Hits from the Bong: The Impact of Recreational Marijuana Dispensaries on Property Values*

Danna Thomas
University of South Carolina

Lin Tian
INSEAD & CEPR

March 31, 2020

Click here for the latest version

Abstract

We exploit a natural experiment in Washington state that randomly allocates recreational marijuana retail licenses to estimate the capitalization effects of dispensaries into property sale prices. Developing a new cross-validation procedure to define the treatment radius, we estimate difference-in-differences, triple difference, and instrumental variables models. We find statistically significant negative effects of recreational marijuana dispensaries on housing values that are relatively localized: home prices within a 0.36 mile area around a new dispensary fall by 3-4.5% on average. We also explore increased crime near dispensaries as a possible mechanism driving depressed home prices. While we find no evidence of a general increase in crime in Seattle, WA, we do find that nuisance-related crimes increase.

keywords: real estate markets, local externalities, drug legalization

---

*We thank Francois Gerard and Wojciech Kopczuk for continued guidance and support throughout this project. We are grateful to Treb Allen, Donald Davis, Jonathan Dingel, Francesc Ortega, Will Strange, and participants at the Applied Microeconomics Colloquium participants at Columbia University and the Annual Meeting of the Urban Economic Association for helpful discussion and comments. Thomas - email: danna.thomas@moore.sc.edu. Tian - email: lin.tian@insead.edu.

1 Introduction

Despite an increasing trend of recreational marijuana liberalization across the United States and in other parts of the world, legalizing cannabis continues to be a contentious issue.[1] While there is general consensus on the benefits of legalization, such as cannabis tax revenues and decreased incarceration for drug-related crimes, strong reservations remain on the *local-level* impacts of marijuana businesses to neighborhoods, evidenced by the widespread city-level restrictions of dispensaries as well as neighborhood resistance to dispensary entry within states that have legalized marijuana.[2] Understanding the local-level consequences of marijuana dispensaries on neighborhoods is, therefore, crucial for assessing the aggregate effects of legalization and designing effective public policies to address the localized impact of the legalization.

This paper studies local responses to marijuana dispensaries, exploring how dispensary entry is capitalized into local housing values. If residents perceive a nearby marijuana dispensary as a disamenity (or as an amenity), they can "vote with their feet;" hence, the opening of a cannabis retailer should lead to a decrease (or increase) in property values, reflecting the residents' willingness to pay to live away (or near) the retailer.[3]

However, the endogeneity of dispensary location creates a challenge in identifying the causal effects of cannabis retailers on neighborhood property values: There may be variables unobserved by the econometrician but observed by the cannabis retailers that are correlated with neighborhood outcomes. To overcome this identification problem, we exploit a natural experiment in Washington that randomly allocates recreational marijuana retail licenses to applicants. Following the 2012 legalization of recreational marijuana in Washington, cannabis license applicants were required to provide potential dispensary sites on their applications, and many retail licenses were allocated via lottery. This enables us to assemble a novel data set that connects license lottery winners, losers, and cannabis

[1] Caulkins et al. (2016) offers a comprehensive summary of the issues and much of the existing academic literature.

[2] Cities such as Pasco, WA and Compton, CA have banned recreational marijuana dispensaries. For an example of neighborhood resistance in San Francisco, California: "Residents fighting to keep marijuana dispensary out of sunset district neighborhood," *kron4.com*

[3] This approach to measuring the capitalization of (dis)amenities has been taken in a many papers across economics subfields. For example, in the education literature, Figlio and Lucas (2004) and Black (1999) study the impact of school quality on housing markets. In the environmental economics literature, examples include Chay and Greenstone (2005) on the impact of the Clean Air Act, Currie et al. (2015) on toxic planting openings and closings, Davis (2005) on cancer clusters, Davis (2011) on power plants, and Greenstone and Gallagher (2008) on hazardous waste.

retailers to nearby property sales in Washington state. Sites that "lost" the license lottery provide a natural comparison group to control for unobservables related to store site choice, thus alleviating the endogeneity concern. Moreover, we control for other neighborhood-level unobservables in the style of Linden and Rockoff (2008), treating properties within the same neighborhood but farther away from the dispensary as a control group. This approach allows us to estimate difference-in-differences and triple-difference models. Further, as the license lottery is a plausible source of exogenous variation in dispensary location, our setting provides a natural instrumental variables framework where we use the addresses in the applications of license winners as an instrument for the actual marijuana dispensaries' locations.

To complement our empirical strategy, we propose a new cross-validation procedure to define the treatment group. One potential concern in studying how amenities impact nearby neighborhoods is determining what constitutes "nearby." In many studies of how (dis)amenities affect property values, researchers often focus their analysis on properties within concentric rings around the (dis)amenity ("ring method").[4] The inner ring defines the treatment group while the outer ring defines the control group. No standardized method of selecting the radius of the treatment ring exists to our knowledge, leaving this important choice to the researcher.

As a result, the radii of the rings are generally chosen in a somewhat subjective manner even though an arbitrary choice of radius may influence the results. For example, if the treatment effect is decreasing in distance, increasing the radius from the (dis)amenity may decrease the magnitude of the estimate, washing out any promising results. On the other hand, increasing the radius adds to the number of observations used, improving the precision of the estimates. To address this issue, Diamond and McQuade (2019) develops a non-parametric difference-in-differences estimator. Complementary to their approach, we propose an easy-to-implement, data-driven procedure to select the optimal radius with which to conduct the analysis, suggesting leave-one-out cross validation to determine the appropriate distance. Our cross validation procedure balances the trade-off between precision and changes in magnitude.

The cross validation procedure yields an optimal radius of 0.36 miles, i.e., property sales that took place within 0.36 miles of a marijuana dispensary are classified into the treated group. We show that

[4] See studies such as Linden and Rockoff (2008), Currie et al. (2015), Muehlenbachs et al. (2015), Autor et al. (2014), Pope and Pope (2015), and Campbell et al. (2011).

sale prices within this distance of a marijuana dispensary decline: the estimated negative price impact is as low as 3.12% in our triple difference model and as high as 4.46% in our instrumental variables model. This decrease particularly effects younger, more diverse neighborhoods. For the average home sale price in our data, this translates to about a $10,373-$14,828 reduction in prices following the entry of a recreational marijuana dispensary. While this magnitude may seem substantial, our results are consistent with other studies on the impact of disamenities on property values found in the public economics literature.

Nevertheless, our result is highly distinct from prior work on the effects of marijuana liberalization and property values. While Adda et al. (2014) observes that a marijuana de-penalization policy in Lambeth, London decreased borough-wide property values, other studies find that marijuana liberalization *increases* home sale prices. For example, Cheng et al. (2018) compares cities in Colorado that allow recreational marijuana businesses to those municipalities that do not, concluding that cities that legalize marijuana businesses have higher property values which the authors attribute to a possible "green boom." At a more localized level, Conklin et al. (2018) studies conversions from medical marijuana dispensaries into recreational retailers in Denver, CO, estimating that homes within 0.1 miles of a medical-retail conversion increase in value relative to those slightly farther away by 8%. Similarly, Burkhardt and Flyr (2018) examines new medical and recreational dispensary entry in Denver. Using home sales within 0.25 miles of a dispensary opening as the treatment group and properties within 0.25 miles of where a new dispensary would open in the subsequent 6-12 months as a control, they find a 7.7% increase in home sale prices.

In contrast to these previous papers, our study uses extensive data on property sales throughout the entire state of Washington and comprehensive administrative data on retailers. Therefore, we are able to compare local neighborhoods both before and after any recreational retailer entry occurs. Moreover, our research design has the advantage of exogenous license distribution statewide as well as plausible counterfactual locations where no retailer enters. These differences may explain the divergence with previous estimates.

A possible driver of the estimated negative price impact is that communities may perceive that marijuana dispensaries cause crime. For instance, because marijuana is still federally illegal, cannabis businesses typically do not have access to banks and, consequently, are cash-only, making them possible targets for robbers.[5] Nonetheless, evidence on the relationship between on local crimes in areas near marijuana dispensaries is mixed.[6] For example, Freisthler et al. (2016) uncovers a positive correlation between dispensary density and violent crime in Long Beach, California while Kepple and Freisthler (2012) does not detect an association between medical dispensary density and violent crime Sacramento. A number of other papers have found correlations between dispensary density and child neglect, marijuana abuse, and youth usage (Freisthler et al., 2015; Mair et al., 2015; Shi, 2016). Using a quasi-experimental approach, Chang and Jacobson (2017) identifies temporary decreases in crimes, particularly property crime, during temporary dispensary closures in Los Angeles likely due to fewer "eyes-on-the-street." Brinkman and Mok-Lamme (2019) uses an instrumental variables approach to establish that dispensaries in Denver, CO decrease crime in the census tracts where they are located.

We add to these quasi-experimental approaches, utilizing data on police reports in Seattle, WA. Leveraging the natural experiment setting from the license distribution lottery, we use the lottery results as an instrument for dispensary location in Seattle census tracts. We estimate that overall crime reports decrease by 13.4 per 10,000 residents though our estimate is not statistically significant at 10%. Despite this, when we analyze categories of crime, we find evidence that dispensary entry increases the number of nuisance crime related reports (e.g. disorderly conduct, loitering) by 4.2 per 10,000 residents but decreases the number of drug-related reports by 2.8 per 10,000 residents. Moreover, we also find that nuisance crime reports and violent crime reports increase in adjoining census tracts by 1.8 and 2.5 per 10,000 residents, respectively. Increased nuisance related crime, therefore, may be one contributing factor to depressed home prices in areas near dispensaries.

The paper proceeds as follows. Section 2 offers details about the setting of our empirical exercise in Washington state. Our data and methodology are described in Sections 3 and 4, respectively. Section 5 details the results, and the model and results for crime are discussed in Section 6. Section 7 concludes.

[5] Abcarian, Robin. "Your Business is Legal, but You Can't Use Banks. Welcome to the Cannabis All-Cash Nightmare." *Los Angeles Times*, January 29, 2017. `http://www.latimes.com/local/abcarian/la-me-abcarian-cannabis-cash-20170129-story.html`

[6] Studies of marijuana liberalization laws on state-wide crime levels include Lu et al. (2019), Huber III et al. (2008), Anderson et al. (2013), and Anderson et al. (2015). Adda et al. (2014) studies the effects de-penalization of marijuana of bourough-wide crime in Lambeth, London.

2 Background and Institutional Details

2.1 Initiative-502

On November 6, 2012, by a statewide the vote of 55.7 percent to 44.3 percent, Washington state voters approved Initiative-502 (I-502), legalizing the possession and consumption of cannabis for adults over twenty-one years of age as well as the production and sale of marijuana in businesses regulated by the state government.[7,8] In order for firms to participate in the legalized recreational marijuana market, I-502 stipulated that a business must hold either a producer (marijuana farmers), processor (creators of joints, edibles, vapor products, etc.), or retailer license. Further, the law allowed state regulators to restrict the number of licenses it issued.

While the state's cannabis market regulator—the Washington Liquor Cannabis Board (WLCB)—opted not to limit the number of licenses issued to upstream firms such as farmers and processors, it capped the number of retail licenses state-wide at 334. It then divided up these licenses among counties using a formula that calculated "the number...[by] minimiz[ing] the population-weighted average" distance from the user to the marijuana retailer.[9,10]

The licenses were then split across the county's incorporated cities according to the proportion of the county's population within the city. The remaining licenses were assigned to the county's rural areas. For example, King County was allocated sixty-one retail licenses to be spread across seventeen incorporated cities and rural King County. Bellevue, which contains about 6.3% of King County's population, was assigned four; and Seattle, which has around a third of the county's population, was assigned twenty-one. Table 1 provides a detailed breakdown of the number of licenses for each jurisdiction.

[7] Initiative Measure No. 502, Session 2011 (WA 2011)

[8] "Initiative Measure No. 502 Concerns marijuana - County Results," Washington Secretary of State, November 27, 2012, https://results.vote.wa.gov/results/20121106/Initiative-Measure-No-502-Concerns-marijuana_ByCounty.html.

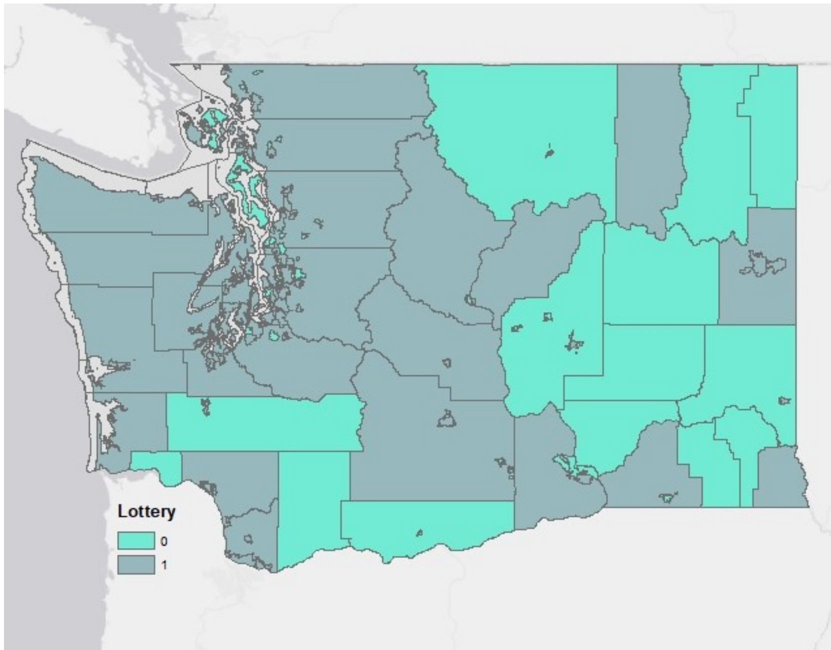
[9] To calculate this average distance, the formula assumes users are spread uniformly across the state and that "stores are placed... to maximize convenience." Hence, the "proxy" for distance is the area of a county divided by the number of stores in the county.

[10] Caulkins, Jonathan P. and Linden Dahlkemper, "Retail Store Allocation," BOTEC Analysis Corporation, Jun. 28, 2013, available from Washington Liquor Cannabis Board.

Table 1: Schedule of License Quotas

City/Locality	# Licenses	# Apps	# Entrants	City/Locality	# Licenses	# Apps	# Entrants	City/Locality	# Licenses	# Apps	# Entrants
Aberdeen	1	2	1	Island County	3	3	3	Port Townsend	1	4	1
Adams County	2	0	0	Issaquah	1	10	1	Pullman	3	14	3
Anacortes	1	3	1	Jefferson County	3	5	2	Puyallup	2	2	0
Arlington	1	3	1	Kelso	1	0	0	Quincy	1	0	0
Asotin County	2	3	2	Kennewick	4	5	0	Redmond	2	3	0
Auburn	2	5	1	Kent	3	9	0	Renton	3	9	2
Bainbridge Island	1	3	1	King County	11	44	10	Richland	3	0	0
Battle Ground	1	2	1	Kirkland	2	12	2	Sammamish	1	0	0
Bellevue	4	19	3	Kitsap County	7	40	4	San Juan Island	1	13	0
Bellingham	6	27	7	Kittitas County	2	4	0	SeaTac	1	2	0
Benton County	2	3	2	Klickitat County	3	3	2	Seattle	21	191	21
Bonney Lake	1	2	0	Lacey	2	9	2	Sedro-Woolley	1	0	0
Bothell	1	1	1	Lake Stevens	1	2	1	Selah	1	0	0
Bremerton	2	16	3	Lakewood	2	4	2	Sequim	1	5	0
Burien	1	0	0	Lewis County	4	4	0	Shelton	1	1	1
Burlington	1	1	0	Lincoln County	2	0	0	Shoreline	2	8	2
Camas	1	8	0	Longview	3	10	3	Skagit County	4	17	1
Centralia	2	4	0	Lopez Island	1	0	0	Skamania County	2	2	1
Chehalis	1	2	1	Lynden	1	0	0	Snohomish County	16	87	12
Chelan County	3	7	3	Lynnwood	2	2	0	Spokane	8	58	6
Clallam County	3	12	3	Maple Valley	1	3	0	Spokane County	7	13	6
Clark County	6	22	1	Marysville	3	9	0	Spokane Valley	3	27	3
Columbia County	1	0	0	Mason County	4	11	3	Stevens County	4	3	4
Cowlitz County	3	8	2	Mercer Island	1	1	0	Sunnyside	1	1	1
Des Moines	1	2	1	Mill Creek	1	1	0	Tacoma	8	44	9
Douglas County	2	6	1	Monroe	1	0	0	Thurston County	6	19	6
East Wenatchee	1	3	1	Moses Lake	2	5	2	Tukwila	1	4	0
Edmonds	2	2	0	Mount Vernon	3	5	2	Tumwater	1	7	1
Ellensburg	2	8	2	Mountlake Terrace	1	16	0	University Place	1	0	0
Ephrata	1	1	1	Mukilteo	1	1	0	Vancouver	6	48	6
Everett	5	27	4	Oak Harbor	1	1	1	Wahkiakum County	1	0	0
Federal Way	3	15	0	Ocean Shores	1	3	1	Walla Walla	2	2	1
Ferndale	1	1	1	Okanogan County	4	3	2	Walla Walla County	2	3	1
Ferry County	1	3	1	Olympia	2	9	3	Washougal	1	6	0
Franklin County	1	0	0	Omak	1	1	1	Wenatchee	3	2	2
Garfield County	1	0	0	Orcas Island	1	1	1	West Richland	1	1	0
Goldendale	1	1	1	Pacific County	2	19	2	Whatcom County	7	15	6
Grandview	1	0	0	Pasco	4	3	1	Whitman County	1	0	0
Grant County	3	2	2	Pend Oreille County	2	1	0	Yakima	5	7	1
Grays Harbor County	3	7	3	Pierce County	17	45	7	Yakima County	6	10	3
Hoquiam	1	2	1	Port Angeles	2	8	2	**State Total**	**334**	**1173**	**212**

Figure 1: Distribution of Jurisdictions in Washington



Notes: The blue-green shaded areas are jurisdictions where the recreational marijuana license quota was not binding, i.e., there are more licenses than the number of applications. The blue-grey areas are jurisdictions where the recreational marijuana license quota was binding and therefore a lottery to allocate the license was carried out within the jurisdiction.

2.2 The Washington Marijuana Retail License Lottery

In November 2013, the WLCB opened a thirty day window during which potential marijuana retailers could apply for a retailer license. Applicants were subject to background checks to determine if they were eligible licensees. Furthermore, as stores were banned from locating within 1000 feet of a "school, playground, ... child care center, public park, public transit center, or library", the license applications required a potential store address so that regulators could determine compliance with the location restrictions.[11]

Prospective firms could submit multiple applications for multiple licenses. However, the state imposed restrictions on the number a firm could obtain: A business could not have more than three licenses and no more than a third of all stores in a jurisdiction. Moreover, while the application fee was a nominal $250, many businesses did not submit several applications. In fact, 99 percent of the 802 individual applicants turned in less than three applications with 68 percent submitting just one,

[11] Initiative Measure No. 502, Session 2011 (WA 2011).

and 51 percent of firms that handed in multiple applications were not petitioning for licenses in the same jurisdiction.

In total, though only 334 retail licenses were available, 1,173 applications were submitted. Table 1 lists the number of applicants and available licenses for each jurisdiction as well as the number of dispensaries in the market by February 2016. Seventy-five jurisdictions had more applicants than licenses, and the WLCB decided to distribute licenses for these areas via a lottery.[12] Figure 1 highlights which jurisdictions allocated licenses via lottery.

The license lotteries were held April 21-25, 2014. Each applicant in a lottery was randomly assigned a number by the accounting firm Kraght-Snell. The numbers—without any identifying information—were then sent to Washington State University's Social and Economic Sciences Research Center, which ranked the numbers from 1 to n, with n being the number of applicants within a jurisdiction. Kraght-Snell then decoded the rankings. If a ranking number was higher than the number of licenses allocated to a jurisdiction, the firm was a lottery "winner." The results of the lottery, which were well-publicized in the state and local press, were made public on May 2, 2014.[13]

2.3 Entry in the Recreational Marijuana Market

Contingent on receiving a license, licensees could begin selling marijuana as early as July 2014. Figure 2 shows the evolution of store entry over time. Seventy percent of lottery winners entered the market, and half of those that did not enter (i.e., 15 percent of the lottery winners) were kept from opening due to local bans on marijuana businesses.[14] We cannot account for the other half of lottery winners that did not enter the market though one possibility is that potential firm owners failed subsequent background checks. However, if a lottery winner could not enter due to a failed background check, etc., Washington awarded the license to the next applicant in the lottery ranking.[15]

Importantly, entrants generally located at or near the address listed on their license applications even though addresses were not legally binding, with 47 percent locating in the exact address, and 28 percent locating within one-third of a mile from the address listed on their application. This suggests

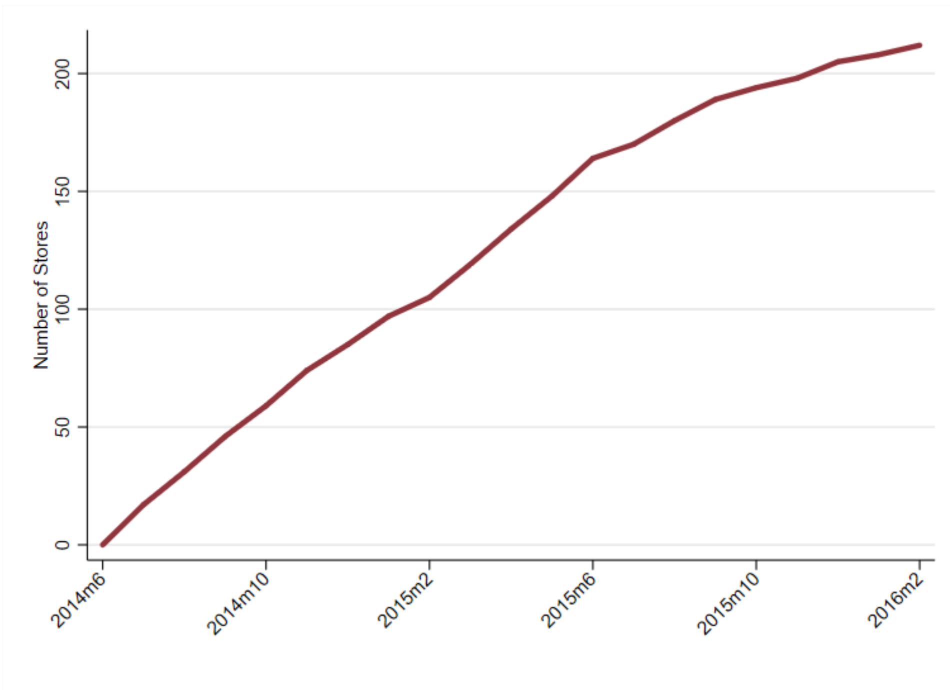
[12] The state did not reallocate the licenses of those jurisdictions whose quota did not bind to other parts of the state which de facto capped the number of licenses at less than 334.

[13] See "Who Won the Pot Shop Jackpot and Where the Stores Might Be," *The Seattle Times*, May 2, 2014.

[14] The WLCB did not reallocate licenses in these areas to other jurisdictions.

[15] Another possible explanation for failure to entry is the potential firm holding the license for speculative purposes.

Figure 2: Entry of Recreational Marijuana Dispensaries Over Time



Notes: The figure shows the cumulative number of recreational marijuana dispensaries between June 2014 and February 2016.

that addresses submitted for the lottery constitute meaningful information and can serve as a good predictor for the actual dispensary location—a fact that is crucial for our empirical strategies.

3 Data

We assemble a novel data set from a variety of sources. For our primary analysis, we first use the results of the marijuana retail license lottery, provided by the WLCB through public records request. The data includes the applicant's tradename, application number, lottery rank, and address for the potential store's location. The WLCB also provided data on jurisdictions' license allocations which enables us to identify lottery "winners"—those applicants whose lottery rank is less than or equal to the number of licenses allocated to the jurisdiction.[16]

Next, we combine the lottery results with information on operating cannabis retailers, allowing us to view which lottery winners eventually entered the market. The retailer data comes from the WLCB's Traceability System, a system that allows the state to track marijuana products through the cannabis supply chain until the products are sold by dispensaries. The data includes retailers' addresses and sales for each day. As not all dispensaries opened on July 2014—the first month of legal sales—we construct the firms' entry dates by defining the entry date as the day of first sale. Figure 3a displays the geographic distribution of retailers and license applicants across Washington.

Our housing sales data is supplied by CoreLogic (previously named DataQuick), a private vendor. The data is generated from public records created by tax assessors as well as from proprietary records created by the multiple listing services. While the data set includes properties from all across Washington, as Figure 3b shows, the coverage is the best along the Interstate 5 corridor, the most populous area of the state. The data includes comprehensive housing characteristics for each property–e.g., square footage, year built, whether the property is a single or multi-family home or condo, and address–and information on the property transaction–e.g., sale date and sale price (which we convert to January 2014 dollars).

We geocode all addresses using the commercial geocoder *Here*, keeping only those property addresses that are identified with high accuracy at the address number level. We then calculate the

[16] We also include those applicants that initially lost the lottery but were awarded licenses due to failed background checks by highly ranked applicants as winners.

geodesic distance from properties to retailers and lottery participant addresses. As we are interested in finding comparable neighborhoods that are attractive to retail applicants but where some firms enter and others do not, we limit the sample to the 75 jurisdictions that allocated retailer licenses via lottery and restrict the data to those properties within one mile of a lottery participant address. We also narrow the sample used in our analysis to home sales occurring from January 2012 to February 2016, as Washington state expands the number of recreational marijuana licenses in 2016. We also drop sales outside the range of $14,700 to $2,500,000 (i.e., the 0.025 percentile and 99.975 percentile in the price distribution). Figure 4 provides a pictorial example of how we construct the data sample.

For additional information about neighborhoods, we connect property locations to their census tracts using the TIGER/Line shapefiles from the U.S. Census Bureau. The shapefiles are linked to the 2014 American Community Survey (ACS) which allows us to incorporate tract-level demographic data into the sample. These data include tract population, median income, and binned counts of age, education level, and race. Using this information, we compute the percentage of individuals in the census tract that are high school and college graduates, the percentage of non-hispanic white individuals in the census tract, as well as the percentage of individuals in the census tract between the ages of 18 and 35.[17]

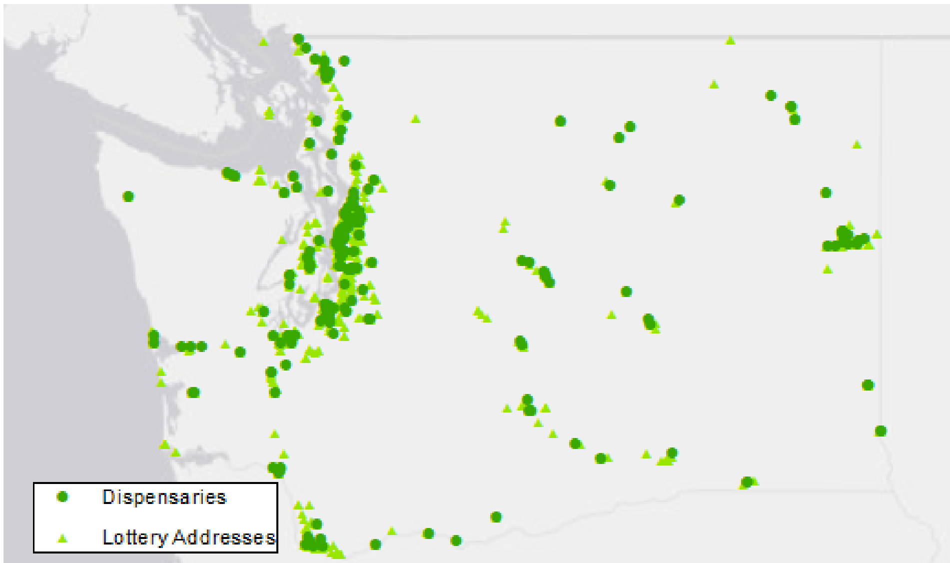
Overall, the data consists 84,166 records from 25 counties in Washington, 216 stores, and 1,125 lottery applications. Table 2 provides the summary statistics of the data. Properties within one mile of a dispensary tend to be less expensive, older, and slightly smaller, while the neighborhoods within one mile of dispensaries tend to be slightly younger, slightly more diverse, and less wealthy. This pattern also repeats itself when comparing areas around “winning” addresses versus “losing” addresses. This is suggestive of selection of stores into neighborhoods. We discuss our strategies to identify causal effects in spite of selection in Section 4 further in Section 5.3.

4 Empirical Methodology

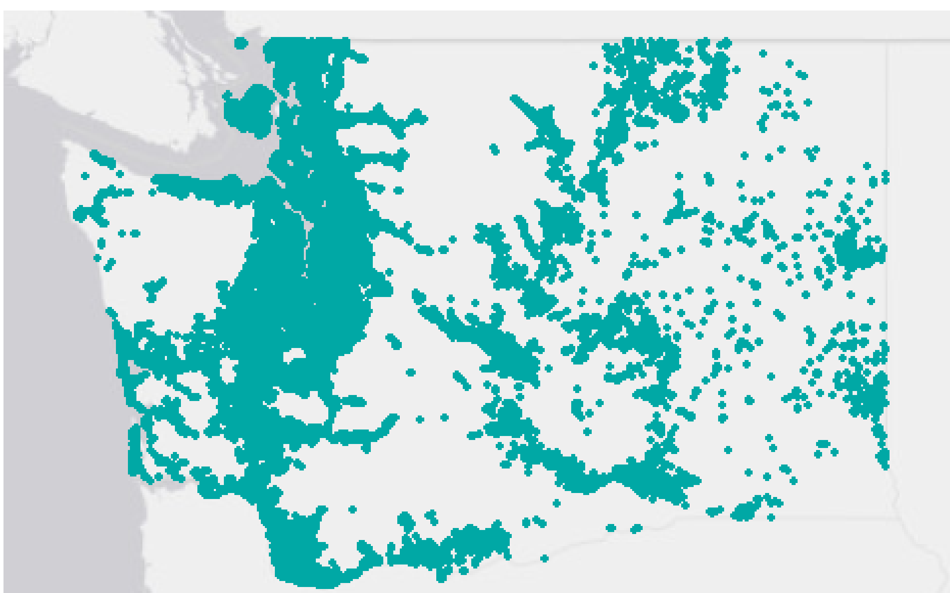
As is well understood, a major challenge in identifying the capitalization of local (dis)amenities such as cannabis retailers into housing prices is that variation in the amenity is rarely exogenous

[17] Azofeifa et al. (2016) finds that people under thirty-five are primary users of marijuana.

Figure 3: Geographic Distribution of Retailers and Properties



(a) Dispensaries



(b) Properties

Notes: In Figure 3a each dot represents a single dispensary that was open from July 2014–February 2016 while in Figure 3b each dot represents a single property in the Corelogic data set that was sold from July 2014–February 2016.

Table 2: Summary Statistics

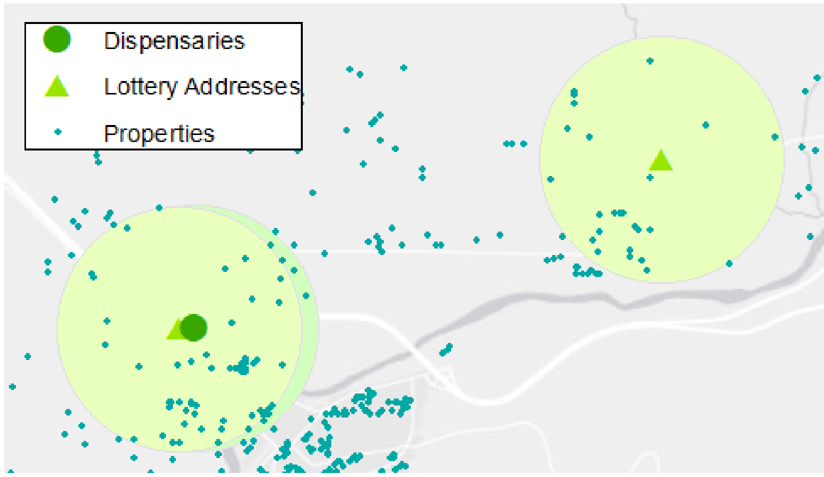
Panel A: Property Characteristics

	Mean (*Standard Deviation*)				
	All Properties	Dispensary	No Dispensary	Winners	Losers
Closing Price	332,486	321,068	350,695	314,868	359,462
	(254,453)	(249,362)	(261,342)	(243,087)	(268,709)
Beds	2.887	2.891	2.88	2.898	2.87
	(1.021)	(1.017)	(1.029)	(.9821)	(1.079)
Baths	1.807	1.789	1.835	1.798	1.821
	(.7774)	(.7802)	(.7722)	(.7754)	(.7803)
Home Age	42.84	43.82	41.29	43.07	42.5
	(34.01)	(34.46)	(33.22)	(33.83)	(34.28)
Square Footage	1,675	1,666	1,688	1,655	1,705
	(864.1)	(911.5)	(782.5)	(720.2)	(1,046)
% 18-34 y.o.	.2858	.2959	.2697	.2801	.2947
	(.106)	(.1097)	(.0978)	(.0979)	(.1169)
% White	.6801	.6723	.6926	.6793	.6814
	(.156)	(.1612)	(.1453)	(.1565)	(.1542)
% H.S. Grad.	.6428	.6363	.6531	.6366	.6523
	(.101)	(.1015)	(.1004)	(.0983)	(.1053)
Median Income	63,561	61,856	66,279	62,592	65,044
	(21,790)	(21,995)	(21,177)	(21,751)	(21,765)
Observations	84,166	51,729	32,437	50,915	33,251

Panel B: Applicants and Entrants

	Entrants	Winners	Losers
Number of Firms	216	257	868
Percentage of Applications	.192	.228	.772

Figure 4: Dispensaries, Lottery Addresses, and Surrounding Neighborhoods



Notes: Buffers around dispensaries and lottery addresses have a radius of one mile.

and is likely correlated with factors unobserved by the econometrician. For example, in our setting, dispensaries will likely locate in neighborhoods that already have high marijuana demand. If latent cannabis demand is correlated with housing prices, then estimates of the effects of dispensary entry will be biased. Hence, we use three alternative empirical strategies to address these selection issues.

4.1 Strategy 1: Difference-in-Differences (DD)

A popular strategy to correct for selection is to compare properties close—within an r mile radius– to a (dis)amentity to properties that are slightly farther away—within $R > r$ miles. The nearby properties are the most likely to be adversely or positively effected by the (dis)amenities, while properties slightly farther away should share many characteristics and price trends of the nearby properties but be less impacted by localized effects, making those properties a desirable control group.

To elaborate further, we define the change in a property's price after dispensary entry as ΔP_k^j with $j = \mathbb{1}(d \leq r)$ and $k = \mathbb{1}(d \leq R)$. The variable d as the distance is to the closest cannabis dispensary. For ΔP_k^j,

$$\begin{aligned}\Delta P_0^1 &= \Delta Neighborhood + \Delta Macro, \\ \Delta P_1^1 &= \Delta Dispensary + \Delta Neighborhood + \Delta Macro.\end{aligned}$$

Price changes can be driven by macroeconomic price trends ($\Delta Macro$) as well as neighborhood-level price trends ($\Delta Neighborhood$). However, changes due to store entry ($\Delta Dispensary$) impact only the closest properties. Hence, the difference-in-differences (DD) calculation identifies the effect of dispensary entry:

$$\Delta Dispensary = \Delta P_1^1 - \Delta P_0^1. \tag{1}$$

This motivates our first empirical specification, a DD regression of the form

$$\begin{aligned}\ln(p_{ijt}) = \beta_0 + \beta_1 \text{Post}_{it} + \beta_2 \mathbb{1}(d_i \le r) + \beta_3 \mathbb{1}(d_i \le r) \cdot \text{Post}_{it} \\ + \beta_5 \mathbb{1}(d_i \le R) + \beta_6 \mathbb{1}(d_i \le R) \cdot \text{Post}_{it} + \beta_7 X_{ijt} + \varepsilon_{ijt}, \quad (2)\end{aligned}$$

where p_{ijt} is the sale price of observation i in area j at time t. The variable Post_{it} is an indicator function that equals one after the announcement of the license winners, consistent with a model of forward-looking consumers. The vector X_{ijt} is a vector of property characteristics (number of bedrooms and bathrooms, age, log square footage, property type) and census tract characteristics (median tract income, percentage of high school graduates, percentage of individuals between 18 and 35, percentage of the tract population that is non-hispanic white), quarter-year fixed effects, and area (city or zipcode) dummy variables. The coefficient of interest is β_3, the treatment effect of dispensary entry.

Cross Validation

To estimate Equation (2), an important decision is selecting the radius, r, that defines the "nearby" treated group. Assuming differential treatment effects in distance, an r that is too small places many observations possibly impacted by marijuana dispensary entry in the control group. On the other hand, an r that is too large includes many properties that are not impacted by dispensary entry in the treatment group, biasing the estimates toward zero. A larger r also includes more observations in the treated group, adding precision to the estimates.

To make ideas more concrete, we consider the following simple model:

$$y_i = \alpha + T_i^r \beta^r + \varepsilon_i,$$

where $T_i^r = \mathbb{I}(d_i \leq r)$. A well-established result is that the ordinary least squares estimate of β is

$$\hat{\beta}^r \quad = \quad \frac{\frac{1}{N}\sum_{i=1}^N T_i \cdot y_i - \left(\frac{1}{N}\sum_{i=1}^N T_i\right) \cdot \left(\frac{1}{N}\sum_{i=1}^N y_i\right)}{\frac{1}{N}\sum_{i=1}^N T_i^2 - \left(\frac{1}{N}\sum_{i=1}^N T_i\right)^2}. \tag{3}$$

By defining $N(r) = \sum_{i=1}^N \mathbb{I}(d_i \leq r)$, Equation (3) can be simplified to

$$\hat{\beta}^r = \frac{1}{N(r)} \cdot \sum_{i=1}^{N(r)} y_i - \frac{1}{(N - N(r))} \cdot \sum_{i=N(r)+1}^{N} y_i. \tag{4}$$

Equation (4) can be interpreted very intuitively: The estimated average treatment effect is the sample average of those i in where $d_i \leq r$ minus the sample average of those observations such that $d_i > r$. However, as mentioned above, if r is increased, then some observations that were originally included in the average for the control group are now included into the average for the treatment group. In another word, $\hat{\beta}^r$ is sensitive to the choice of r.

Further, increasing r reduces the variance of $\hat{\beta}^r$. The heteroskedastic robust standard error of the parameter estimates $\hat{\theta} = (\hat{\alpha}, \hat{\beta}^r)$ is

$$V_{\hat{\theta}} = \left(\bar{T}'\bar{T}\right)^{-1} \left(\bar{T}'D\bar{T}\right) \left(\bar{T}'\bar{T}\right)^{-1},$$

where $\bar{T}' = \begin{pmatrix} \mathbf{1}'_{N\times 1} \\ \mathbb{I}(d \leq r)' \end{pmatrix}$, and $D = \text{diag}(\sigma_1, ..., \sigma_N)$. Some matrix algebra reveals that the variance of $\hat{\beta}^r$ is

$$V_{\hat{\beta}^r} = \left(\frac{1}{N(r)^2} - \frac{1}{(N - N(r))^2}\right) \cdot \sum_{i=1}^{N(r)} \sigma_i^2.$$

As $N(r)$ increases, $\left(\frac{1}{N(r)^2} - \frac{1}{(N-N(r))^2}\right)$ decreases because $\sum_{i=1}^{N(r)} \sigma_i^2 \geq 0$. Thus, it follows that as long as $\left(\frac{1}{N(r)^2} - \frac{1}{(N-N(r))^2}\right)$ decreases faster than $\sum_{i=1}^{N(r)} \sigma_i^2$, increasing r increases precision of the estimate $\hat{\beta}$.

Therefore, picking an appropriate r is critical in estimating the treatment effect. The conventional approach in dealing with this choice, however, is somewhat arbitrary. The most careful approaches such as Currie et al. (2015) on air pollution use a selection process that is informed by the science.[18] However, when the scientific literature offers less direction, others such as Linden and Rockoff (2008) and Muehlenbachs et al. (2015) guide the choice of radius by plotting the non-parametric price gradient and finding a "sensible" distance at which to define the nearby properties.

We aim to formalize a selection process by developing an easy to implement, *data-dependent* rule for choice of r. The considerations in distance selection echos the classic "bias-variance trade-off" in econometric analysis. We can characterize this "bias-variance trade-off" by denoting the "optimal" predictor of the treatment effect as

$$Y \quad = \quad m^*(T) + \epsilon, \tag{5}$$

where $\mathbb{E}(\epsilon)^2 = \sigma^2$ while $\hat{m}_r(T) = \hat{\alpha} + T^r \hat{\beta}^r + \varepsilon$. The mean squared error of the predictor is

$$\mathbb{E}(Y - m_r(T)) = \mathbb{E}\left(m^*(T) - \hat{m}_r(T)\right)^2 + \mathbb{E}\left(\hat{m}_r(T) - \mathbb{E}(\hat{m}_r(T))\right)^2 + \sigma^2. \tag{6}$$

The first expression in on the right side of Equation (6) is the squared bias of the estimator while the second expression is the variance of the estimator.

The best choice of r is the one that minimizes the mean squared prediction error. We choose r such that r minimizes the sample mean squared prediction error.[19] Specifically, we use leave-one-out cross validation so that

$$\min_r \sum_i \left(y_i - \hat{y}_{-i}\right)^2, \tag{7}$$

where y_{-i} is the leave-one-out predicted value of y_i.[20]

[18]Specifically, Currie et al. (2015) use data on ambient levels of hazardous air pollution to define the treatment radius.

[19]This is analogous to cross-validation procedures found in the non-parametric regression literature.

[20]While $k-$fold cross validation can be used, leave-one-out cross validation is relatively inexpensive from a computa-

4.2 Strategy 2: Difference-in-Difference-in-Differences (DDD)

While the difference-in-differences specification controls for many neighborhood-level unobservables that effect marijuana demand, there could still exist systematic unobserved price trends between properties near dispensaries and properties further away that are not adequately captured. However, our setting in Washington allows for us to account for these differences because we observe properties in our data that were located in a location attractive to marijuana firms but miss out on being close to a dispensary due to the license quota.

To further illustrate our approach, we now denote the change in a property's price after dispensary entry as $\Delta P_{k,n}^{j,m}$ with j and k defined as in Section 4.1, and $m = \mathbb{1}(\tilde{d} \leq R)$ and $n = \mathbb{1}(\tilde{d} \leq r)$. The variable $\tilde{d}$ is the distance to the closest lottery address or dispensary.[21,22] For $\Delta P_{k,n}^{j,m}$,

$$\begin{aligned}
\Delta P_{0,0}^{1,1} &= \Delta Neighborhood + \Delta Macro, \\
\Delta P_{1,1}^{1,1} &= \Delta Dispensary + \delta + \Delta Neighborhood + \Delta Macro, \\
\Delta P_{0,1}^{0,1} &= \delta + \Delta Neighborhood + \Delta Macro \\
\Delta P_{0,0}^{0,1} &= \Delta Neighborhood + \Delta Macro,
\end{aligned}$$

where δ denotes price trends common to areas that are/would be close to marijuana dispensaries. A triple difference identifies the treatment effect, i.e., $\Delta Dispensary$.

$$\Delta Dispensary = (\Delta P_{1,1}^{1,1} - \Delta P_{0,0}^{1,1}) - (\Delta P_{0,1}^{0,1} - \Delta P_{0,0}^{0,1}). \tag{8}$$

tional standpoint due to the fact that

$$\sum_i (y_i - \hat{y}_{-i})^2 = \sum_i \left(\frac{y_i - \hat{y}_i}{1 - h_{ii}}\right)^2$$

where h_{ii} is the leverage value of the hat matrix $\mathbf{P} = \mathbf{T}(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'$.

[21] Using all lottery addresses or only lottery "losers" does not result in a statistically significant change in our estimates.

[22] In order to keep sharp distinctions between treatment and control groups, we exclude properties where $\mathbb{1}(d_i \leq R) \cdot \mathbb{1}(\tilde{d}_i \leq r) = 1$ but $\mathbb{1}(d_i \leq r) = 0$.

This motivates the following triple-difference estimating equation:

$$\begin{aligned}\ln(p_{ijt}) = \alpha_0 + \alpha_1 \text{Post}_{it} + \alpha_2 \mathbb{1}(d_i \leq r) + \alpha_3 \mathbb{1}(d_i \leq r) \cdot \text{Post}_{it} \\ + \alpha_5 \mathbb{1}(d_i \leq R) + \alpha_6 \mathbb{1}(d_i \leq R) \cdot \text{Post}_{it} \\ + \alpha_7 \mathbb{1}(\tilde{d}_i \leq r) + \alpha_8 \mathbb{1}(\tilde{d}_i \leq r) \cdot \text{Post}_{it} + \alpha_{10} X_{ijt} + u_{ijt} \quad (9)\end{aligned}$$

where X_{ijt} is the same as in Section 4.1 and α_3 is the triple-difference estimate.

4.3 Strategy 3: Instrumental Variables (IV)

Our setting also provides a source of plausibly exogenous variation, the license lottery, in which some properties are assigned to treatment and others to control. However, as seen in Figure 4, though licensees generally stayed close to their original stated addresses, stores were allowed to move locations. This sort of randomization with partial-compliance implies a natural instrumental variables framework. Therefore, we modify our difference-in-differences model from Section 4.1 into a instrumental variables model.

Following in the spirit of Chaisemartin and D'Haultfoeuille (2018), we denote those properties in the same neighborhood as a lottery winner as $G_i^R = \mathbb{1}(d_i^W \leq R)$ and $G_{it}^R = \mathbb{1}(d_i^W \leq R) \cdot Post_{it}$ where d_i^W is distance to the closest "winning" lottery address. The variables denoting properties near license winners, G_i^r and G_{it}^r, are defined similarly. As the number of available licenses varies across cities, and many homes within the same city may be relatively close to multiple lottery addresses, some properties have a higher chance of being assigned to treatment than others. Hence, G_i^R and G_i^r are random conditional on W_i, a vector of variables that control for the probability of treatment. To construct W_i, we create dummy variables for the quartile of the number of license applications within r miles and within R miles along with their interactions.[23] We also interact these dummy variables with city/zipcode variables.

We denote realized treatment after dispensary entry as $D_{it}^r = \mathbb{1}(d_i \leq R) \cdot Post_{it}$ with D_{it}^R defined analogously. Because of non-compliance, $G_{it}^r \neq D_{it}^r$. Therefore, we use G_{it}^r and G_{it}^R as instruments for

[23] In practice, we use an interaction between G_i^R and the dummy variables for the quartile of the number of license applications within r rather than simply G_i^r.

D_{it}^r and D_{it}^R. The IV model is

$$\begin{aligned}
\ln(p_{ijt}) &= \tau_0 + \tau_1 D_{it}^r + \tau_2 D_{it}^R + \tau_4 G_i^r + \tau_5 G_i^R + \tau_6 W_i + \tau_7 X_{ijt} + \nu_{ijt}, && (10)\\
D_{it}^r &= \lambda_0 + \lambda_1 G_{it}^r + \lambda_2 G_{it}^R + \lambda_4 G_i^r + \lambda_5 G_i^R + \lambda_6 W_i + \lambda_7 X_{ijt} + \mu_{ijt}, && (11)\\
D_{it}^R &= \pi_0 + \pi_1 G_{it}^r + \pi_2 G_{it}^R + \pi_4 G_i^r + \pi_5 G_i^R + \pi_6 W_i + \pi_7 X_{ijt} + \xi_{ijt}. && (12)
\end{aligned}$$

We also study the intent-to-treat (ITT) model

$$\ln(p_{ijt}) = \sigma_0 + \sigma_1 G_{it}^r + \sigma_2 G_{it}^R + \sigma_4 G_i^r + \sigma_5 G_i^R + \sigma_6 W_i + \sigma_7 X_{ijt} + \chi_{ijt} \qquad (13)$$

The coefficients of particular interest are τ_1 in (11), the average treatment effect on the treated, and σ_1 in (13), the intent-to-treat effect. However, Chaisemartin and D'Haultfoeuille (2018) cautions that τ_1 may not identify the average treatment on the treated without the strong assumption that local average treatment effects are homogeneous between treatment and control groups. The researchers suggest using control groups where the distribution of treatment does not change over time. Thus, following the suggested approach, we also define an alternative control group: where for $G_i = 0$, $G_{it} = D_{it} = 0$.

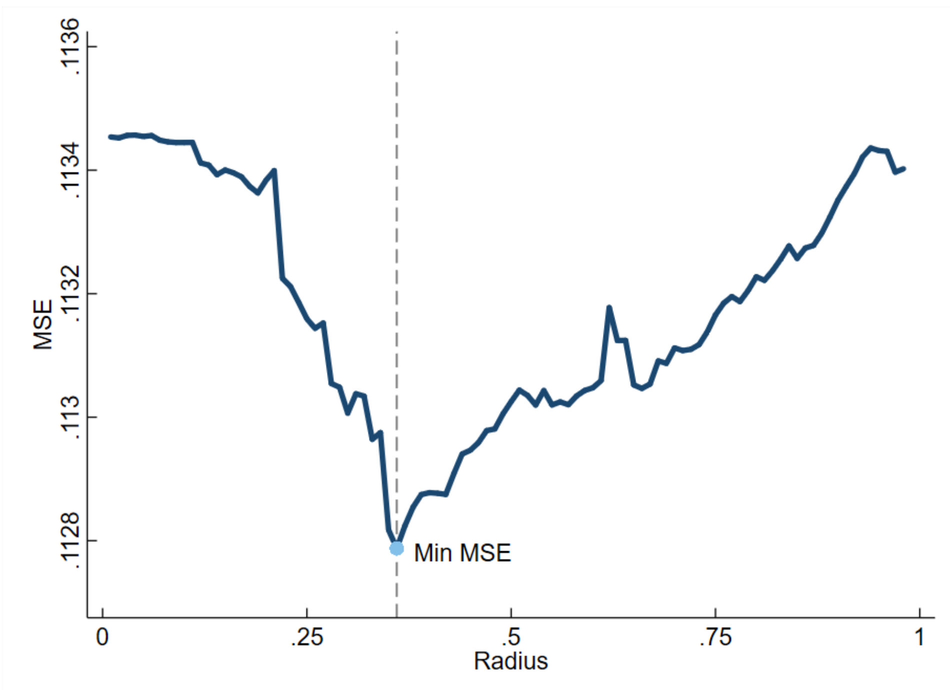
5 Empirical Results

5.1 Cross Validation

The results of the cross-validation procedure on Equation (2) are reported in Figure 5. The estimated mean squared error takes the classic U-shape and is minimized at a distance $r^* = 0.36$ miles. To assess the reasonableness of this result, we follow Linden and Rockoff (2008) and Muehlenbachs et al. (2015) and estimate local linear regressions of log home sale price on distance to nearest dispensary (or lottery address). The results are displayed in Figure 6. "Treatment" in Figure 6 refers to those properties within one mile of a marijuana retailer while "Control" references properties that were within one mile of a lottery address but missed being in the same neighborhood as a dispensary.

Figure 6a plots the pre-lottery prices with respect to distance, showing that prices around dispen-

Figure 5: Cross Validation Results



Notes: This figure plots the mean squared errors, as specified in Equation (7), under different treatment radius (in miles).

saries and around lottery addresses have prices that evolve similarly prior to retailer entry. Figure 6b shows the post-license lottery period. After the license lottery, sale prices in areas near dispensaries decrease. The effect seems particularly pronounced between 0.15-0.4 miles. The vertical grey dashed line denotes the 0.36 miles, which on visual inspection draws a sensible boundary between properties affected and unaffected by the marijuana dispensary location.

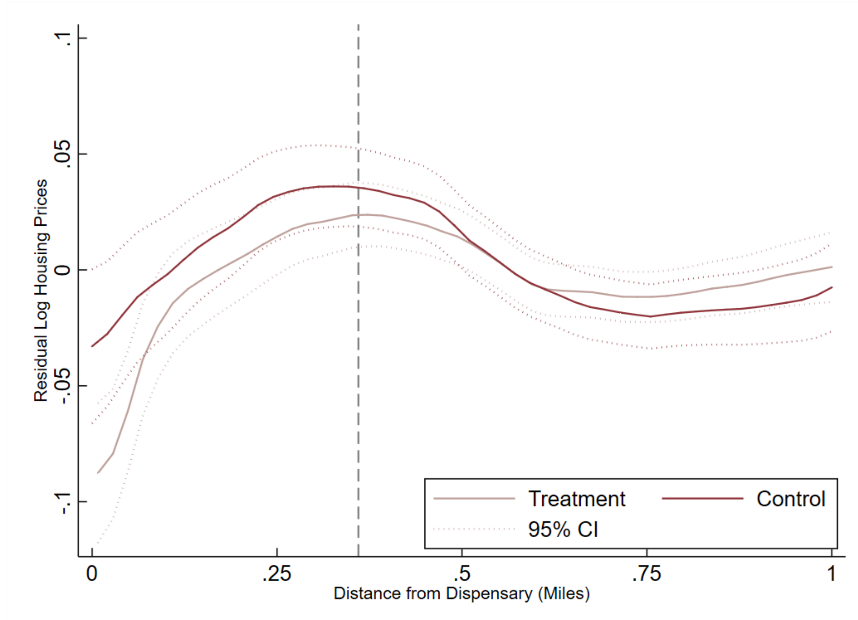
5.2 Parallel Trends

With the cross-validation results in hand, we can study the identification assumptions of our empirical approaches. In a difference-in-differences design, the primary assumption is that absent of the treatment, the treated and control groups would evolve along parallel trends. We examine price trends directly by estimating a local polynomial regression of log home sale prices on days before and after the license lottery results are announced. Figure 7 displays the results.

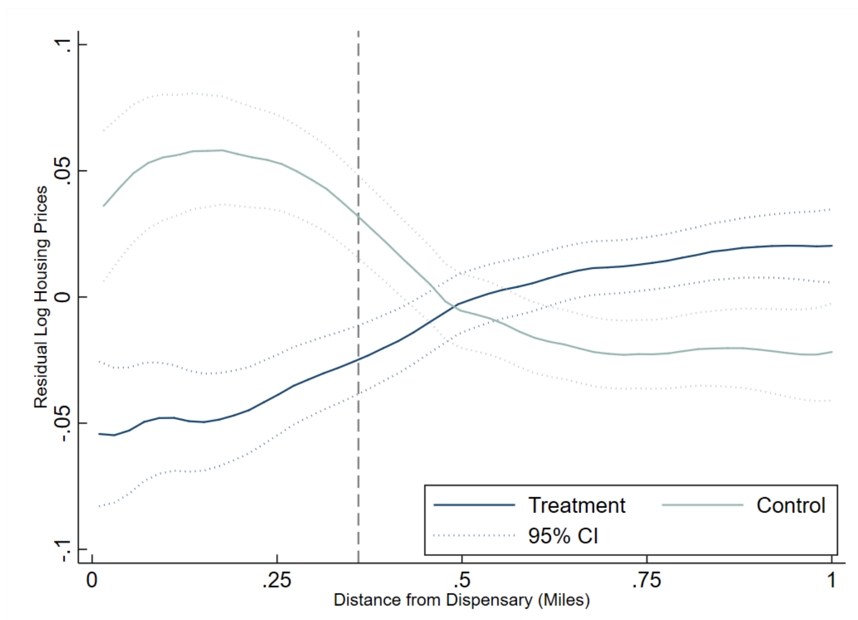
Figure 7a shows the evolution in prices of properties within 0.36 miles of a dispensary or control lottery address while Figure 7b displays the same information for properties greater than 0.36 miles away. Price trends for farther away properties are almost identical. For properties within 0.36 miles, prices evolve similarly to farther away properties pre-lottery, and there is little evidence of differential trends between treated properties and control properties. After the lottery announcement, however, prices diverge. This change is most noticeable around 200 days after the lottery winners announcement. (This pattern is mirrored in our event study graph Figure A.1.) As seen in Figure 2, only a few stores enter after July 2014, which may contribute to the delayed response.

Together, Figures 7a and 7b lend credibility to the critical parallel trends assumption. Further, though systematic differences in prices may exist between properties close to dispensaries and properties farther away (see Figure 6), both within-neighborhood and within-location time trends are differenced out in the DDD specification. In another words, the DDD estimate is immune to both neighborhood-specific shocks—such as transitory demand shocks that affect the property prices in a particular neighborhood—and location-specific shocks—such as fluctuations in demand for properties very close to a dispensary. Hence, the identification assumption for consistency of the DDD estimate is that there was no shock during our study period that differentially affected prices of only the treatment location in the treatment neighborhood. Given the random nature of the license allo-

Figure 6: Sale Price Gradient of Distance from Closest Dispensary



(a) $Post = 0$



(b) $Post = 1$

Notes: The figures display the results of a local polynomial regression of log sale prices (de-meaned) and the distance to the nearest dispensary or lottery address. "Treatment" properties refer to those properties in the same neighborhood as a dispensary while "Control" refer to properties in neighborhoods that missed out on dispensary entry. The bandwidth is 0.125 miles. The gray dashed line denotes the $r^* = 0.36$ found by the cross-validation procedure.

cation lottery, we believe there are reasonable grounds to believe that the identification assumption is unlikely to be violated.

5.3 Covariate Balance

For the IV specification, causal inference rests on the assumption that G_i^r is random conditional on vector W_i. To investigate this further, we study differences in the pre-randomization characteristics of properties and neighborhoods. Table 3 reports the results of the covariate analysis.

Simply looking at simple differences between the groups $G_i^r = 0$ and $G_i^r = 1$ reveals that properties near lottery winners are smaller, older, and in younger neighborhoods in the pre-lottery period. However, raw differences do not control for the fact that the probability of treatment differs across cities and neighborhoods. Therefore, we estimate the following regression:

$$y_{ijt} = \rho_0 + \rho_1 G_i^r + \rho_2 W_i + \epsilon_{ijt} \tag{14}$$

where y_{ijt} are the property and neighborhood characteristics found in X_{ijt}. Zipcode fixed effects are also included in W_i. We report the p-value of the regression in the third column of Table 3. Reassuringly, for all but one variable, the median income of the census tract, we are unable to reject the null of balance for the characteristics.

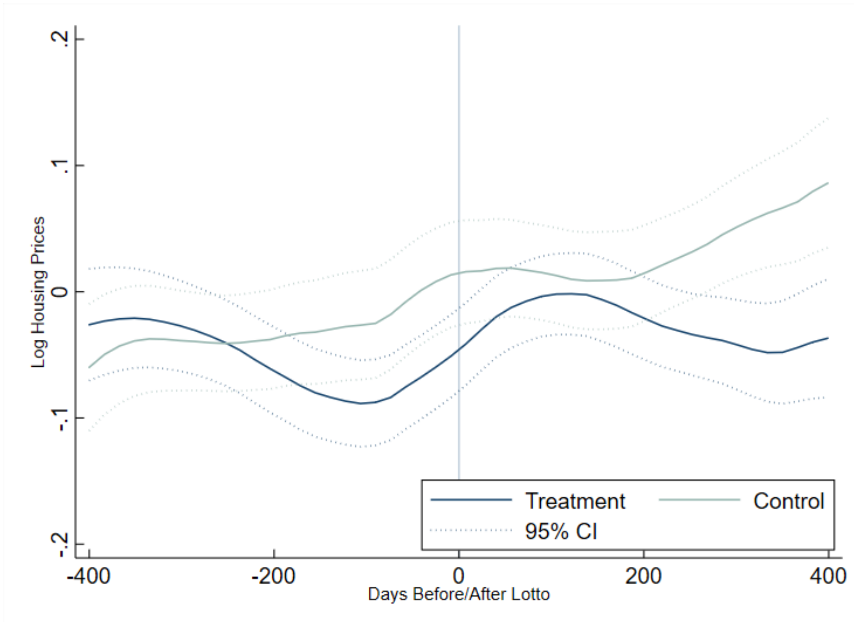
We also conduct a similar analysis for G_i^R. The results are shown in the sixth column of Table 3. One variable, the percentage of the census tract population between the ages of 18 and 35, has a marginally significant coefficient on G_i^R. However, we cannot reject the null of balance for all other variables.

5.4 Estimation Results

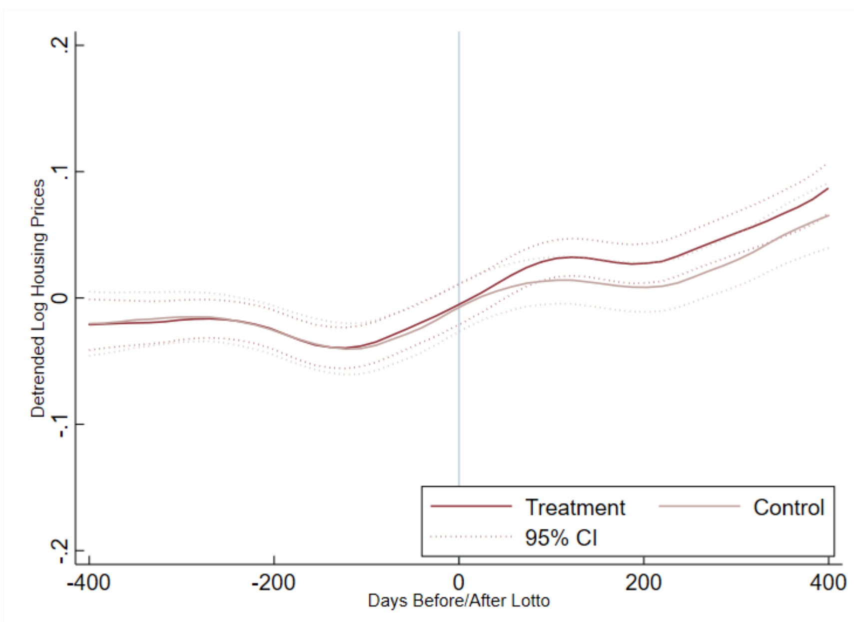
The estimation results for Equation (2), the difference-in-differences model, are reported in Columns (1) and (2) of Table 4. Even after controlling for zipcode fixed effects, the estimated effects of dispensary entry on nearby property values are very small and insignificant.

However, as shown in Figure 6, our DD model may not adequately control for unobservables in

Figure 7: Time Trends



(a) $\mathbb{1}(d \leq r^*) = 1$



(b) $\mathbb{1}(d \leq r^*) = 0$

Notes: The figures display the results of a local polynomial regression of log sale prices (de-meaned) and days before and after the license lottery. The bandwidth is 75 days. "Treatment" properties refer to those properties where $\mathbb{1}(d_i < R) = 1$.

Table 3: Covariate Balance

	G_i^r			G_i^R		
	0	1	p-value	0	1	p-value
Beds	2.91	2.655	.4837	2.87	2.898	.3428
	(1.021)	(.9991)		(1.079)	(.9821)	
Baths	1.819	1.693	.6767	1.821	1.798	.3037
	(.7791)	(.7511)		(.7803)	(.7754)	
Home Age	42.76	43.66	.7831	42.5	43.07	.3538
	(34.04)	(33.71)		(34.28)	(33.83)	
Square Footage	1,691	1,513	.1171	1,705	1,655	.7633
	(883.8)	(615.9)		(1,046)	(720.2)	
% 18-34 y.o.	.2841	.3027	.3316	.2947	.2801	.0542
	(.1055)	(.1095)		(.1169)	(.0979)	
% White	.6803	.6788	.6087	.6814	.6793	.1463
	(.1573)	(.1374)		(.1542)	(.1565)	
% H.S. Grad.	.6414	.6566	.7418	.6523	.6366	.8443
	(.1014)	(.1002)		(.1053)	(.0983)	
Median Income	63,532	63,843	.017	65,044	62,592	.8114
	(19,534)	(19,534)		(21,751)	(21,751)	
N	76,422	7,744		33,251	50,915	

Notes: The p-value is based on standard errors clustered at the jurisdiction level.

Table 4: Difference-in-Differences and Triple Difference Regressions

	DD		DDD	
	(1)	(2)	(3)	(4)
$\mathbb{1}(d_i \leq r) \cdot \text{Post}_{it}$	-.0043	-.0052	-.032**	-.0312**
	(.0124)	(.0109)	(.0144)	(.0137)
$\mathbb{1}(d_i \leq r)$	-.0954***	-.0672***	-.0851***	-.0414
	(.0204)	(.0239)	(.0309)	(.0259)
$\mathbb{1}(d_i \leq R) \cdot \text{Post}_{it}$	.0127	.0115	.0186*	.017*
	(.0093)	(.0079)	(.0101)	(.0093)
$\mathbb{1}(d_i \leq R)$	-.0475*	-.0234	-.0496*	-.0282
	(.0285)	(.018)	(.0263)	(.0178)
$\mathbb{1}(\tilde{d}_i \leq r) \cdot \text{Post}_{it}$			.0277***	.026***
			(.0081)	(.0082)
$\mathbb{1}(\tilde{d}_i \leq r)$			-.0102	-.0259*
			(.0215)	(.0151)
City FEs	Yes	No	Yes	No
Zipcode FEs	No	Yes	No	Yes
N	83,607	83,522	83,607	83,522
R^2	.776	.803	.776	.804

Notes: The estimating equation for Columns (1)-(2) is (2). The estimating equation for Columns (3)-(4) is (9). Observations are for each property transaction within the study period. Other controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type) and census tract characteristics (share of white, median income, median age, share of high school graduates). Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

nearby properties, suggesting a need for the triple difference model. The estimates of our DDD model are found in Columns (3) and (4) of Table 4. After controlling for zipcode fixed effects, we find a statistically significant decrease of 3.12% in properties within 0.36 to marijuana dispensaries.[24] The difference in coefficient estimates between DD and DDD model further underscores the importance of having a valid control group to address all endogeneity concerns, especially the potential selection bias in site selection *within* a neigborhood.

The estimation results of our IV model are described in Table 5. Sanderson–Windmeijer test

[24]Rather than cross-validating each model separately, we only report estimates using $r = 0.36$, the results of the CV procedure on Equation 2 as this is the model typically used in the literature. This keeps treatment groups consistent across models.

Table 5: Intent to Treat and Instrumental Variables Regressions

	ITT		IV		
	(1)	(2)	(3)	(4)	(5)
G_{it}^r	-.0227	-.0279**			
	(.0153)	(.0139)			
G_{it}^R	.0137	.0136			
	(.0086)	(.0085)			
D_{it}^r			-.0367	-.0446**	-.0296*
			(.0245)	(.0221)	(.0169)
D_{it}^R			.039	.038	.022*
			(.0258)	(.0257)	(.0119)
F-stat (D_{it}^r)			336.1	328.2662	601.8022
F-stat (D_{it}^R)			66.0029	66.0242	704.784
City FEs	Yes	No	Yes	No	No
Zipcode FEs	No	Yes	No	Yes	Yes
N	83,607	83,522	83,607	83,522	76,034

Notes: The estimating equation for Columns (1)-(2) is (13). The estimating equation for Columns (3)-(4) is (11). Observations are for each property transaction within the study period. Other controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type) and census tract characteristics (share of white, median income, median age, share of high school graduates). Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

statistics for first stage-weak instruments are shown in Columns (3)-(5). First stage results are significant and robust across all instrumental variables regressions. After controlling for zipcode fixed effects and instrumenting for treatment by using assignment to the treatment group, the estimated effect is -4.46%, qualitatively and quantitatively consistent with the DDD result.

As previously mentioned in Section 4.3, Chaisemartin and D'Haultfoeuille (2018) suggest using control groups where the distribution of treatment does not change over time in IV difference-in-differences models. Column (5) reports estimates of (11) using an alternative control group where for $G_i = 0$, $G_{it} = D_{it} = 0$. The estimates are smaller than in Column (4) though not significantly different and is almost exactly the DDD result. Furthermore, the intent-to-treat effect also is statistically significant at $p = 0.05$, implying a decrease of 2.79% for properties near lottery winning addresses.

Together, the results imply an estimated negative price impact of around 3% to 4.5%. For the average home sale price in our data, of $332,486, this implies a willingness-to-pay to avoid the disamentity of \$10,373-\$14,828, a non-trivial amount. It is however worth-noting that, while sizeable, the magnitude is comparable to other estimates in the economics literature. For example, Linden and Rockoff (2008) identifies a 4.1% drop in property values after the arrival of a sex offender in neighborhoods, an implied decrease of \$5,500 given median home prices in their area of study: Mecklenberg County, North Carolina. In Davis (2011), neighborhoods within two miles of a power plant experience a 3-7% decreases in housing values and rents while toxic plants lead to declines in property values by 11% for homes within 0.5 miles of the plants in Currie et al. (2015).

An important caveat to this analysis is that we only observe market prices for homes that sell and do not have data on the changing composition of neighborhoods after dispensary entry. Kuminoff and Pope (2014) state that using time series variation as in difference-in-difference estimation could fail identify the hedonic price function if neighborhood composition changes over time. Along similar lines, if the types of individuals in the housing market changes after cannabis firm entry, the observed prices are not may not reflect average willingness to pay. Rather, property sellers may be those that have a higher than average WTP to avoid dispensaries while buyers have a lower than average WTP. Data such as long-run demographic data or data on buyer and sellers would be needed study any neighborhood compositional changes.

However, as our analysis is short-term, it is unlikely that neighborhoods experienced huge changes in the time period studied. Nonetheless, identifying longer-run effects and studying changes in neighborhood composition remains an important area of future study, particularly if as attitudes evolve over time and citizens become more accustomed or hostile to nearby recreational marijuana firms.

Additional Results

To examine how varying the distance used to define the treatment group will impact the estimated effects, we report results with differing treatment radii for the DDD and IV models. Table 6 displays estimates for models that use 0.25, 0.30, 0.4, and 0.45 miles to define the treatment group. The estimates for both models vary in magnitude and precision based on the radii (though this is less of an issue for the IV model), showing the need for data-driven procedures such as cross-validation to define optimal radius.

We also explore how our choice of *post* period impacts our estimates. We define the post period

Table 6: Estimated Effects Varying r

	DDD				IV			
	0.25	0.30	0.40	0.45	0.25	0.30	0.40	0.45
D_{it}^{r}	-.0072	-.0069	-.0229*	-.0257**	-.0478	-.0427*	-.052**	-.0573**
	(.0225)	(.0204)	(.0131)	(.0112)	(.0422)	(.0248)	(.0252)	(.0262)
D_{it}^{R}	.0115	.0116	.0152*	.0158*	.033	.0361	.0407	.0435
	(.0086)	(.009)	(.0089)	(.0088)	(.0244)	(.0248)	(.0272)	(.0269)
F-stat (D_{it}^{r})					58.1278	141.458	270.802	265.23
F-stat (D_{it}^{R})					95.8151	75.795	82.965	78.507
N	83,522	83,522	83,522	83,522	83,522	83,522	83,522	83,522
R^2	.803	.803	.803	.803				

Notes: The estimating equation for Columns DDD is (9). The estimating equation for Columns IV is (11). Each column includes zip code fixed effects. Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

for our initial analysis as after the announcement of lottery winners, consistent with a model of forward-looking consumers. However, even though winning locations were well publicized, it could still be the case that dispensary location is not salient until after dispensary entry. (Response in the post-period appears delayed in Figure 7.) Therefore, we estimate our models using actual store entry as the post-period. A challenge in implementing this approach lies in defining the post period for neighborhoods that do not have a store: While the data has counterfactual locations, it does not have counterfactual entry dates. Therefore, we use the date of the first dispensary entrant in a jurisdiction as the counterfactual entry date for all lottery losers in that jurisdiction.

The results, as reported in Table 7, are qualitatively and quantitatively consistent with our baseline estimates. Specifically, while the magnitude of the coefficients all increase indicating that perhaps dispensary location is not initially salient, the estimates are not statistically significantly different from the original point estimates.

Differences by Neighborhood Demographics

To analyze how dispensary entry may have heterogeneous effects across neighborhoods, we stratify the sample by demographic variables and estimate Equation (9). The results are reported in Table

Table 7: Using Dispensary Entry as the Post Period

	DD		DDD		IV	
	(1)	(2)	(3)	(4)	(5)	(6)
D_{it}^{r}	-.001	-.0046	-.0512**	-.0544**	-.0543	-.0668**
	(.0134)	(.0129)	(.0202)	(.0254)	(.0377)	(.0339)
D_{it}^{R}	.0158**	.0154**	.0268***	.0261***	.0738*	.0755*
	(.008)	(.0065)	(.0103)	(.0092)	(.0416)	(.0401)
F-stat (D_{it}^{r})					268.0279	265.6987
F-stat (D_{it}^{R})					61.6181	62.7215
N	83607	83522	83607	83522	83607	83522
R^2	.775	.803	.776	.803		

Notes: The estimating equation for Columns DDD is (9). The estimating equation for Columns IV is (11). Each column includes zip code fixed effects. Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

8. We first study a sample split across education level, dividing by whether the property is located in a census tract where greater than 50 percent of the population has a Bachelor's degree. While the coefficient estimates suggest that properties in more educated tracts are the ones affected by dispensaries, an F-test cannot reject the null hypothesis of equality between the DDD coefficients. An F-test between the DDD estimates when the sample is divided by whether the property is in a tract that is above the 2014 median income of the state (around $61,000) also cannot reject the null of equality between the regression estimates.

Statistically significant differences between the DDD coefficients emerge when the sample is divided by whether or not the property's census tract is greater than 70 percent non-hispanic white. (The population of Washington is about 76 percent non-hispanic white.) Those properties in more diverse census tracts seem to drive the estimated decrease in property values while mostly white census tracts experience no change in housing prices. Moreover, census tracts where the tract's median age is below the state-wide median age of 37 also seem to drive the decline in sale prices after dispensary entry. Properties in older census tracts do not see a decline in prices.

Table 8: Estimated Effects by Demographic Group

	> 50 % Bachelor's		> Median Income		> 70% White		< Median Age	
	0	1	0	1	0	1	0	1
$\mathbb{1}(d_i \leq r) \cdot \text{Post}_{it}$	-.0258	-.0386***	-.0176	-.0467***	-.0552**	.0012	.0049	-.0462**
	(.0226)	(.0137)	(.0222)	(.0147)	(.0232)	(.0194)	(.0143)	(.0182)
$\mathbb{1}(d_i \leq R) \cdot \text{Post}_{it}$	.0123	.0163	.0142	.0101	.014	.0091	.0237**	.0065
	(.0125)	(.013)	(.0142)	(.0087)	(.0143)	(.0109)	(.0107)	(.013)
N	43905	39588	41516	41977	41819	41667	31004	52484
R^2	.66	.824	.683	.825	.798	.806	.818	.793
$F_{1,77}$	0.2		1.08		2.83		4.56	

Notes: The estimating equation is (9). Columns under the heading "> 50 % Bachelor's" splits the sample by whether or not the property is in a census tract where > 50 percent of individuals have a bachelor's degree. '> Median Income" refers to whether properties in census tracts where the median income exceeds the statewide median income. "> 70% White" references whether properties are in census tracts that have a population share of non-hispanic white that exceeds 70 percent. "< Median Age" divides the sample by whether the property is in a census tract where the median age exceeds the statewide median age. Observations are for each property transaction within the study period. Other controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type) and zip code fixed effects Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

6 Crime

Thus far we have remained agnostic to what could be driving any changes in home sale price. To that end, we analyze crime as a possible mechanism for depressed prices and use census tract-level police response data from the city of Seattle (which we discuss at length in Section A.2 in the Appendix) to estimate the following instrumental variables model:

$$crime_{jt} = \gamma_0 + \gamma_1 \cdot Dispensary_j \cdot Post_t + \gamma_t + \gamma_j + \omega_{jt}, \tag{15}$$

$$Dispensary_j = \rho_0 + \rho_1 \cdot Winner_j \cdot Post_t + \rho_t + \rho_j + \upsilon_{jt}, \tag{16}$$

where $crime_{jt}$ is the number of police responses per 10,000 residents of census tract j at month-year t. The variables γ_t and γ_j are month-year and census tract fixed effects, respectively. The time fixed effects control for any overall cyclical trends in crime while the census tract fixed effects control for any within tract crime trends. The variable $Dispensary_j$ is an indicator function equal to 1 if a dispensary locates in census tract j, and $post_t$ is equal to 1 after the license lottery announcement.

The coefficient of interest is γ_1.

As before, it is likely that dispensary entry in a census tract is correlated with unobserved variables that impact crime rates. Therefore, we instrument for $Dispensary_j$ with $Winner_j$, an indicator function equal to 1 if a license applicant address in j won a retailer license.

The results are reported in Table 9. We find that the number of police response reports decreases by around 13 per 10,000 census tract residents—a magnitude consistent with Brinkman and Mok-Lamme (2019), but this decrease is not statistically significant at 10% level. After sub-dividing crimes by category, drug-related crimes experience a small but statistically significant decrease while no significant change is found for property crimes and violent crimes. In constrast, nuisance-related crime reports such as loitering, disturbing the peace, or traffic crimes increase by 4.2 per 10,000 residents.

We also estimate Equation (15) using only census tracts that neighbor those tracts with dispensaries, to study possible spillover effects. These adjacent tracts see small increases in nuisance crimes as well as a small increase in violent crimes.

As these changes to crime in adjacent tracts do not directly correspond with decreases in crimes of directly treated census tracts, our findings are not consistent with the some-what prevalent belief that crimes are "displaced" by marijuana dispensaries. Rather, they are more consistent with a theory where marijuana dispensaries increase the number of people in an area, creating more unwanted disturbances. These perceived nuisances may make neighborhoods more unpleasant places to live, which in turn may induce a drop in property values. However, if the greater number of people is a consequence of increased economic activity generated by dispensaries, then these costs and benefits must be carefully weighed when implementing any local policies intended to mitigate any negative externalities caused by marijuana firms.

7 Conclusion

While legalization is often voted on by citizens and legislatures at the state- or national-level, municipalities and neighborhoods are left to deal with any localized negative effects of marijuana businesses. Therefore, significant research on the causal effects of cannabis businesses on neighbor-

Table 9: Crime Rates

	All	Nusiance	Drug	Property	Violent
Panel A: Treated Tracts					
Treated	-13.39	4.216**	-2.78***	-10.88	2.308
	(9.543)	(2.104)	(.9859)	(7.378)	(2.045)
N	2728	2728	2728	2728	2728
F	387.4583	387.4583	387.4583	387.4583	387.4583
Panel B: Adjacent Tracts					
Adjacent	2.692	1.813*	.888	-1.221	2.46**
	(4.697)	(.9297)	(.7157)	(3.519)	(1.018)
N	6200	6200	6200	6200	6200
F	454.58	454.58	454.58	454.58	454.58

Notes: The estimating equation is (15). Each column includes month-year as well as census tract fixed effects. Robust standard errors are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

hoods is needed to inform local public policy.

To this end, we study the impact of cannabis dispensaries on surrounding property values. A recreational marijuana retail license lottery held in Washington state provides plausibly exogenous variation to neighborhoods that were affected by marijuana retailer entry. Further, because participants in the license lottery were required to submit potential addresses due to location restrictions, we have novel data on both actual entrants' addresses and the addresses for license lottery winners and losers. This allows us to estimate difference-in-differences, triple difference, and instrumental variables empirical models. In addition, we develop a data-driven cross validation procedure to determine the optimal radius from the dispensary to study any effects.

Our estimates indicate a marijuana dispensary decreases property values by 3%-4.5% for homes within 0.36 miles of the retailer, a decline of about $10,000-$15,000 based on the average home values in Washington. These results imply a high willingness to pay to avoid the local negative externalities.

A hypothesized mechanism driving the decline in property values is crime around dispensaries. To investigate this, we study changes in police reports after recreational dispensary entry in the Seattle, WA. While we find limited evidence of a general decrease in overall crime and statistically significant evidence of a decrease in drug-related reports, we estimate that nuisance-related crime reports increase by about 4.2 per 10,000 census tract residents. Moreover, we find evidence that violent crime slightly increases in census tracts adjoining those where dispensaries locate.

Our findings suggest that crime risk could be a contributing factor to the negative price impacts of dispensaries, but that is likely only a partial explanation. Research on the localized impact of cannabis businesses, particularly studies that focus on long-run effects, remains a vital area of research.

References

Adda, J., B. McConnell, and I. Rasul (2014). Crime and the Depenalization of Cannabis Possession: Evidence from a Policing Experiment. *Journal of Political Economy 122*(5), 1130–1202.

Anderson, D. M., B. Hansen, and D. I. Rees (2013). Medical Marijuana Laws, Traffic Fatalities, and Alcohol Consumption. *Journal of Law and Economics 56*(2).

Anderson, D. M., B. Hansen, and D. I. Rees (2015). Medical Marijuana Laws and Teenage Marijuana Use. *American Law and Economics Review 17*(2).

Autor, D. H., C. J. Palmer, and P. A. Pathak (2014). Housing market spillovers: Evidence from the end of rent control in cambridge, massachusetts. *Journal of Political Economy 122*(3), 661–717.

Azofeifa, A., M. E. Mattson, G. Scauer, T. McAfee, A. Grant, and R. Lyerla (2016). National Estimates of Marijuana Use and Related Indicators – National Survey on Drug Use and Health, United States, 2002-2014. *MMWR Surveillance Summaries 65*(11).

Black, S. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics 114*(2), 577–599.

Brinkman, J. and D. Mok-Lamme (2019). Not in My Backyard? Not So Fast. The Effect of Marijuana Legalization on Neighborhood Crime. *Regional Science and Urban Economics 78*.

Burkhardt, J. and M. Flyr (2018). The effect of marijuana dispensary openings on housing prices. *Contemporary Economic Policy*.

Campbell, J. Y., S. Giglio, and P. Pathak (2011). Forced sales and house prices. *American Economic Review 101*(5), 2108–31.

Caulkins, J. P., B. Kilmer, and M. A. Kleiman (2016). *Marijuana Legalization: What Everyone Needs to Know?* (Second ed.). Oxford University Press.

Chaisemartin, C. d. and X. D'Haultfoeuille (2018). Fuzzy Differences-in-Differences. *The Review of Economic Studies 85*(2), 999–1028.

Chang, T. Y. and M. Jacobson (2017). Going to pot? the impact of dispensary closures on crime. *Journal of Urban Economics 100*, 120–136.

Chay, K. Y. and M. Greenstone (2005). Does Air Quality Matter? Evidence from the Housing Market. *Journal of Political Economy 113*(2), 376–424.

Cheng, C., W. J. Mayer, and Y. Mayer (2018). The Effect of Legalizing Retail Marijuana on Housing Values: Evidence from Colorado. *Economic Inquiry 56*(3).

Conklin, J., M. Doup, and H. Li (2018). Contact High: The External Effects of Retail Marijuana Establishments on House Prices. *Real Estate Economics*.

Currie, J., L. Davis, M. Greenstone, and R. Walker (2015). Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings. *American Economic Review 105*(2), 678–709.

Davis, L. W. (2005). The Effect of Health Risk on Housing Values: Evidence from a Cancer Cluster. *American Economic Review 94*(5), 1693–1704.

Davis, L. W. (2011). The Effect of Power Plants on Local Housing Values and Rents. *Review of Economics and Statistics 93*(4), 1391–1402.

Diamond, R. and T. McQuade (2019). Who wants affordable housing in their backyard? an equilibrium analysis of low-income property development. *Journal of Political Economy*.

Figlio, D. N. and M. E. Lucas (2004). What's in a Grade? School Report Cards and the Housing Market. *American Economic Review 94*(3), 591–604.

Freisthler, B., P. J. Gruenewald, and J. P. Wolf (2015). Examining the Relationship Between Marijuana Use, Medical Marijuana Dispensaries, and Abusive and Neglectful Parenting. *Child Abuse and Neglect 48*, 170–178.

Freisthler, B., W. R. Ponicki, A. Gaidus, and P. J. Gruenewald (2016). A micro-temporal geospatial analysis of medical marijuana dispensaries and crime in long beach, california. *Addiction 111*(6), 1027–1035.

Greenstone, M. and J. Gallagher (2008). Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program. *The Quarterly Journal of Economics 123*(3), 951–1003.

Huber III, A., R. Newman, and D. LaFave (2008). Cannabis Control and Crime: Medicinal Use, Depenalization and the War on Drugs. *The B.E. Journal of Economic Analysis and Policy 16*(4).

Kepple, N. J. and B. Freisthler (2012). Exploring the ecological association between crime and medical marijuana dispensaries. *Journal of Studies on Alcohol and Drugs 73*(4), 523–530.

Kuminoff, N. and J. C. Pope (2014). Do "Capitalization Effects" for Public Goods Reveal the Public's Willingness to Pay? *International Economic Review 4*(4), 1227–1250.

Linden, L. and J. Rockoff (2008). Estimates of the Impact of Crime Risk on Property Values from Megan's Laws. *American Economic Review 98*(3), 1103–1127.

Lu, R., D. Willits, M. K. Stohr, D. Makin, J. Snyder, N. Lovrich, M. Meize, D. Stanton, G. Wu, and C. Hemmens (2019). The Cannabis Effect on Crime: Time-Series Analysis of Crime in Colorado and Washington State. *Justice Quarterly*.

Mair, C., B. Freisthler, W. R. Ponicki, and A. Gaidus (2015). The Impacts of Marijuana Dispensary Density and Neighborhood Ecology on Marijuana Abuse and Dependence. *Drug and Alcohol Dependence 154*, 111–116.

Muehlenbachs, L., E. Spiller, and C. Timmins (2015). The Housing Market Impacts of Shale Gas Development. *American Economic Review 105*(12), 3633–59.

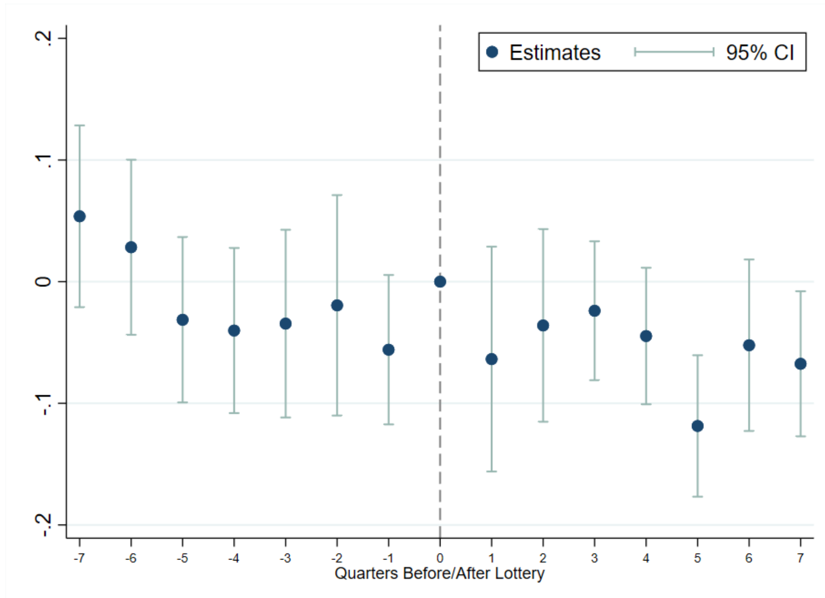
Pope, D. G. and J. C. Pope (2015). When Walmart Comes to Town: Always Low Housing Prices? Always? *Journal of Urban Economics 87*(5), 1–13.

Shi, Y. (2016). The Availability of Medical Marijuana Dispensary and Adolescent Marijuana Use. *Preventative Medicine 91*, 1–7.

Appendix

A.1 Additional Tests of Pre-Trends

Figure A.1: Event Study



Notes: Controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type) and zip code fixed effects Robust standard errors clustered at the jurisdiction-level are in parentheses.

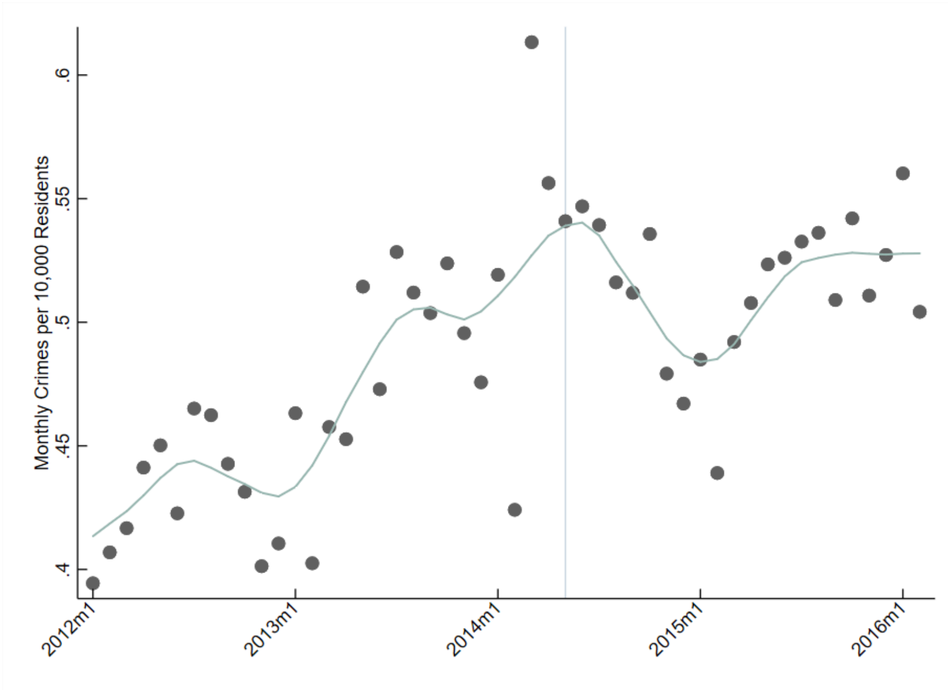
A.2 Seattle Crime Data

Data on reported crimes comes from the City of Seattle Open Data Portal. This comprises all reports to the Seattle Police Department (SPD) and includes the type of offense (e.g., burglary), along with a summarized description of the event (e.g., vehicle robbery), date of the police report, the date the crime occurred, and the location of the offense at the census tract level and block level.[25] We merge the crime data with the previously described data 2014 ACS data from the U.S. Census Bureau in order to calculate crimes per 10,000 residents.

Using the TIGER/Line shapefiles, we also match lottery applicants and operating dispensary addresses to their corresponding census tracts. This enables us to find crime rates in Seattle census

[25]The City of Seattle recently stripped the publicly available data of block and census tract information.

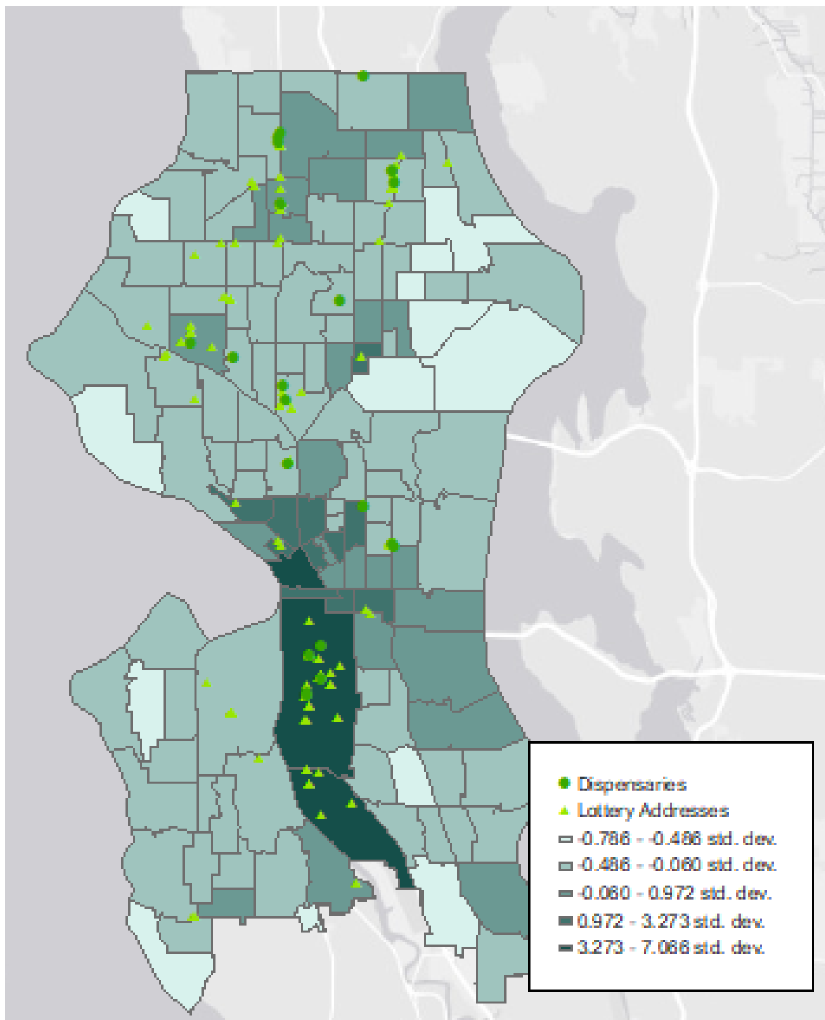
Figure A.2: Crime Rates in Seattle



Notes: The dots show monthly crime rates per 10,000 residents of Seattle. The solid line corresponds to a two-month moving average of crime rates. The vertical gray line denotes the lottery announcement month.

tracts that are near marijuana retailers. Further, we limit the analysis to census tracts that contain a lottery address or an operating dispensary or adjoin a census tract contain a lottery address or dispensary. Figure A.2 shows the evolution of crime rates over time while Figure A.3 provides a heat map of police reports in Seattle census tracts along with the locations of dispensaries and lottery addresses. As seen in the map, marijuana dispensaries tend to locate in areas with a higher number of police reports per 10,000 residents. We also separate out crime rates by types of crime separating out violent, property, drug-related, and nuisance crimes. Table A.1 explains how we categorize these crimes in detail.

Figure A.3: Crime Rates in Seattle Census Tracts



Notes: The heat map displays standardized crime rates (police reports per 10,000 residents) for each census tract in Seattle.

Table A.1: Crime Classifications

Offense Type	Violent	Property	Drug-related	Nuisance
Animal				X
Assault	X			
Burglary		X		
Disorderly Conduct				X
Disturbance				X
Drive-by		X		
DUI			X	
Fireworks				X
Gambling				X
Harassment				X
Homicide	X			
Illegal Dumping				X
Liquor Law Violation		X		
Loitering				X
Narcotics (Possession, Production, etc.)			X	
Pornography				X
Property Damage		X		
Property Stolen		X		
Prostitution				X
Reckless Burning				X
Robbery	X			
Theft		X		
Threats	X			
Traffic				X
Trespass		X		
Public Urination/Defecation				X
Vehicle Theft		X		
Weapon	X			