.553	290.376	-3.533	-0.038	-15.860
80	274.515	306.677	312.026	308.293	304.874	278.257	274.698	291.485	-3.742	-0.183	-16.970
100	274.515	306.677	313.409	309.361	304.929	278.427	274.915	292.209	-3.912	-0.400	-17.694
120	274.515	306.677	314.112	309.613	305.140	279.045	275.066	292.806	-4.530	-0.551	-18.291
150	274.515	306.677	314.513	-	305.368	279.590	-	293.613	-5.075	-	-19.098
200	274.515	306.677	314.609	-	305.862	280.290	-	294.467	-5.774	-	-19.952

oc = number of original controls / muc = number of unique matched controls
SZ = school zone / SY = sale year

Table 8: Estimation results for regression-adjusted matching estimator (1000$'s)

	unforced				same SZ				same SY			
M	low	mean	high	range	low	mean	high	range	low	mean	high	range
1	-3.027	2.405	7.399	10.426	2.406	7.456	12.643	10.237	-15.681	-9.948	-4.442	11.239
2	-1.634	2.240	6.238	7.872	0.771	4.679	8.667	7.896	-12.818	-7.971	-3.665	9.153
4	-5.873	-2.579	0.603	6.476	-3.639	-0.520	2.866	6.505	-11.085	-7.624	-4.126	6.959
6	-4.877	-1.998	0.823	5.700	-5.068	-2.101	0.738	5.806	-10.670	-7.698	-4.524	6.146
8	-4.333	-1.585	1.008	5.341	-4.217	-1.586	1.237	5.454	-9.483	-6.705	-3.970	5.513
10	-4.287	-1.898	0.660	4.947	-4.045	-1.511	0.935	4.980	-7.859	-5.289	-2.669	5.190
20	-6.424	-4.393	-2.400	4.024	-6.665	-4.583	-2.516	4.149	-5.770	-3.714	-1.668	4.102
40	-6.828	-5.158	-3.502	3.326	-6.607	-4.826	-3.117	3.490	-5.946	-4.272	-2.633	3.313
60	-6.388	-4.883	-3.365	3.023	-6.311	-4.676	-3.169	3.142	-5.718	-4.193	-2.718	3.000
80	-6.140	-4.696	-3.300	2.840	-6.294	-4.702	-3.343	2.951	-5.101	-3.650	-2.245	2.856
100	-5.735	-4.378	-3.081	2.654	-5.817	-4.368	-2.985	2.832	-5.120	-3.738	-2.401	2.719
120	-5.854	-4.532	-3.280	2.574	-5.855	-4.481	-3.160	2.695	-4.754	-3.439	-2.118	2.636
150	-5.765	-4.498	-3.249	2.516	-	-	-	-	-4.358	-3.077	-1.750	2.608
200	-5.568	-4.372	-3.163	2.405	-	-	-	-	-3.536	-2.316	-1.047	2.489

SZ = school zone / SY = sale year
mean = posterior mean
low [high] = lower [upper] bound of 95% highest posterior density interval

Table 9: Adjustment analysis

	housing characteristics portion			fixed effects portion			total
M	T	C	diff	T	C	diff	adjustment
				unforced			
1	226.172	222.614	3.558	45.929	39.225	6.704	10.262
2	265.604	261.911	3.693	6.649	4.642	2.007	5.700
4	289.622	285.079	4.543	-12.513	-12.603	0.090	4.633
6	263.513	259.079	4.434	13.014	12.306	0.708	5.142
8	263.751	259.301	4.450	12.371	11.362	1.009	5.459
10	281.160	276.174	4.986	-4.761	-5.134	0.373	5.359
20	273.511	268.792	4.719	5.396	6.384	-0.988	3.731
40	294.167	289.914	4.253	-14.479	-12.563	-1.916	2.337
60	299.283	295.637	3.646	-19.882	-17.586	-2.296	1.350
80	296.270	293.039	3.231	-17.065	-14.787	-2.278	0.953
100	293.579	290.709	2.870	-14.686	-12.283	-2.403	0.467
120	295.521	292.616	2.905	-16.467	-13.564	-2.903	0.002
150	291.162	288.439	2.723	-12.148	-8.848	-3.300	-0.577
200	288.286	285.973	2.313	-9.397	-5.682	-3.715	-1.402
				same SZ*			
1	240.949	238.382	2.567	26.103	18.494	7.609	10.176
2	264.234	261.466	2.768	5.579	0.612	4.967	7.735
4	266.315	262.547	3.768	8.736	5.692	3.044	6.812
6	243.413	238.509	4.904	33.218	29.490	3.728	8.632
8	241.247	236.989	4.258	34.876	30.614	4.262	8.520
10	262.690	258.679	4.011	13.314	9.821	3.493	7.504
20	264.678	260.171	4.507	14.437	11.951	2.486	6.993
40	278.132	275.220	2.912	1.213	-1.202	2.415	5.327
60	289.078	286.825	2.253	-9.887	-12.271	2.384	4.637
80	284.603	282.851	1.752	-5.381	-8.147	2.766	4.518
100	279.275	278.382	0.893	-0.388	-3.462	3.074	3.967
120	276.642	275.843	0.799	2.350	-0.782	3.132	3.931
				same SY			
1	248.639	241.053	7.586	35.811	41.340	-5.529	2.057
2	260.213	255.735	4.478	22.247	24.872	-2.625	1.853
4	260.430	257.719	2.711	21.698	25.902	-4.204	-1.493
6	254.602	251.156	3.446	27.624	33.710	-6.086	-2.640
8	247.213	244.688	2.525	34.027	40.293	-6.266	-3.741
10	242.183	240.544	1.639	37.612	44.702	-7.090	-5.451
20	252.224	252.932	-0.708	26.010	33.653	-7.643	-8.351
40	257.536	258.924	-1.388	21.252	30.059	-8.807	-10.195
60	258.905	261.206	-2.301	19.797	29.164	-9.367	-11.668
80	261.776	264.875	-3.099	16.383	26.604	-10.221	-13.320
100	264.341	267.653	-3.312	13.916	24.560	-10.644	-13.956
120	266.865	270.661	-3.796	11.085	22.142	-11.057	-14.853
150	270.809	275.212	-4.403	6.777	18.395	-11.618	-16.021
200	274.027	279.543	-5.516	2.803	14.923	-12.120	-17.636

X*b portion: portion of regression prediction based on X*b
fixed effects portion: portion of regression prediction based on SY and / or SZ effects
T = treated / C= controls / diff = difference between T and C
SZ = school zone / SY = sale year
*M=150 or 200 infeasible for SZ matching