The Effect of Correctional and Rehabilitation Facilities on Property Values

Allyson Emblom-Hooe

North Carolina State University

Abstract

The number of correctional facilities has surged by over 240 percent since 1980. Yet little research has focused on their unintended consequences, particularly how this affects nearby property values. To address this gap, I create a novel dataset that combines data on correctional and rehabilitation facilities with micro-level repeated sales data that overcomes omitted variable bias. I use the hedonic method in the form of a triple-difference to investigate the impact of these facilities on property values. I find that a new facility significantly decreases nearby property values by an average of 20 percent. Moreover, I find heterogeneous effects by the type of facility. These findings highlight the complex interplay between criminal justice policies and the unintended repercussions of prison expansion.

JEL Classifications: O1, R1, R2, R3 Q5

December 8th, 2024

1 Introduction

Between 1960 and 1980, violent crime rates surged by 270 percent (BJS, 2019), leading to a significant increase in prison construction. In response, over 2,000 correctional and rehabilitation facilities have been built across the country. This rapid expansion has sparked widespread debates about their effects on local communities, raising questions about both the economic and social consequences of these institutions.

In the last four decades, the growth of the prison population in the United States has been the fastest in the country's history. With an almost five-fold increase from 1980, the U.S. prison system in 2010 held over 1.6 million individuals. Simultaneously, federal and state governments have funded massive prison infrastructure projects to meet the growing supply of prisoners (BJS, 2010).

Moreover, both state and federal governments responded by implementing a series of legislative measures that significantly lengthened sentences and intensified law enforcement efforts targeting lower-level crimes and drug-related violations. Key policies, such as mandatory minimum sentencing and the abolition of parole also played pivotal roles in fueling the prison boom of the 1980s. Overall, the tough-on-crime policies of the 1980s set the stage for a rapid expansion of the prison system.

But surprisingly, few studies have attempted to measure how property values are impacted by this rapid expansion, and none at the micro-level. The decision to open a facility has understudied social costs and benefits. In this paper, I look closer at these externalities and determine whether households view a new facility as an amenity that will generate economic growth or as a disamenity that will lower property values and push households to move away.

Opponents of the prison boom cite repercussions such as falling property values as well as racial and income inequities. Lower income and minority communities for example tend to have lower mobility that would lead to a geographic concentration of poverty (Kirchhoff, 2010). On the other hand, proponents of more prisons cite positive externalities such as more job opportunities and investments. This is known as the boomtown phenomenon. New infrastructure would increase the population, encourage businesses, and even improve complementary infrastructure such as roads (Cherry & Kunce, 1991).

My goal is to assess whether living near correctional and rehabilitation facilities is perceived negatively by nearby residents. I measure this by analyzing changes in property values when a facility opens in the area. A decline in property values would indicate reduced demand for housing in those neighborhoods.

Additionally, I explore heterogeneous effects of different types of facilities on property values. The effect of a prison opening may be more or less pronounced based on the size of the facility. For example, a halfway house is likely to be viewed differently to a maximum-security prison. A detention center is likely to be viewed as being different from a drug treatment facility.

I use the hedonic model in the form of a repeated sales triple-difference design. Each repeated sale is at the house level. They comprise a novel dataset that I collected by emailing individual counties in the Southeastern United States. I find that on average, households near a newly opened facility reduce in value by 20 percent. This magnitude changes depending on the type of facility a house is nearby and whether a local area is classified as rural or urban.

Although the property values of households near a facility decrease, I also show that as the distance increases, property values increase relative to far away. This reflects a short-term change in market conditions. Since facilities have the potential to increase employment, there may be an increase in demand for houses closer to a facility but not right next to it. Furthermore, an increase in property values further away from a facility may be indicative of households that are near moving further away which is also reflected in higher property values.

It may seem intuitive why property values near a facility would decrease but numerous behavioural reasons underlie why this may be the case. These include perceived safety concerns, Not in My Backyard (NIMBY) attitudes, noise and visual concerns, and civic pride. That is, homeowners have expressed that they do not want to live in areas where non-law-abiding citizens reside (Engel, 2007).

One significant psychological aspect is the fear of escaped convicts. Schicor explains that local residents near a prison often fear escaped inmates, believing that the escapees will incite harm (1992). Although these concerns are grossly exaggerated, media exposure can fuel this fear when escapes do occur.[1]

Additionally, there is the concern that a large number of families of convicted individuals will move into the local area, known as "camp following". However, research has also found that there is not a large influx of camp followers as might be predicted (Millay, 1989).

Another factor potentially decreasing property values is that a large facility alters the local community's perception of what the community should look like. In rural areas especially, a newly built facility may seem to indicate increased urbanization, moving away from the pristine countryside that local residents value. Furthermore, McGee states that many rural residents do not want to live near individuals who are not considered law-abiding citi-

[1]This argument against prison siting is purely psychological. Some studies have even shown that crime rates have actually decreased near facilities (Hawes, 1985).

zens (1981). Overall, many concerns lead to a decrease in property values, heavily influenced by subjective considerations.

Lastly, NIMBY attitudes and Locally Undesirable Land Uses (LULUs) also play a role in decreasing property values. LULUs are land uses perceived as risky when conducted near residential areas, such as mining, waste incinerators, or hazardous landfills. While research has shown that the construction of new correctional facilities is popular there is a strong "emotional" revulsion against having these facilities nearby (Krause, 1991).

Chirakijja (2022) examines the impact of prison opening between 1990 and 2000 on local communities across the US. The main outcome variables are housing values and income at the county and census-tract level. The data used for the outcome variables comes from the Decennial Census for 1990 and 2000, looking at the Tract-level median value.

This paper departs from Chirakijja's work in three important ways. The first difference is that my data consists of individual transactions, which include the sale price, the sale date, and the address of the property. Thus, I am able to map the precise location of where the transaction takes place and how close the property is to a correctional or rehabilitation facility. This means that I am able to test out different distances to see at what point being "near" a facility loses its effect.

This difference in units of observations is similar to the difference between Greenstone & Gallagher (2008) and Gamper-Rabindran et al. (2011), who looked at the effects of the Superfund cleanup program on housing values. This program initiated the cleanup of the most severely contaminated sites on the National Priorities List. Gamper-Rabindran & Timmins look at the within-tract distribution of house prices using fine geographies. In contrast, Greenstone & Gallagher look at only the median price of the tract. These data differences affected their conclusions. Where Greenstone & Gallagher find that there was little to no effect on housing values as a result of this cleanup, Gamper-Rabindran & Timmins (2013) find that housing values appreciated between 19 and 24 percent.

Additionally, Banzhaf & Farooque (2013) test the strength of certain types of housing data that are commonly used to estimate the value of a local public good. That is, whether median values, self-reported values, or transactions data reflect the value of a local public good. The authors find that median housing values are the weakest among these types of data and do not represent local conditions.

Second, the Decennial Census data used by Chirakijja is collected by survey. The Census Bureau defines the value as the respondent's estimate of how much the property (house and lot, mobile home and lot (if lot owned), or condominium unit) would sell for if it were for sale." That is, this data is not reflective of actual transactions. It is reflective of how the owner of a house responds to what the house is worth. This likely causes unknown bias, especially when looking at housing values in the 1990s when sites such as Zillow, Trulia, or Redfin did not exist.

Third and finally, I look at heterogenous effects based on what type of facility it is and who the facility is owned by. As stated above, people's preferences may differ based on what type of facility they are living next to. In my results, I find evidence that the type of facility does influence the effect on property values.

My paper also overlaps with literature on the effects of amenities and disamenities on property values. For example, Currie et al (2015) use the opening and closing of 1,600 toxic plants and compare housing prices within different distances. Using a similar triple-difference design, Muehlenbach et al. (2015) estimate the impact of shale gas development on nearby property values. Although this development allows for extracting natural gas more efficiently, nearby homes are strongly negatively impacted by groundwater contamination from the extraction process.

2 Data

I estimate the effect of facility openings between 2005 and 2023 in the Southeast region. due to the availability of real estate data, as I will discuss in more detail in the following pages. Specifically, I will focus on Washington DC, Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, and Tennessee, which makes up the Southeastern region of the US. Kirchoff (2010) notes that the much of the increase in correctional facilities especially takes place in the Southeast region.

There are two types of data I use for my analysis. The first data required is information on correctional facilities and rehabilitation facilities. This includes information on the location of the facilities, when the facilities opened, as well as the type of facility.

The second crucial data needed are data on real estate transactions at the micro-level for the scope of the analysis (Southeast region). This requires information on the location of the single-family transaction, the price, the date the home was sold, as well as any additional housing characteristics that can help describe the home and the neighborhood. For the latter type of data, much of the data collected lacks rich information on each house. I overcome this issue by using repeated sales for single family homes to be able to observe a home with the same characteristics overtime.

2.1 Data Collection on Facilities

I gathered data on correctional facilities from the Census of State and Federal Adult Correctional Facilities (CCF) for the years 2005, 2012, and 2019.[2] The CCF survey is conducted by the Bureau of Justice Statistics approximately every 5 to 7 years [3].

According to the CCF survey documentation, facilities that were excluded from the census included privately operated facilities that were not exclusively for state or federal inmates, military facilities, Immigration and Naturalization Service facilities, Bureau of Indian Affairs facilities, facilities operated and administered by local governments, facilities operated by the U.S. Marshals Service, and hospital wings and wards reserved for state prisoners.

The data that I collected from the CCF includes the facility name, the address, the city and county it is located in, the zip code, and the state. Unfortunately, data on who it is owned by (private, State, Federal, or local), the type of facility, and when every facility has opened or if it has closed is incomplete and had to be manually checked.

I first classified each type of facility as minimum, medium, or maximum security; work camps; or rehabilitation facilities. There were many facilities such as non-profits, probation offices, and rehabilitation centers that the survey included that were not yet coded despite being classified in the survey documentation. The classifications I made for each facility were based on security features discussed in the CCF survey documentation. Additionally, when searching each facility on the internet, the classification was often given via government (state, local, federal) or private website.[4] In the results section, I separate work camps and also group work camps as a minimum facility to compare the results.

In the survey, community-based facilities included halfway houses, residential treatment centers, non-profits, restitution centers, and pre-release centers. I chose to call this category of facilities "rehabilitation facilities."

I grouped minimum, medium, and maximum security facilities together, as the definition of these types of facilities are closely related and depend mostly on security features and not on size. For example, maximum security facilities have a double-fenced perimeter and an armed tower or armed patrols. Medium security facilities have a single or double fence perimeter with armed coverage by towers or patrols. Minimum security facilities have a fence or "posted" perimeter. Some facilities have multiple classifications. For robustness checks, I separate each facility into their own category to look at how property values are affected. The idea is to see whether locals nearby care about the type of facility or if the facility is

[2] I also check data for the years 1995 and 2000 to make sure that my list is comprehensive. I originally considered going back prior to 1995 but the data is not as rich.

[3] https://www.icpsr.umich.edu/web/NACJD/series/67

[4] Some states had better documentation on each facility classification than others. Florida for example had the best documentation on all classifications of correctional facilities

any type of correctional facility.

Lastly, jails are categorized as separate groups, and I classify work camps as a minimum security facility. Jails should be its own category because many jails are adjacent to court houses and police stations. The amenity/disamenity effect could be ambiguous if a jail is attached to these types of facilities. The data for jails comes from the Census of Jail Facilities (2006, 2013, and 2019). I classify work camps as minimum security facilities because each state has their own language for what a work camp is. In some states, a work camp is denoted as “work camp” and in other states a work camp is denoted as a “transitional center” or something similar.[5] I show in the results section that classifying work camps as a minimum security facility leads to more intuitive results.

After classifying each facility in my dataset, an important next step is to find out when every facility in my dataset opened and if the facility has closed. This step is important because my core research question asks whether nearby residents respond to a facility opening nearby that is measured via property values. To check for openings, I looked up each individual facility and started with the correctional department website for this information. For larger facilities, this information tended to be well-documented on the correctional department website or in the local news of each county. I was also able to confirm the opening of large correctional facilities using Google Earth Pro and Google maps where I could see the before and after satellite images for any facility built between 1990 and present day.

For smaller facilities, I found the open and closed dates through several methods. For each facility, I looked up if there was an affiliated website. A website tended to give the date that the facility was opened. I cross-checked this information with sales date information from Zillow or similar real estate sites. If the sales date was within a year that the website stated the facility opened, then I assumed that the date was correct. Opening dates tended to be well documented on the websites of larger non-profit organizations.

For smaller facilities that did not have websites, such as halfway houses or rehabilitation centers, I exploited the fact that these facilities often had signs on the outside of the property. I used Google Street View to find when the facility had a sign and when it did not have a sign. This gave an approximation of when these smaller facilities opened or closed. I cross-checked this also with sales dates. Note that this method could only be used for facilities that opened between 2004 and the present day due to the availability of the Google Street View feature. This limitation determined the time-span of my sample.

Determining closing dates was more difficult and could not necessarily be confirmed with Google Earth Pro or Google Street View because closures do not necessarily mean that the

[5] This is also true for jails. In the state of Georgia and Kentucky, jails are denoted as detention centers. This emphasizes the importance of checking each type of facility in order to be correctly categorized.

infrastructure is torn down immediately. I relied on white paper reports from each state and county announcing prison closures as well as local news reports from each county.

Despite the extensiveness of this dataset, the CCF data does not include immigration detention facilities or youth detention facilities. Accordingly, I found a list of centers in each state and manually checked when they opened and closed using the same process explained above.

2.2 Summary Statistics of Facilities

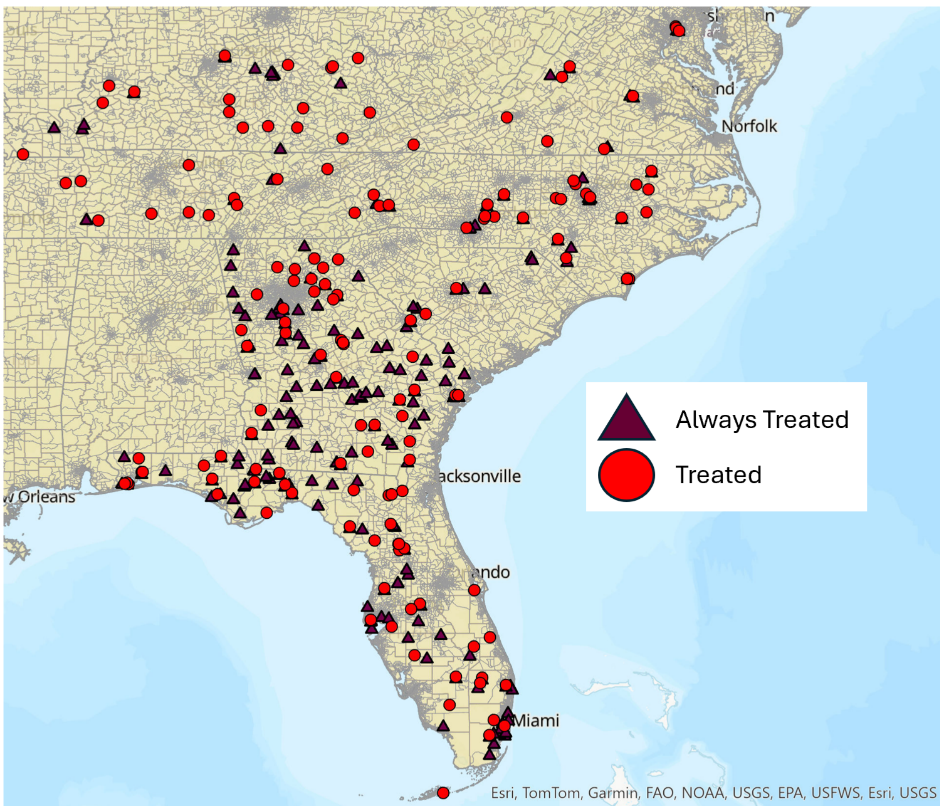
Since the core research question is how property values change when a facility opens nearby, I need a group of facilities that the treatment group can be compared to. That is, I need to match the treated facilities to a comparison group.

The treatment group consists of all facilities that have opened since 2005. As a way to match the treated group to a similar group that has not been treated in the treatment time period, I include what I call an "always treated" group that consists of all facilities that opened prior to 2005 and have remained open since, in the same county as the facility that has opened. This is a carefully designed method of comparison that will be able to directly incorporate a spatial component to compare property values which is discussed further in the methods section. If I compare a treated facility to an area with no facility, I have no baseline and no way to incorporate a spatial element to see to what extent homeowners nearby are impacted by facilities opening.

This method of defining a treatment group and an always-treated group is used because the placement of facilities is not random; counties with facilities differ systematically from those without. To address this, I compare the impact of a recently opened facility with a facility that has been operational for a longer time within the same county. This approach helps to control for county-level differences. By comparing property values within the same county, I can directly assess the difference in property values between a newly treated facility and an established, always-treated facility. Essentially, this comparison allows me to evaluate property value changes in areas where the values have already adjusted, focusing on the impact of the new facility compared to those that have been present for a longer period.

Figure 1 shows the distribution of treated and always treated facilities across the Southeast region. The red circles are treated facilities and the purple triangles are always treated facilities. This figure also shows that the majority of facilities are located in Florida, Georgia, and North Carolina. Many counties have both always treated and treated facilities within the same county and some of the counties have either a treated facility or an always treated facility. In this map, I show all facilities even if only one group resides in a single county.

Figure 1: Distribution of Facilities in the Southeast



This image shows the distribution of facilities across the Southeast. Red circles are treated facilities and Purple triangles are always treated facilities. The states included are Washington DC, Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, and Tennessee. Some red circles are directly on top of the purple triangles indicating that there are counties with both treated and always treated facilities.

The always treated group has a facility that opened prior to 2005. This means that an area with an always treated facility if opened prior to 2005 has already had time to adjust in the market. On the demand side, households have already made their decisions to stay or move and on the supply side, houses have already been adjusted in their value and have been sold on the market. The market is at a new equilibrium. Thus, the house values reflect the long-run effect from a facility opening nearby. When a facility opens in the treated group, this reflects short-run effects, as it catches up to the long run effects already reached in the comparison group.

Table 1 compares counties with an opened facility to counties that have never had a facility between 2005 and 2023 in the Southeast region using the ACS 2013 5-year estimates at the county level. This table motivates why using an always treated group should be used rather than a never treated group, as the means are different for many characteristics. The table includes counties that I was unable to obtain real estate data for and counties with only an always treated facility. The treated group in this case are all counties with facilities in the treated group and/or always treated group. The never treated group (control group in this case) are the counties that do not have facilities.

Table 1 shows that there is fundamentally a difference in the demographics between the two types of counties. The treated counties have a smaller white population, and a larger Black and Hispanic population.[6] Additionally, counties with a facility have a higher percentage of those in poverty and a lower employment percentage.

Table 2 similarly compares counties that have opening vs. closing facilities. It should be noted that some of the counties have both an open and closed facility.[7]

Counties with a closed facility are more densely populated, have a higher median income, a higher employment rate, a lower poverty rate, and are more educated. These summary statistics provide an overview on the difference in counties, but as stated above, county-level data is an aggregate of micro-data and does not paint a complete picture of what each county actually looks like. Even with many similarities at the aggregate, it is likely that the different types of counties have much more stark differences. The difference in county characteristics between counties with opened vs. closed facilities once again raises an interesting question on mechanisms. What pushes a facility to close? Is it a sign of economic growth if counties with a closed facility seem to be better off, or is there some unobservable causing the stark differences in characteristics? This is further explored in the results section.[8]

[6] Some of the differences in characteristics between counties could also be due to the large sample size of never treated vs. the smaller sample of treated.

[7] In the t-tests, I counted counties with both a opening and closing facilities twice. One was put in the treated category and one was put in the control category so that it would not affect the outcome.

[8] Chirakijja (2023) also explores the impact of prison closures on local communities and uses county-level

Table 1: T-test Results for Treated and Never Treated Groups

Variable	Mean (Never Treated)	Mean (Treated)	Difference in Means
Population Density	327.88	289.47	38.42
White	79.80	72.71	7.10***
Black	15.67	21.85	-6.18***
Hispanic	4.59	6.59	-2.00***
Less than HS	19.15	19.66	-0.51
High School	35.30	34.68	0.62
Bachelors	11.43	11.50	-0.07
Doctorate	0.76	0.77	-0.01
Not Enrolled	5.82	7.74	-1.91***
Labor Force	57.04	55.74	1.30*
Employed	50.49	48.86	1.63***
Unemployed	6.00	6.34	-0.34**
Not in LF	42.96	44.26	-1.30*
Employed in Business	11.27	11.29	-0.01
Employed in Fishing	1.41	1.81	-0.40**
Employed in Construction	10.84	10.17	0.67***
Median Income	42922.12	41276.10	1646.02
Public Assistance	2.44	2.22	0.22**
Gini Index	0.45	0.45	-0.01***
Housing Vacant	16.81	17.37	-0.56
For Rent	15.09	16.43	-1.34
For Sale	9.94	10.09	-0.16
Year Built	1980.89	1982.55	-1.66***
Gas	29.90	26.92	2.98**
Electricity	60.87	67.20	-6.33***
Coal	4.59	2.95	1.63***
No Fuel	0.31	0.47	-0.16***
In Poverty	17.96	19.56	-1.60***

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2: T-test Results for Counties with Opened or Closed Facilities

Variable	Mean (Closed)	Mean (Opened)	Difference in Means
Population Density	638.52	223.32	415.21***
White	69.31	73.17	-3.86
Black	23.69	22.15	1.54
Hispanic	9.69	6.17	3.53***
Less Than HS	16.59	20.37	-3.78***
High School	31.78	35.79	-4.02***
Bachelors	14.48	10.78	3.70***
Doctorate	1.10	0.74	0.36**
Not Enrolled	7.24	8.18	-0.94
Labor Force	58.61	54.59	4.02***
Employed	51.36	47.96	3.40***
Unemployed	6.59	6.27	0.32
Not in LF	41.39	45.41	-4.02***
Employed in Business	12.56	10.99	1.57***
Employed in Fishing	1.54	1.76	-0.23
Employed in Construction	9.25	10.28	-1.03**
Median Income	44846.65	40257.56	4589.08***
Public Assistance	2.08	2.26	-0.18
Gini Index	0.46	0.45	0.01**
Housing Vacant	16.62	17.60	-0.98
For Rent	18.48	15.99	2.49
For Sale	10.24	10.21	0.03
Median Year Built	1983.11	1982.85	0.26
Gas	24.63	27.81	-3.17
Electricity	70.56	67.51	3.05
Coal	1.67	2.92	-1.25**
No Fuel	0.84	0.40	0.44***
In Poverty	18.40	20.04	-1.65

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3: The Number of Facilities by State and Group

State	# of Always Treated	# of Treatment Facilities
DC	3	4
FL	79	48
GA	94	40
KY	14	17
NC	19	28
SC	15	4
TN	4	11
VA	5	7
Total	233	159
# of Counties	130	121

The tables show demographic characteristics among counties in different groups to motivate the comparison group and the difference between openings and closures. The following tables now provide a summary of what types of facilities are in my dataset. This includes a breakdown of facilities by state and facilities by owner that will be explored more in the robustness checks.

Table 3 provide a summary of the total number of facilities. This table is broken down by state, to show which states have the largest number of facilities, and by treatment and the always treated group.

Table 3 shows that there are 121 "treated" counties that have a facility that opened or closed between 2005 and 2023.[9] The "always" treated facilities are the facilities that opened prior to 2005 and have remained open. Most of the always treated facilities are located in the same county as the treatment facilities.

In addition to Table 3, which shows the number of facilities in each state, I also provide summary statistics of how many facilities have opened or closed in each state for the treatment group. This is important because openings may result in lower neighboring property values while closings may result in higher neighboring property values or null changes. This table is located in the appendix (Table A3).

One contribution of this paper is to show how market values change based on the type of facility that opens. Thus, Table A1 provides a breakdown of the sample size of each type of facility that has opened. For robustness checks, I also look at how property values change when a facility closes. Table A2 provides summary statistics of the owners of each facility (private, state, federal, or local)

data with employment data as the outcome variable whereas I focus on property values.

[9]In some cases, I counted "opened" as when a facility was renovated and an entire new building was added on.

2.3 Real Estate Data Collection

In addition to rich data on the specific facilities, I need single-family residential sales records from each individual county in my dataset that has a correctional or rehabilitation facility.

To each county in the Southeast region, I sent an email to the GIS department, Property Assessor, Tax department, or the recorder of deeds to ask if there is publicly available data on single-family residential sale transactions in the county three years before a facility was opened/closed and three years after.[10] I specified that the transactions must be only qualified or arms-length transactions to eliminate transactions that occurred between family members for example. Of the 212 counties that I sent an email to in the Southeast region, 181 counties responded with data. To all counties that did not respond to my first email, I sent a second email and a third email. After the third email, I categorized that the county did not have the data.

Single-family homes are ideal for a hedonic model because it is easy to compare attributes and their effects on prices. They also represent a major market segment with clear consumer preferences based on number of bedrooms, bathrooms, square footage, etc. On the other hand, multi-family and commercial properties are less suitable due to their varied sizes, layouts, and uses, making it harder to pinpoint how specific attributes influence their values. Additionally, the market dynamics for these properties are more complex, involving factors like rental income and occupancy rates that differ from single-family homes.

Rental rates in particular may be more complicated to work with due to ambiguity about key rental-contract features, such as which party pays for utilities and maintenance. The short-term nature of rental contracts may also weaken the incentive for renters to become fully informed about local amenities prior to entering the market (Bishop et al., 2020). An assumption of this analysis is that homeowners are aware that a facility is opening (or closing) nearby.

I use data three years before and three years after to capture the before and after effect of a facility opening and closing on property values. This broader time-span would also account for whether there was an announcement made of a facility opening or closing that would give a household time to make a decision to move or stay before the policy shock even happened. Chirakijja (2022) suggests that the announcement of a facility opening took place around 1-2 years prior.[11] Lastly, since I exploit distances between facilities and properties, I show the sample sizes of real estate data at different distances to each facility. These summary

[10] I sent emails to larger counties across the US and counties from the Southeast had the highest response rate.

[11] Local neighborhoods are informed of these openings by public hearing meetings, local news and media, direct notifications, and online postings.

statistics can be found in the appendix, specifically Table A8 and Table A9

2.4 Summary Statistics of Real Estate Data

A potential source of selection bias is that the 31 counties that did not respond to my data request could have different county characteristics than the counties that did respond. Fortunately, using 5-year estimates from the 2013 American Community Survey, I find that this is not the case. Table 4 shows county characteristics of the counties that responded to my inquiry for data and the 31 counties that did not across the Southeast region. Surprisingly, there is no statistical difference in population density or most other variables between the two types of counties. In Table 4, Group 0 are the counties with no real estate data and Group 1 are the counties with available real estate data.[12]

With all combined and cleaned real estate data, I next find the percent of data that has housing characteristics.[13] This is important because, as I will discuss in my methodology section, the hedonic model relies on having a complete dataset of housing characteristics.

The downside of sending out emails to different counties is that the data collection process is not centralized. That is, I do not receive the same types of files from each county. The extent of the uniformity of the data that I received what that all data included the address of the home, the sales price, and the sales date. However, the number of household characteristics often varied as well as the file type and the variable names. Additionally, as I discuss in the appendix, some files did not include the city or zip code of the address, which is vital in order to accurately geocode each property. The data cleaning section in my appendix discusses how I addressed this issue.

Table 5 shows that the percentage of available housing data is quite poor. The best housing characteristics are Year Built at 58 percent available data and Acres at 78 percent available data. The appendix also shows housing characteristic availability broken down by state. The tables will show that the availability of housing characteristics across different states was random. Several counties within each state had rich data, but this was not statewide. To overcome this issue of poor housing characteristics, I use repeated sales in my main estimation sample. That is, sales that occurred more than once in the time period of my analysis. I have over 557,000 sales that were repeated in my dataset. Using repeated

[12] Some counties that had real estate data sent this data via PDF and ended up not being usable. Though not included in this analysis, many counties in Arkansas had the data but sent real estate data via PDF. Additionally, some states had the data but required me to be a resident of that state or to show up to buildings in-person to receive the data.

[13] I dropped all transactions below $10,000, and eliminated bedrooms and bathrooms that were coded as 0 or above 10. Additionally, I eliminated any anomaly numbers for bedrooms and bathrooms such as .25 or .45.

Table 4: T-test Results for Variables Comparing Real Estate and No Real Estate Groups.

Variable	Mean (Group 0)	Mean (Group 1)	Difference in Means
Population Density	146.10	314.16	-168.05
White	74.08	72.47	1.61
Black	18.95	22.35	-3.40
Hispanic	6.04	6.68	-0.64
High School Less	19.56	19.68	-0.12
High School	32.42	35.06	-2.64**
Bachelors	11.89	11.43	0.45
Doctorate	0.60	0.79	-0.19
Not Enrolled	6.01	8.04	-2.02
Labor Force	57.98	55.35	2.63
Employed	50.34	48.61	1.73
Unemployed	6.50	6.32	0.18
Not in LF	42.02	44.65	-2.63*
Employed in Business	11.44	11.26	0.18
Employed in Fishing & Agriculture	1.91	1.79	0.12
Employed in Construction	10.64	10.09	0.55
Median Household Income	42,455.03	41,073.06	1,381.97
Public Assistance	2.14	2.23	-0.09
Gini Index	0.45	0.46	-0.00
Housing Vacant	16.55	17.51	-0.96
For Rent	17.55	16.23	1.32
For Sale	10.31	10.06	0.25
Median Year Structure Built	1981.35	1982.75	-1.40
Gas	25.09	27.23	-2.14
Electricity	63.32	67.86	-4.55
Coal	4.29	2.72	1.57**
No Fuel	0.31	0.50	-0.18
In Poverty	18.22	19.79	-1.57

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Group 0 is no real estate data and Group 1 is real estate data group.

Table 5: Percent of Available Data for Each Housing Characteristic

Variable	Total Observations	% of Available Data
Year Built	903,234	58
Acres	1,227,272	78
Square Feet	536,028	34
Bathroom	408,512	26
Bedroom	316,980	20
Total Rooms	282,721	18
Stories	493,858	32

sales, I overcome the issue of time invariant unobservables.

3 Hedonic Literature: From Theoretical To Empirical

To start the discussion on the methodology design of this paper, I discuss the evolution of the hedonic model from its theoretical origins to empirical applications, transitioning from structural to reduced-form approaches.

The hedonic model has evolved significantly over the past few decades, shifting from structural methods and utility functions to more adaptable reduced-form approaches. Recent literature has expanded its use, incorporating not just difference-in-differences but also triple-difference methods.

To begin, as Palmquist (1992) reiterates from Rosen (1974), the hedonic model is split into two stages. In theory, Rosen's hedonic model of market equilibrium can be used to assess the welfare implications of changes in goods and services that are not explicitly traded in formal markets. The first stage of the hedonic method estimates the hedonic price schedule. In the first-stage, the theory begins with the intuition that different household types compete against each other for entry into the most "desirable" neighborhoods. This competition takes place through bids on housing. Thus, hedonic regressions can tell us something about the way different household types sort across locations.

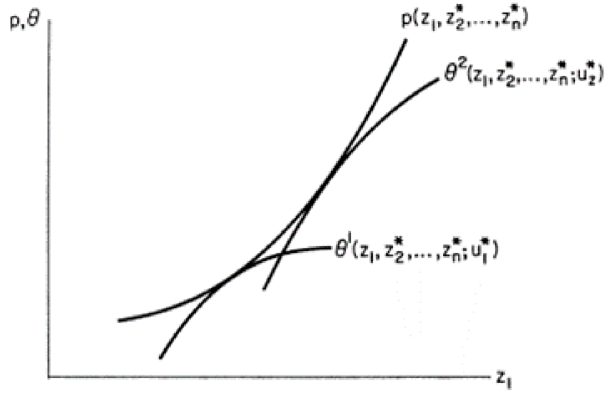
In this case, a bid function is the amount a household type would pay for housing at different levels of an amenity, holding utility constant. The hedonic envelope is the set of winning bids where each point is tangent to one of the underlying bid functions. This can be best seen in Figure 2. To summarize, the first stage is to estimate a hedonic regression (the envelope) and differentiate the results to find the implicit or hedonic price of the amenity.

The hedonic price function is estimated using information about the sales price of a differentiated good (housing in this case) and the characteristics.

In Figure 2, z is the amenity. In my methodological design, the amenity or disamenity to be estimated would be the distance to a correctional or rehabilitation facility holding constant the level of all other characteristics. θ indicates a bid function and p is the hedonic price function or the implicit price function. Rosen's framework is designed to consider heterogenous households or households with different demand for the amenity. If all households are alike, the hedonic price function is the same as the bid function. But with heterogenous households, we only observe one point – the tangency point – on each household type's bid function.

The slope of the hedonic price function at a point is the implicit price of z for a household choosing that point. Thus, maximizing households set the marginal benefit from z equal to

Figure 2: The Bid and Price Functions



their marginal cost. Households with different demand traits will have tangency points at different places on the hedonic price function.[14]

In Rosen's first stage, a major challenge to overcome is omitted variable bias. Many variables influence house values and leaving out key variables will bias estimated implicit prices and coefficients of interest. The problem is that there are many characteristics that matter to the household but that cannot be observed by the researcher. There are two ways to overcome this. The first approach is to use house fixed effects. Another is to collect extensive information on housing and neighborhood traits (Taylor, 2017). For my model, I use house fixed effects represented through repeated sales. As explained above, this is necessary since my data collection of single-family transactions does not include rich housing characteristics.

Kuminoff et al. (2010) discuss the concern of omitted variable bias and the importance of using the best functional form. The authors test whether omitted variables seriously undermine the method's ability to accurately identify economic values. The results suggest that moving from the standard linear specifications for the price function to a more flexible framework that uses a combination of spatial fixed effects, quasi-experimental identification, and temporal controls for housing market adjustment shows promising results. This conclusion is supported by Bishop & Timmins (2019). This is something that I also adhere to.

To estimate a hedonic model, quasi-experimental designs such as regression discontinu-

[14]Figure 2 depicts two types of households. Type 2 households have a strong taste for amenity z and choose to consume a high level of z. Type 1 households have weaker tastes for z and choose a lower level. In my case, z is the distance to a facility. This indicates that property values will be higher further away from a facility and lower when closer to the facility. If distance to a facility is the amenity, then Type 2 households have a stronger taste for z (distance) and choose higher levels. Type 1 households that live closer to a facility will bid at lower levels.

ity and difference-in-differences have become increasingly popular to estimate the change in housing prices as a result of an amenity or disamenity (Bishop et al., 2020). A primary concern with the hedonic method is that there are unobserved spatial characteristics that might correlate with amenities of interest. If there is selection into treatment based on unobservable differences between the treatment and control groups, estimated treatment effects will be biased. With quasi-experimental methods, the researcher can rely on a naturally occurring process, such as a natural event or a policy decision that assigns homes to treatment in an exogenous manner (i.e., in a way that is unrelated to the error term in the hedonic regression) (Taylor, 2017).

In line with the use of quasi-experimental methods to estimate housing price changes from environmental factors, the localized impact of toxic emissions from Currie et. al (2015) provides a key case study. The authors estimate that toxic air pollutants from industrial plants have highly localized effects, primarily within a one-mile radius. Plant openings lead to an 11% decrease in housing prices and a 3% increase in low birth weight incidence within this range. These findings reinforce the need for careful empirical strategies, like difference-in-differences, to isolate the effects of environmental disamenities.

Muelenbacks & Timmins (2015) for example combine hedonic theory with a quasi-experimental design to show the relationship between shale gas development and property values. The authors point out that a major endogeneity issue is that shale gas wells are not located randomly. Instead, they tend to be placed in areas with features that aid in the drilling process. The authors create a triple-difference (DDD) estimator. The first difference is the change in price of a particular home that has a new well pad. The second difference is the change in price for public water service properties adjacent to shale gas development to the change in prices of public water service properties not adjacent to shale gas development. The last difference must difference away the effects across public water service areas and ground water areas. The results demonstrate that groundwater-dependent homes are negatively affected by nearby shale gas development.

Linden & Rockoff (2008) research the relationship between property values and local crime risk. This is useful for measuring the willingness of individuals to pay to reduce their exposure to crime risk. Specifically, the authors look at the willingness to pay of households to not live near registered sex-offenders. The quasi-random experiment in this study is unique because it provides the ability to not only estimate a treatment effect on an implicit housing price, but also a reversal treatment effect. Like Linden & Rockoff (2008), Pope (2008) show that the implementation of Megan's Law also has a negative impact of housing values. This paper adds an additional robustness check of what happens to housing values when a sex offender moves away. Overall, they find that housing values do improve when this happens.

Kuminoff and Pope (2014) discuss what such quasi-experimental designs can identify. Typically, they measure the change in housing prices from a treatment relative to a counterfactual. Historically, such "capitalization effects" have been used interchangeably with measuring willingness to pay for a public good, but this is a misconception. When an exogenous shock affects the spatial distribution of a public good, the hedonic price function adjusts to clear the housing market. This creates a wedge between average capitalization effects and the average household's willingness to pay. Capitalization effects conflate willingness to pay with changes in the shadow price and other unobserved housing and neighborhood attributes. To accurately interpret price changes over time, it is essential to understand the evolution of the hedonic price function. Unfortunately, many studies have assumed a time-constant gradient, typically fixed for 10-20 years (Linden and Rockoff, 2008; Pope, 2008; Greenstone and Gallagher, 2008, etc.).

As noted by Kuminoff and Pope, adopting a capitalization framework alters the economic interpretation of these estimates, as the price schedule shifts over time. The total change in prices from an exogenous change in characteristics is not equivalent to willingness to pay; instead, it conflates willingness to pay with changes in the price function (Klaiber and Smith, 2013; Banzhaf, 2021).

If the equilibrium hedonic price function for a housing market changes endogenously due to a shock to amenities, house prices can change even if the amenities themselves have not. This scenario violates the Stable Unit Treatment Value Assumption (SUTVA), which posits that the price or outcome of an untreated house should not be affected by changes in other treated houses. When SUTVA is violated, the effect on untreated prices is termed the indirect effect (Banzhaf, 2021).

Banzhaf (2021) bridges the gap between quasi-experimental designs and the underlying structural model. The author demonstrates that the difference-in-difference hedonic model can reveal the "average direct effect" on prices resulting from changes in amenities, relative to a baseline that includes the indirect effect (where spillover effects may influence prices). This estimation is a conservative or lower-bound estimate on total welfare, suggesting that the actual welfare impact could be greater than indicated by the results.

Building on these results, this paper demonstrates that quasi-experimental designs, such as difference-in-differences, can provide meaningful welfare measures without relying on the hedonic second stage.

In my results section, I discusses potential SUTVA violations. For instance, when a facility is built near a treated area, homeowners may sort themselves further away, influencing the untreated or control areas used for comparison. As Banzhaf (2021) states, the estimates are a lower bound estimation on welfare which means that the results that I find are a more conservative estimate of the amenity's (or disamenity) impact.

The following Methodology section builds upon this hedonic literature review by validating the model using empirical analysis.

4 Methodology

To estimate the effect of the opening of correctional and rehabilitation facilities on property values, I use the hedonic method as discussed in detail in the previous section. More specifically, I design a triple-difference that expands on the traditional difference-in-differences model by incorporating an important spatial component that captures that micro effect of a facility opening.[15] The spatial component means that I can gauge the exact distances that individuals value or don't value living near a facility.

I start with providing equation (1) which is my triple-difference baseline equation. This has a triple interaction that finds the average effect on single-family sales prices. I break down each component of this triple-difference in detail.

$$\begin{aligned}\ln(P_{it\rho}) = \beta_0 + \rho_\rho + \mu_t \\ + \beta_1 \text{After}_t + \beta_2 \text{Near}_i + \beta_3 \text{Near}_i * \text{After}_t + \beta_4 \text{Treat}_{it} * \text{Near}_i + \beta_5 \text{Treat}_{it} * \text{After}_i \\ + \beta_6 \text{Near}_i * \text{After}_t * \text{Treat}_{it} + \beta_7 X_i + \epsilon_{it\rho}.\end{aligned} \tag{1}$$

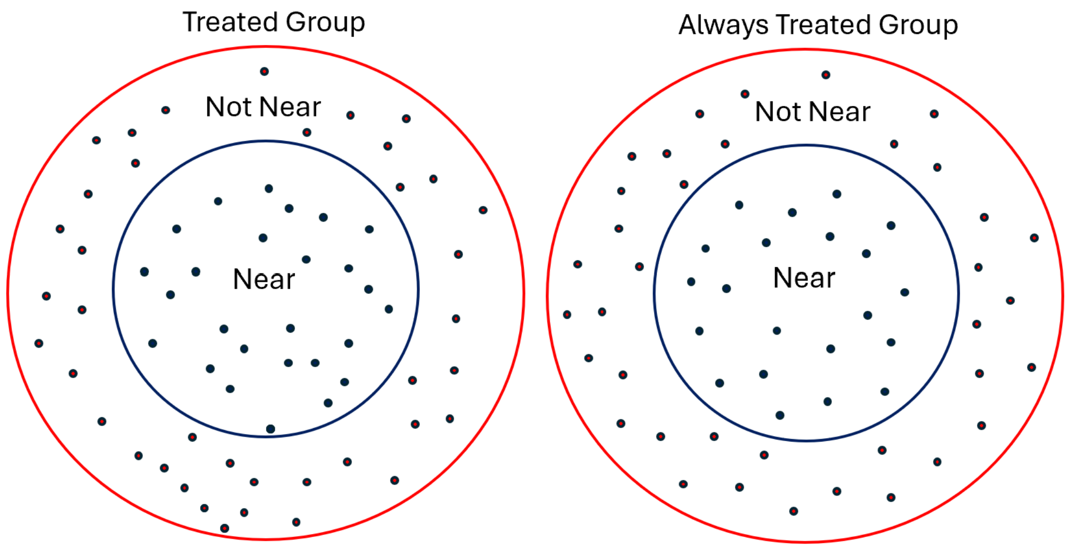
In this equation, i represents individual houses, t is a time subscript, and ρ is a facility subscript. The dependent variable is the natural log of the price that each household sold for. ρ_ρ represents facility fixed effects which absorb the Treat variable. μ_t are year-fixed effects that is represented by the year a sale took place.[16] X_i are housing characteristics such as acres, year built, the number of bedrooms, bathrooms, and stories.

As discussed earlier, a traditional hedonic model includes rich data on housing characteristics which as discussed in the data section, I do not have. This issue is overcome by using repeated sales. That is, houses that were sold multiple times in the time period of my sample. I assume that the composition of the house has not changed over time - that there have been no additions or torn down and rebuilt structures at each address.

[15]I use this same model for facility closures and other robustness checks.

[16]The results are the same if I incorporate month and/or day

Figure 3: Visualizing the Triple-Difference



Each facility is within the near circle. The comparison group to the near circle is the not near boundary. This compares similar areas - the only difference is that the smaller circle is close to a facility and the outer boundary is farther away. The before and after elements are not depicted in this image. There are 81,488 transactions that occurred both before and after a facility opens within these bounds.

Equation (2) is my repeated sales baseline equation.

$$\begin{aligned}\ln(P_{it\rho}) = \beta_0 + \lambda_i + \mu_t + \\ + \beta_1 \text{After}_t + \beta_2 \text{Near}_i * \text{After}_t + \beta_4 \text{Treat}_{it} * \text{After}_i \\ + \beta_6 \text{Near}_i * \text{After}_t * \text{Treat}_{it} + \epsilon_{it\rho}.\end{aligned} \tag{2}$$

The λ_i are house fixed effects represented by repeated sales. This fixes the issue of omitted variable bias with limited data on housing characteristics. λ_i absorbs Near, Treat, and Treat*Near. The μ_t captures time fixed effects, representing unobserved time-specific factors that affect the housing prices.

4.1 Components of the Triple-Difference Equation

To complement the above triple-difference equations, the following figure provides a visual explanation of each triple-difference component - Near, Treat, and After.

4.1.1 Near vs. Not Near

Each correctional and rehabilitation facility has been geocoded along with the location of every real estate transaction. This means that I can find the precise distance between a facility and a transaction. The first difference of near and not near compares the sales price of homes near a facility and not near a facility. In this case though, not near must be far enough away that it can be assumed there is no effect from the policy shock but not so far that the local characteristics would be different.

The definition of near depends on population density - that is, whether an area is considered to be urban or rural. For my analysis, I define urban as having a population density of more than 1,000 people per square mile and rural as having a population density less than 1,000 people per square mile. For example, if a house is within 1 mile of a facility where the Census Tract population density is greater than 1,000 people per square mile, a house is considered to be near a facility. If a house is within 5 miles of a facility where the Census Tract population density is less than 1,000 people per square mile, a house is also considered to be near a facility.

Not near also measures the distance between a facility and a home but is not so far away that the characteristics between the local area are different. For example, the distance between a facility and a house is considered not near if a house is between 1 and 3 miles away from a facility when the population density is greater than 1,000 people per square mile (urban). A house is also not near a facility if a house is between 5 and 10 miles away from a facility when the population density is less than 1,000 people per square mile.

In my analysis, I consider variations on how I define "near" and "not near" for rural areas keeping the definition the same for urban areas.

4.1.2 After vs. Before

After is a dummy variable that is 1 if a house is sold t+1 to t+3 years after a facility opens. After is 0 (representing before) if a house is sold t-3 to t (inclusive) years before a facility opens. This allows for homeowners to be informed about a facility opening before it even opens due to announcements that took place a year or two prior to opening. That is, even though a facility opens in period t, the decision could have been announced in period $t-2$. For the main results, I include t as "before". As a robustness check, I consider different definitions of after.

Note that there are 81,488 house transactions of the same house that occurred both Before and After a facility opens within a 20 miles radius.

4.1.3 Treated vs. Always Treated

Treat refers to all facilities that opened between 2005 and 2023. Treat is a dummy variable that is 1 if a facility opened in the study period or 0 if the facility is always treated. An always treated facility refers to all facilities that opened before 2005 that have remained open.[17,18] Every treated facility is matched to an always treated facility in the same county.

This design has been carefully constructed around using treated versus an always treated facility. If I used a never-treated area compared to a treated area, the spatial component would be arbitrary. That is, how would I compare distances between a facility opening and an area with no facility whatsoever?

Observations in the always treated group are near or not near a facility that opened prior to 2005. This means that this area has had time to adjust in the market. On the demand side, households have already made their decisions to stay or move and on the supply side, houses have already been adjusted in their value and have been sold on the market. The market is at a new equilibrium. Thus, the house values reflect the long-run effect from a facility opening nearby prior to 2005. When a facility opens in the treated group, this reflects short-run effects.

The difference when comparing a treated group and an always treated group reflects the difference in property values in the short run and long run. By using an always-treated control group, I can compare Near/Not Near and Before/After Difference-in-Differences across groups for an overall triple-difference analysis, similar to the approach used in Currie et al. (2015). This comparison is not feasible with a never-treated group.

4.2 Event-Study

Equation (3) shows how my results change before and after the "policy shock." That is, Equation (3) is a triple-difference event-study with staggered openings.

[17]There is very little overlap between treated and always treated. I used block group data and census tract codes from 2020 to check any overlap. Any facility with overlap was not included in the analysis at the block group level and should not have impact on the results.

[18]In my results, I change the always treated sample to cutoff after the year 2000 to avoid any overlap with properties. This has no impact on the results.

$$
\begin{aligned}
\ln(P_{it\rho}) = &\beta_0 + \lambda_i + \mu_t \\
&+ \sum_{\tau=-3}^{3} \alpha_\tau \text{After}_{t=\tau} + \sum_{\tau=-3}^{3} \beta_\tau (\text{After}_{t=\tau} \times \text{Near}_i) \\
&+ \sum_{\tau=-3}^{3} \gamma_\tau (\text{After}_{t=\tau} \times \text{Treat}_{it}) \\
&+ \sum_{\tau=-3}^{3} \delta_\tau (\text{After}_{t=\tau} \times \text{Near}_i \times \text{Treat}_{it}) + \epsilon_{it\rho}.
\end{aligned} \tag{3}
$$

Equation (3) follows Sun and Abraham (2020), Callaway and Sant'Anna (2020), and Goodman-Bacon (2021) for a staggered treatment event-study. Equation (3) will show how the coefficients change in the triple-difference both before and after a facility opens.

In this equation, τ represents time periods (years before and after). The coefficients δ_τ capture the interaction effects over time. Specifically, the terms $\sum_{\tau=-3}^{-1}$ and $\sum_{\tau=1}^{3}$ measure the effect of being near the treated site, both before and after the event. A significant δ_τ for post-treatment periods (e.g., $\tau = 1$) would indicate a notable change due to the treatment after it occurs. This helps assess dynamic effects over time, showing any anticipation or lasting impacts (in the short-run).

The results from equation (3) offer valuable insights into the effects of a facility opening, both before and after the event. They potentially reveal whether individuals begin to react to the announcement of a facility opening prior to its actual operation. Furthermore, the results shed light on the duration of the significance of these effects following the facility's opening.

4.3 Continuous Triple-difference

In this model, the spatial dimension is important. However, there are two differing implications when a facility opens nearby.

I hypothesize that property values decrease near a facility. However, I have not talked about whether housing values increase as they get further away. That is, there could be a positive effect as the distance increases for two reasons. The first reason could be population growth and job creation which creates more demand for houses in the local area - just not right next to a facility. The second reason is that perhaps houses near a facility moved away enough to no longer by near the facility, but close enough that they are still in the local area. Equation (4) shows a continuous version of Equation (2) to show how distances as a continuous variable change property values.[19]

Equation (4) presents distances as continuous instead of binary.

$$\begin{aligned}\ln(P_{it\rho}) = {} & \beta_0 + \lambda_i + \mu_t + \\ & + \beta_1 \text{After}_t + \beta_2 \text{LogDistance}_i * \text{After}_t + \beta_4 \text{Treat}_{it} * \text{After}_i \\ & + \beta_6 \text{LogDistance}_i * \text{Treat}_{it} * \text{After}_t + \epsilon_{it\rho}.\end{aligned} \tag{4}$$

The interpretation of this equation requires careful attention. β_4 would suggest that these are the results when a house is sold after a facility opens right next to it (when distance equals zero).[20]

Since it is also likely that decisions are spilling over in this experimental design I follow Banzhaf (2021) and interpret my estimates as a lower bound on welfare. One reason for spillovers is that facilities are a source of employment and it could be that workers come to live in the local area - just not right next to the facility (in the not near buffer).

5 Results & Discussion

5.1 Full Sample Results With and Without Characteristics

The following results come from Equation (1) (full sample results), which includes key housing characteristics such as year built, acreage, number of bedrooms and bathrooms, square footage, and the number of stories. Table 6 presents these findings.

In the first column of the table, there are 3,020 observations that reflect a rich dataset of housing characteristics. Although the coefficient for the triple-difference interaction - Near*Treat*After is statistically significant, the limited number of observations constrains my ability to draw strong conclusions regarding the impact of facility openings on property values.[21] This highlights a notable issue of statistical power within these results. To bolster the findings from this full sample, I employ a repeated sales model, which provides a more robust framework for analysis.

In the second column, I regress the same model without housing characteristics, yet the sample remains unchanged. Here, we observe a substantial decrease in statistical power, further complicating the analysis. By the time we reach the third column, which does

[19] In the results section, I show results for logged and not logged distances.

[20] Note that there is no distance of zero in my dataset - only distances very close to zero. The closest distance to zero in my dataset is .0008 miles. I also demean or center the continuous distance variable around the mean as done by Bakhtiar et. al (2023). This makes the effects easier to interpret and more stable across specifications. The results do not change.

[21] These results also only come from 6 Tracts

not include any housing characteristics (regressed on the full sample), the results show no significant relationships, making it difficult to ascertain any meaningful link between the opening of facilities and property value changes. This progression illustrates the necessity of incorporating rich housing characteristics, as the omitted variables lead to inconclusive results.

Despite these challenges, the coefficients for Model 1 indicate that property values decline by approximately 20%. All results are clustered at the tract level, in line with the methodological recommendations of Bishop et al. (2020).

The varying R-squared values across the models provide additional insights into their explanatory power. Model 1 demonstrates the highest R-squared, indicating that the inclusion of housing characteristics significantly enhances our understanding of the data. Conversely, Models 2 and 3 exhibit lower R-squared values, underscoring the drawbacks of omitting relevant variables, which ultimately diminishes the models' explanatory capacity.[22]

In summary, while the results from the full sample indicate interesting trends, the analysis reveals important limitations due to statistical power and variable inclusion. The following sections will address these issues by using Equation (2).[23]

5.2 Repeated Sales Results for Different Distances

Recall that the variable near is a dummy variable that categorizes different distance ranges based on whether a tract is classified as rural or urban. In the results presented below, I offer various definitions for how I distinguish between near and not near.[24]

While the coefficients from Equation (1) provide valuable insights, there remains potential for enhancement through the use of a repeated sales approach. This method examines properties that have undergone multiple transactions within the project time frame, specifically from 2002 to 2023.[25]

The forthcoming results incorporate both house fixed effects and year fixed effects. The house fixed effects are repeated sales of single-family homes - allowing for a more robust analysis of the data.

[22]Note: In Model 3, Treat is not absorbed. It could be that the variable Treat might have enough variation across observations in Model 3.

[23]There is very little overlap between types of facilities at the block-group level. Even so, I removed all facilities within the same block-group.

[24]It's important to note that the sample size for urban areas within the 0-0.5 mile range is too small, which is why this definition of "Near" is not applied to urban regions. Instead, I consistently define Near for urban areas as being within 0-1 mile.

[25]I opted to use data starting in 2002 instead of 2005 to ensure access to sales data from three years prior to the opening of any facility in 2005. It's worth mentioning that many facilities did not open after 2020, and by that year, several facilities in my dataset had already begun to close.

Table 6: Regression Results for Models 1, 2, and 3

	Model (1)	Model (2)	Model (3)
Near*Treat*After	-0.203***	-0.423	0.217
	(0.0411)	(0.494)	(0.139)
Near*Treat	-0.136	-0.231	-0.446**
	(0.164)	(0.240)	(0.184)
Near*After	-0.104***	-0.116***	-0.259*
	(0.0233)	(0.00697)	(0.133)
Treat*After	0.557**	0.950***	0.00324
	(0.169)	(0.0978)	(0.0891)
Near	0.0815**	0.171***	0.280*
	(0.0272)	(0.0102)	(0.152)
After	0.0831**	0.140***	-0.0266
	(0.0268)	(0.0212)	(0.0744)
Treat	-	-	0.143
			(0.154)
Year Built	0.0138**	-	-
	(0.00362)		
Acres	-0.203***	-	-
	(0.0184)		
Square Feet	0.000394***	-	-
	(2.85e-05)		
Bedroom	-0.00842	-	-
	(0.0388)		
Bathroom	0.0171	-	-
	(0.0134)		
Stories	0.0213	-	-
	(0.0407)		
Observations	3,020	3,020	174,250
R-squared	0.727	0.460	0.293

Robust standard errors in parentheses. Model (1) is the full sample with housing characteristics. Model (2) is a simplified sample without housing characteristics. Model (3) is the full sample without housing characteristics.

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Results for Different Distances

Variables	Distance (1)	Distance (2)
After	-0.147	-0.231***
	(0.0970)	(0.087)
Near*After	0.103	0.181***
	(0.0673)	(0.050)
Treat*After	0.0573	0.119*
	(0.0730)	(0.0696)
Near*Treat*After	-0.200**	-0.220***
	(0.0893)	(0.07995)
Observations	41,775	48,157
R-squared	0.786	0.792

Table 7 shows results when the distances vary. Distance (1) is defined as Near (0-1 Miles Urban, 0-5 Miles Rural) and Not Near (1-3 Miles Urban, 5-10 Miles Rural). Distance (2) is defined as Near (0-1 Miles Urban, 0-10 Miles Rural) and Not Near (1-3 Miles Urban, 10-20 Miles Rural).

In the following table, the definition of an urban area remains the same for each model. In an urban area, a house that is considered near to a facility is within 1 mile of the facility. Not near for an urban area is 1-3 miles.

Table 7 shows results based on how I define near and not near in a rural area keeping the urban definition the same. Distance (1) defines near as between 0-5 miles and not near as 5-10 miles. Distance (2) defines near as 0-10 miles and not near as 10-20 miles. Showing these different results highlight the distance definitions that I chose. I also provide sensitive tests that strengthen the explanatory power at different distances.

Table 7 shows Near (0-1 Miles Urban, 0-5 Miles Rural) and Not Near (1-3 Miles Urban, 5-10 Miles Rural) and Near (0-1 Miles Urban, 0-10 Miles Rural) and Not Near (1-3 Miles Urban, 10-20 Miles Rural) are both definitions that yield significant results. The coefficient of interest is Near*Treat*After. Both coefficients show that opening a facility has the unintended consequence of reducing housing values by around 20 percent. This shows that these types of facilities are in fact a disamenity and not an amenity to household that live nearby.[26]

These results have two interesting policy implications that provide the potential for future research. Firstly, if facilities reduce property values by this much, this has interesting policy implications on the location decisions of facilities not just in the Southeast region of the US but anywhere in the world. Rather than having facilities spread apart throughout a county

[26]These results are driven by Florida, Georgia, Kentucky, Tennessee, North Carolina, and Virginia. Many of the facilities in Washington DC closed between 2005 and 2023 and are included in the closed facility sample.

for example which in-turn reduces property values, maybe these results suggest that all these disamenities should be placed near each other rather than spread apart. That way the disamenity effect is contained to one area. However, I cannot make any judgements on the full effect of having a facility nearby without also looking at the positive impacts. For future work, I would also look at how employment has changed near a facility since job growth is one of the reasons why larger facilities especially are valued for the local community.

The second issue that these results showcase is that if the rate of incarceration continues to rise, this creates more demand for all types of facilities. My results establish that an unintended consequence is that nearby property values suffer. That is, households outside of the correctional facility system are affected negatively.

A policy implication from this could be that the demand for non-violent offenders should be reduced. According the the Federal Bureau of Prisons, around 62 percent of those incarcerated in 2024 were incarcerated for non-violent crimes. These category of crimes include Banking and Insurance, Counterfeit, Embezzlement, Burglary, Larceny, Property Offenses, Continuing Criminal Enterprise, Courts or Corrections, Drug Offenses, Extortion, Fraud, Bribery, Immigration, Miscellaneous, and Robbery (2024).

Although the results above lead to some interesting policy implications, the model uses an arbitrary definition to define what is considered Near and what is considered Not Near. Thus, I also provide a continuous distance model to show when the effect on property values disappears.

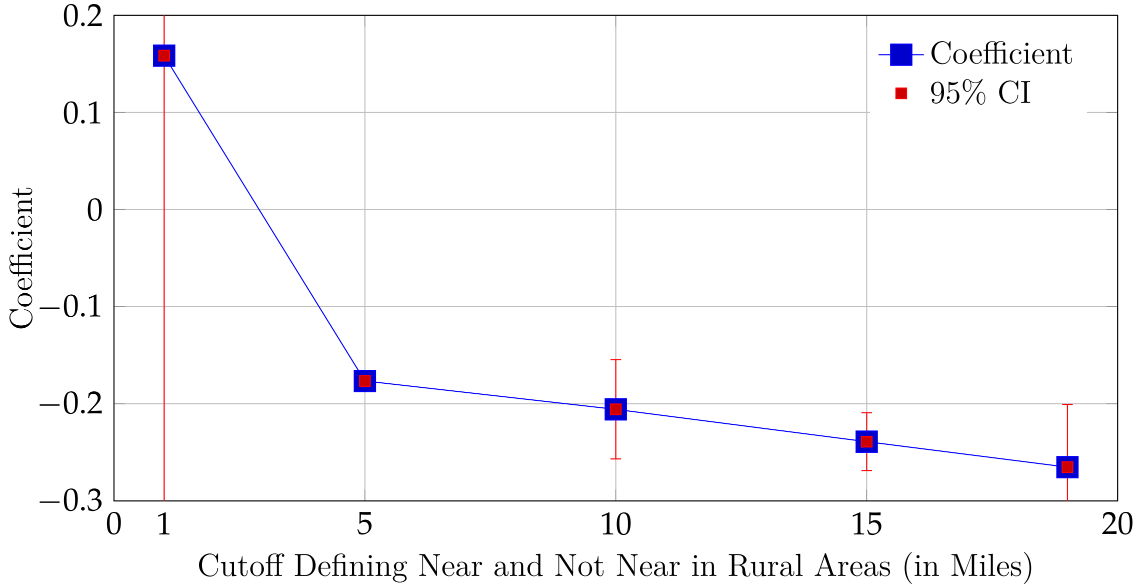
5.2.1 Sensitivity Tests

I also show how the coefficients change when I define different cutoffs for how I define near and not near for rural areas (leaving the definition of urban the same). What is near and not near in rural areas can be subjective so the following figures provide a good sensitivity check that the main results are sound.

To start, Figure 4 illustrates that as the sample qualifying as near increases for rural areas, the coefficient consistently hovers around negative 20 percent. Each point along the horizontal axis represents a cutoff distance used to distinguish between "Near" and "Not Near." For example, at 5 miles on the x-axis, "Near" is defined as within 0 to 5 miles, and "Not Near" as within 5 to 10 miles. Note that in Figure 4, distances beyond 20 miles are excluded from the model.[27] At the cutoff of 1 mile, the coefficient is around .15. However, this is not statistically significant and due to there not being enough sample in the "Near" group.

[27] Near and not near definitions for urban areas remain the same in this model. Furthermore, the definition of urban remain the same for all models.

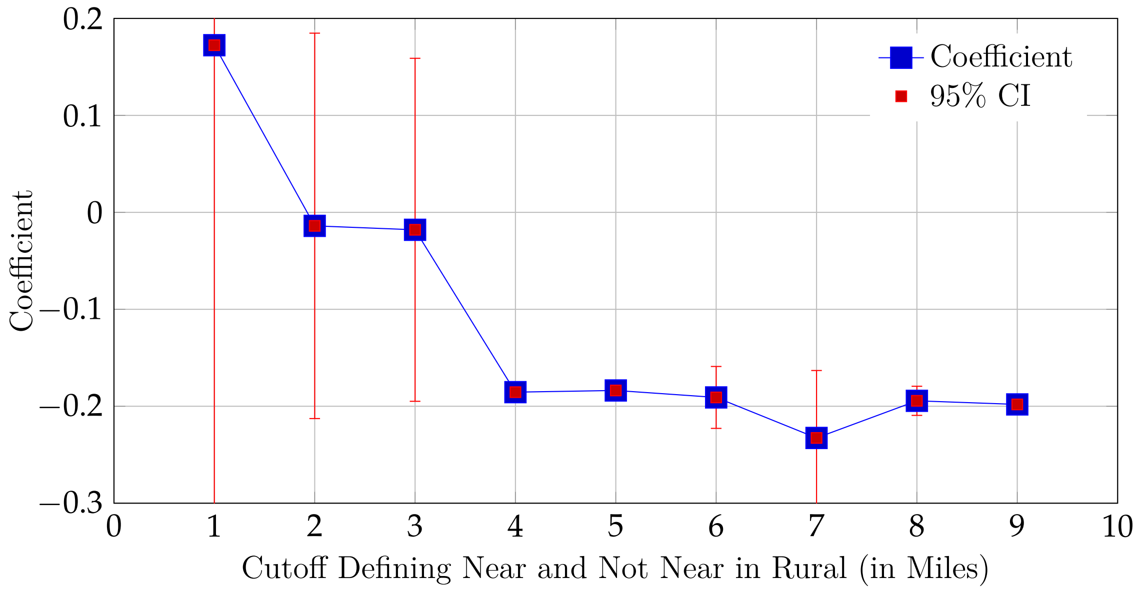
Figure 4: Equation (2) Results With Different Cutoff Definitions



Note: Households defined in rural areas are included out to 20 miles in sample. Definition of urban areas remains the same (0-1 miles is near and 1-3 miles is not near)

The next figure is the same figure, but with only 10 miles of distance in the sample compared to Figure 4 that has 20 miles of distances in the sample.

Figure 5: Equation (2) Results With Different Cutoff Definitions



Note: Households defined in rural areas are included out to 10 miles in sample. Definition of urban areas remains the same (0-1 miles is near and 1-3 miles is not near)

For both figures, after 4-5 miles, the results stabilize. If the cutoff for near is less than 4

Table 8: Facility Clustering Results

Variables	Distance (1)	Distance (2)
After	-0.147	-0.231***
	(0.0979)	(0.0765)
Near*After	0.103	0.181***
	(0.0861)	(0.0548)
After*Treat	0.0573	0.119*
	(0.0918)	(0.0698)
Near*After*Treat	-0.200**	-0.220***
	(0.0980)	(0.0737)
Observations	41,775	48,157
R-squared	0.786	0.792

miles, there is not enough sample for "Near" which leads to large standard errors. Based on these images, I choose cutoffs for rural regions after 5 miles.

In addition to clustering at the tract level, I also cluster at the facility level to show the difference in results. I cluster the results of Distance (1) and Distance (2) from Table 7 at the facility level. Table 8 shows that the standard errors vary slightly but do not affect the results.

5.2.2 Continuous Distance Results

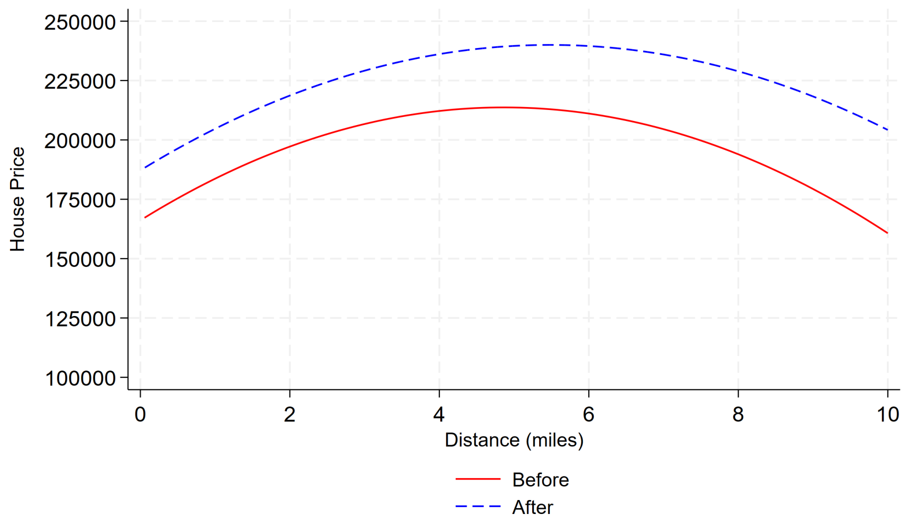
The previous section demonstrated that the difference between property values for homes near and not near newly opened facilities decline relative to the same distance around old facilities. This section now examines how house values change as the distance between facilities and single-family properties increases.

Figure 6 plots trends using the raw data, comparing how property values change before and after a facility opens between the treated and always treated areas. Before a facility opens, there is a slight increase in property values at varying distances. This is likely due to the fact that this slight increase is picking up macroeconomic effects such as increasing house prices at different years. However, after facilities open, there is a large increase in property values until around 3 miles away from the property with a plateau followed by levelling off.

Running the model also shows that there is a positive relationship between price and distance. Figure 7 plots $\beta_4 + \beta_6$*Distance (not logged) from Equation (4) to show how housing values change at different distances on average relative to the same distances for the always treated group. This shows the average increase in property values as houses get further from a facility. This figure is showing that at each mile away from a facility,

Figure 6: Plotting Data of House Prices Before and After Facility Opens

(a) House Prices Near Always Treated Facilities



(b) House Prices Near Treated Facilities

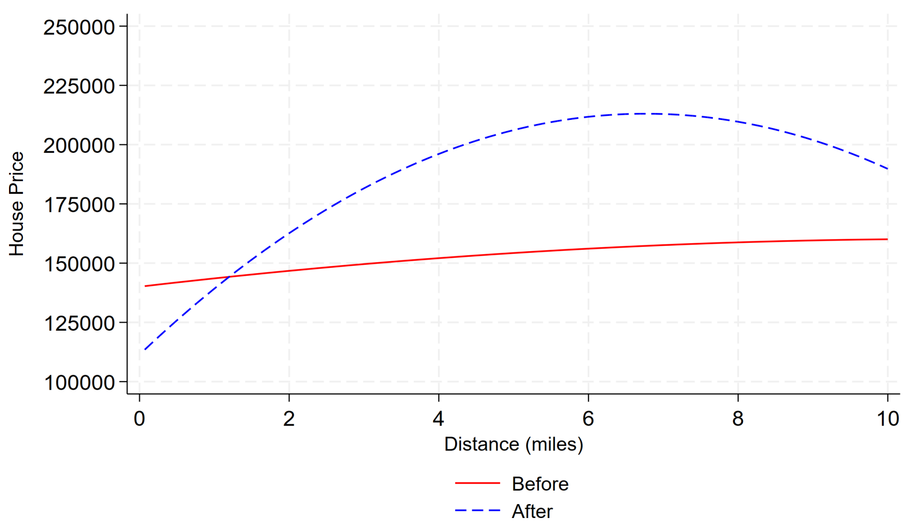
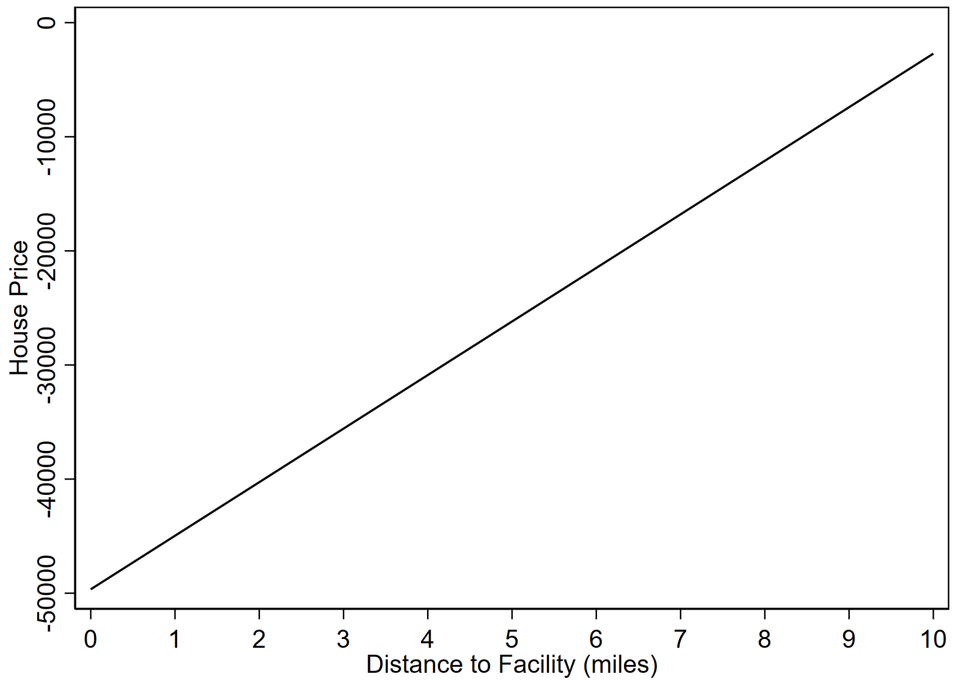


Figure 7: Plotted Housing Price Values



properties will increase by an average of $4,694.

Table 9 is the basis for Figure 7. The values for After*Treat also provide interesting insight on when the distance is 0 miles.[28] At 0 miles, property values fall $50,000 after treatment compared to the always treated group.[29]

Table 9 shows that as distance increases, the sales price of homes also increases relative to far away. These results do not include any transformations in the first column. In the second column, the sales price and distance are in natural log form[30]. One motivation of transforming the values is that the standard errors in the first column are huge.[31]

Table 9 show that the spatial dimension is important. Even though property values are the lowest right next to a facility, the negative effect is not permanent and dissipates as properties get further and further away. There could be an amenity effect at a certain distance away from a facility for several reasons. The first reason could be population growth and job creation which creates more demand for houses in the local area- just not near. The

[28] My smallest distance is .0008 which is very close to zero.

[29] The reference group here is 9 to 10 miles.

[30] To avoid negative natural logs, all distances (in miles) have a 1 added to the value since many of the distances in my dataset are less than 1 mile to a facility. Taking the natural log of a number below 0 would lead to a negative value

[31] Other papers such as Bellemare & Wichman (2020) and Chen & Roth (2024) discuss some of the issues that arise with transformations.

Table 9: Continuous Distance Results

	Not Logged (In $)	Logged
Distance	-305,761	-8.412
	(3,972,963)	(46.90)
After	28,909***	-0.230***
	(9,739)	(0.0334)
After*Distance	-2,058*	0.00495
	(1,214)	(0.0162)
Treat*Distance	141,845	-11.08
	(3,993,286)	(47.55)
After*Treat	-49,657***	0.0369
	(10,571)	(0.0406)
After*Treat*Distance	4,694***	0.0482**
	(1,717)	(0.0219)
Observations	84,684	84,684
R-squared	0.748	0.827

Standard errors in parentheses. In the first column, distance is not logged and presented in dollar amounts. In the second column, distance is in log form, where distance values are transformed by adding 1 to avoid negative values. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

second reason is that perhaps houses Near a facility moved away enough to no longer by near the facility, but close enough that they are still in the local area.

Figure 9: Results of Coefficients at Different Distances With Confidence Intervals

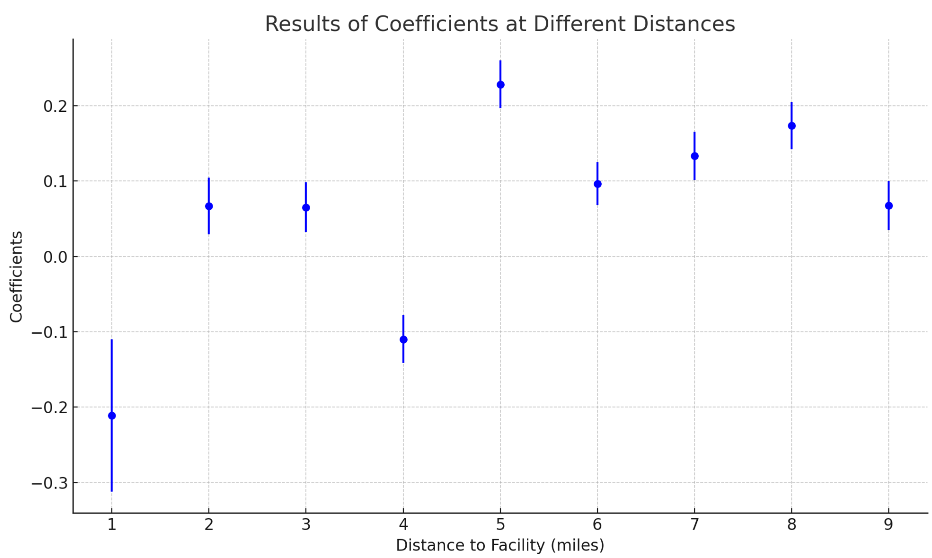
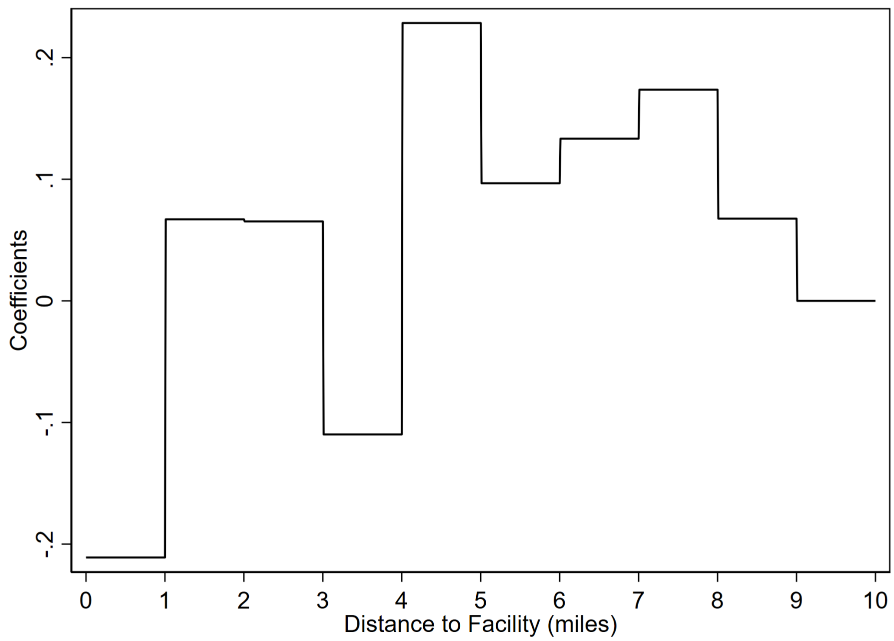


Figure 8 shows different bands on how property values change using results from equation

Figure 8: Results of Coefficients at Different Distances



The sample sizes for each distance range are as follows: 0-1 miles: 705; 1-2 miles: 1,692; 2-3 miles: 2,393; 3-4 miles: 2,344; 4-5 miles: 2,460; 5-6 miles: 3,244; 6-7 miles: 2,450; 7-8 miles: 2,723; 8-9 miles: 2,122; 9-10 miles: 2,133. All bands are statistically significant.

(4). The results also support the previous tables and figure. Between 0-1 miles, property values are 20 percent lower compared to 9-10 miles away, relative to the same comparison for always treated. This figure shows a steady increase in property values as properties are further away. This also supports the idea that some households may move to be within a few miles to the facility, just not right next to the facility or in view. Figure 9 shows the confidence intervals from Figure 8 and shows that each distance band is statistically significant.

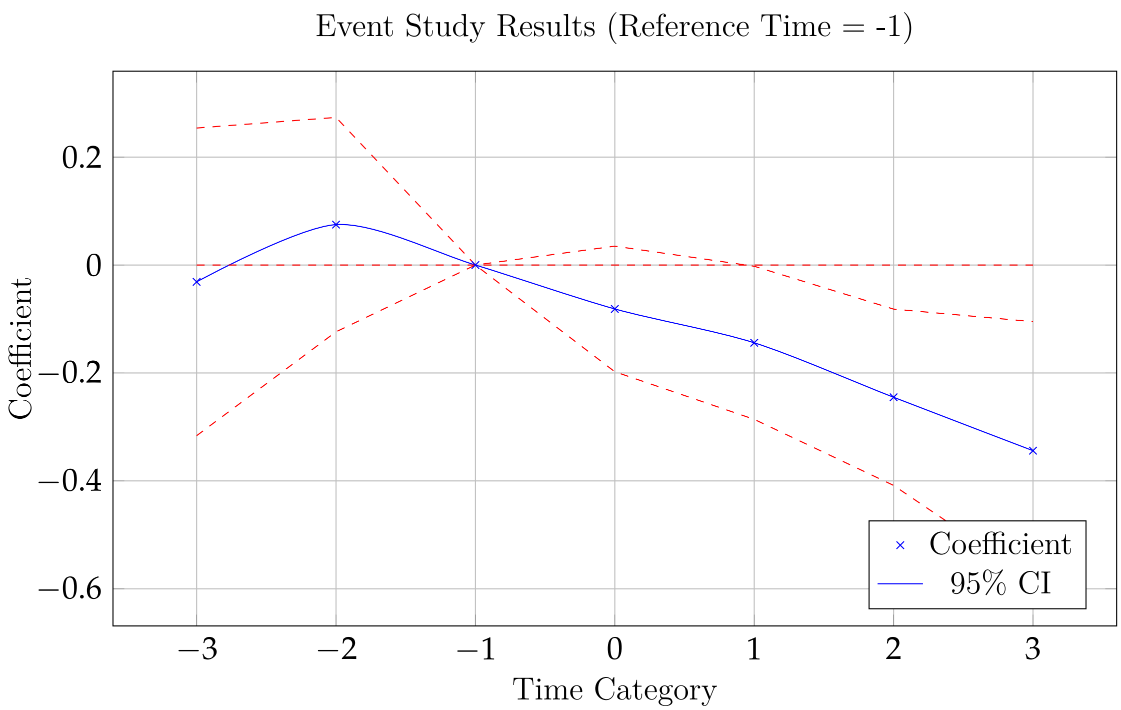
In the appendix, I add additional function forms for robustness that all show the same relationship that as the distance increases, the sales price of single-family homes increases relative to far away.

5.2.3 Event-Study Results

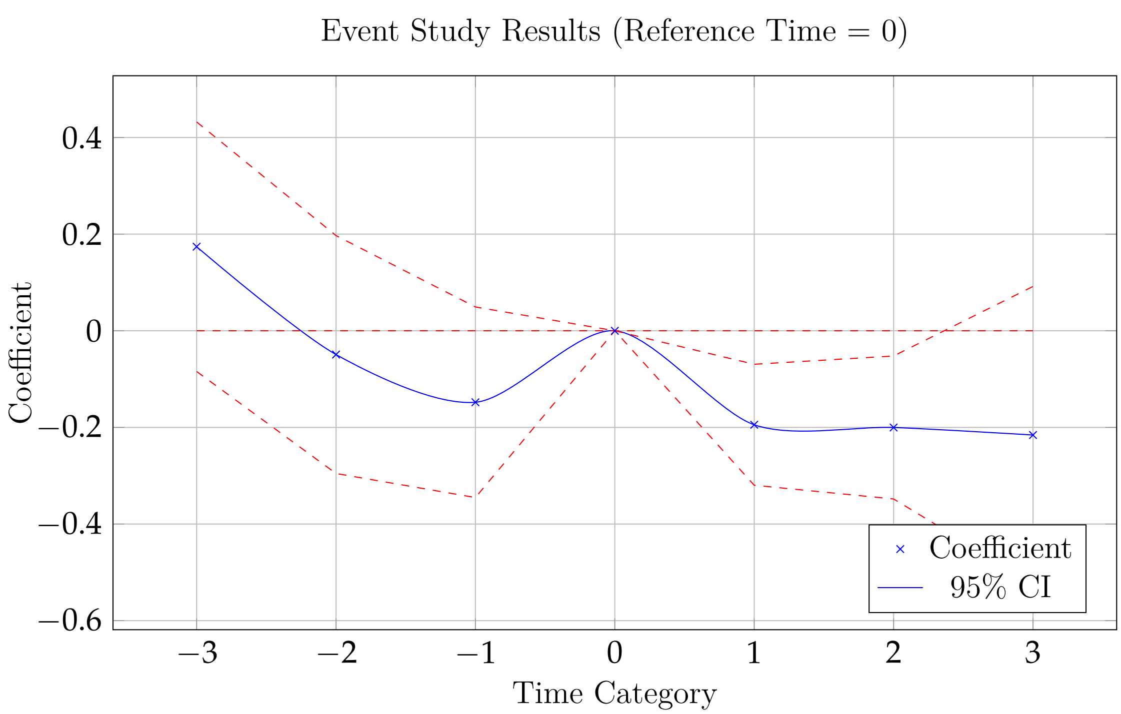
This section now highlights the changes in coefficients before and after a facility opens at different times while also accounting for the fact that homeowners can make a decision at t-1 to account for the announcement of the facility.

The above figure illustrates the coefficients from the triple-difference analysis over time, depicting housing prices three years before and after a facility opens. This staggered event

Figure 10: Comparison of Event Study Results with Different Reference Times



(a) Reference Time = -1



(b) Reference Time = 0

study, adapted from Goodman-Bacon (2021) for triple-difference rather than difference-in-difference, highlights this trend.

Figure 10 (a), with a sample size of 48,157 observations, reveals a significant drop in property values 1-2 years after a facility opens. The coefficient three years after the facility opens is significant at -0.34, while the coefficients at t-2 and t-3 are small and insignificant, indicating a lack of pre-trends before the ”policy shock.”

Similarly, Figure 10 (b), based on 41,775 observations, shows a significant impact on property values 1-2 years post-opening. The comparison between the event study results using different reference times sheds light on pre-trends before the facility opening. In both panels (with reference times at -1 and 0), the coefficients for the pre-treatment periods are small and statistically insignificant, especially in the years t-2 and t-3. This indicates a lack of significant pre-trends in property values before the facility opens.

The small and insignificant coefficients suggest that property prices remained stable in the years leading up to the facility's opening, supporting the assumption that no anticipatory effects or systematic differences occurred prior to the policy change. This strengthens the internal validity of the analysis by showing that the observed post-opening effects on housing prices are not driven by pre-existing trends.

6 Analysis of Heterogeneity

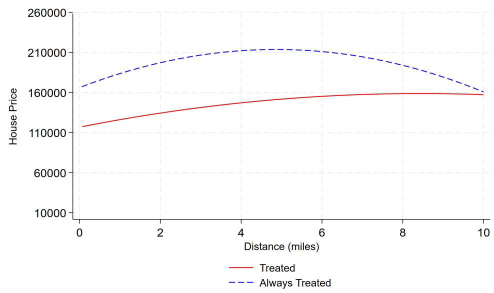
6.1 Rural and Urban Coefficients Separated

To further interpret the results, it's crucial to explore the differing perceptions of living near a facility between urban and rural residents. Urban dwellers may view proximity to a facility more negatively due to concerns about safety, increased traffic, and stigma, given the higher population density and close-knit communities. In contrast, rural residents might be less concerned about these issues but more sensitive to disruptions in the natural landscape, noise pollution, and the potential for decreased privacy. These factors collectively influence property values differently in urban and rural settings.[32]

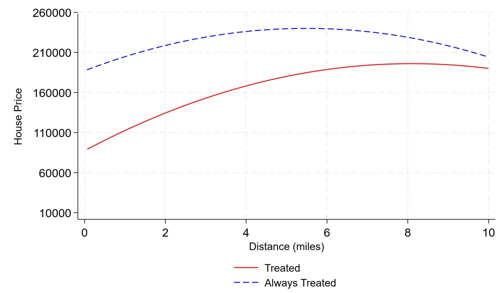
Before showing the results, Figure 11 paints a picture of the difference between single-family homes in rural areas vs. urban areas in the raw data. The first thing to notice is that housing prices in urban areas are higher than prices in rural areas. For both urban and rural areas, the change in price is relatively flat before any facility opens for both the treated and always treated figures. After a facility opens, housing prices are the lowest near the

[32] I tested numerous definitions of rural vs. urban with different cutoffs between 700 and 1200. All results were similar.

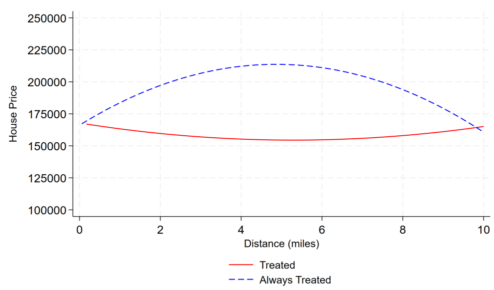
Figure 11: Plotting House Values for Urban vs. Rural Separated



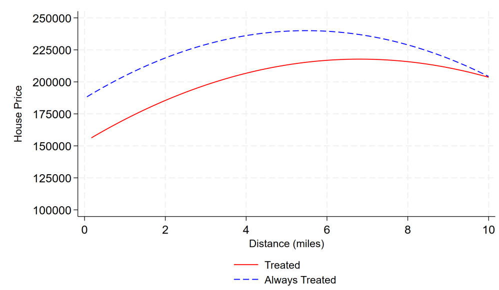
(a) Rural Before



(b) Rural After



(c) Urban Before



(d) Urban After

facility but then show an upward trend before leveling off. Like previous summary statistics provided by graph, this does not control for anything - it is merely showing the change in prices before and after. The following table shows the results from equation (2) broken up by urban vs. rural areas.[33]

The results above indicate that the main effects are driven by households in urban areas. Although both rural and urban locations experience declines in property values, the results for rural areas are not statistically significant. However, the magnitude of the decrease in rural property values is still substantial at 8 percent. As discussed, factors such as proximity to the facility play a role. Additionally, the fact that larger, more secure facilities are typically located in rural areas, while smaller, less secure ones are found in urban areas, may contribute to the psychological impact which will affect property values differently.

Figure 12 and Figure 13 show that housing prices plummet right next to a facility (between 0 and 1 miles) relative to 9-10 miles. As the distance increases, the results are either statistically insignificant or the values of homes are increasing. This shows that the relationship of near relative to not near is important in urban areas when a facility is right

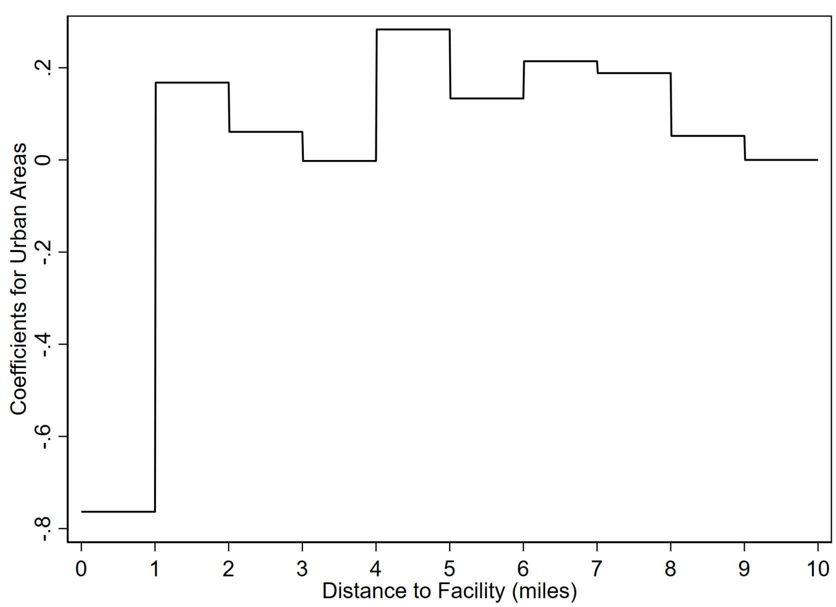
[33]Recall that the definition of urban is when the population density is greater than 1,000 people per square mile and the definition of rural is less than 1,000 people per square mile. Changing this definition resulted in very little if any change to the main results and to the robustness checks.

Table 10: Urban and Rural Coefficients Separated

	Urban	Rural
After	-0.375***	-0.154*
	(0.0937)	(0.0899)
Near*After	0.741*	0.0799
	(0.368)	(0.0614)
After*Treat	0.245***	-0.0196
	(0.0410)	(0.0910)
Near*After*Treat	-0.879**	-0.0801
	(0.388)	(0.0793)
Observations	10,059	38,203
R-squared	0.800	0.776

Robust standard errors clustered at the tract level
*** p<0.01, ** p<0.05, * p<0.1

Figure 12: Price Coefficients in Urban Areas at Different Distances



Note: In this regression, both the distance and sales price variables are logged. The y-axis represents the results for the interaction terms at various distances, with the reference group being properties located 9-10 miles from the facility. The x-axis shows the distance (in miles) between a property and the facility. The results for distances of 0-1 mile, 4-5 miles, 6-7 miles, and 7-8 miles are statistically significant. The sample sizes for each distance range are as follows: 0-1 miles (153), 1-2 miles (885), 2-3 miles (1,724), 3-4 miles (1,793), 4-5 miles (1,877), 5-6 miles (2,581), 6-7 miles (1,793), 7-8 miles (1,432), and 8-9 miles (1,348).

Figure 13: Price Coefficients in Urban Areas at Different Distances With Confidence Intervals

next to a home. This contrasts to the results seen in Figure 14 and Figure 15 that have no statistically significant results. On one hand, the presence of a correctional facility in rural areas might be perceived positively as it provides a stable source of employment. On the other hand, the stigma associated with such facilities could have negative effects. These opposing influences may contribute to ambiguous conclusions in rural areas, where the net impact remains unclear.

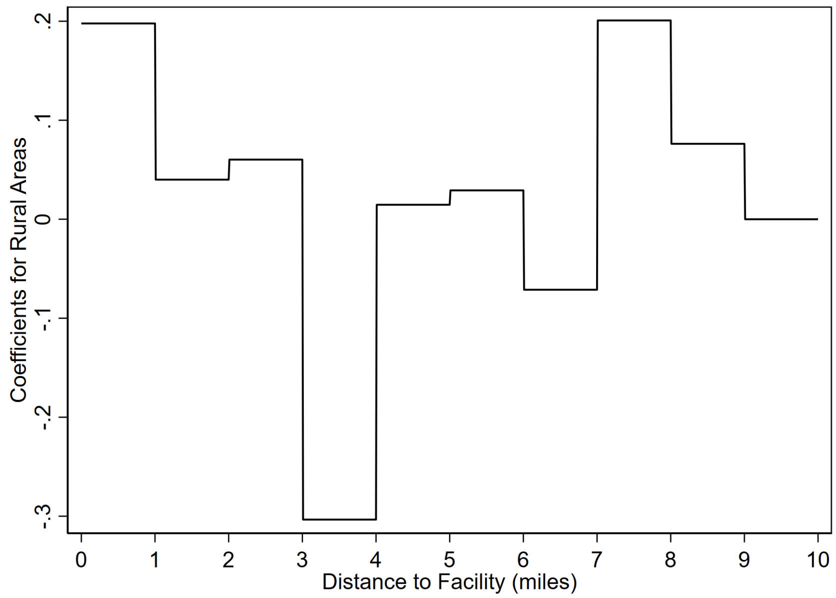
Although the results for rural areas are not statistically significant, there is still an interesting story as to why the coefficients are positive right next to the facility. It could be that this strengthens the idea that rural areas welcome facilities - especially larger facilities due to the fact that they bring employment and perhaps new infrastructure.

6.2 Closed Facility Results

While all the above results focus on the effects on property values when a facility opens, I also show results for when a facility closes.

One might expect property values to rise after a facility closes. However, I cannot reject a null hypothesis of no effect. Often, closed facilities are either repurposed under a different name or left abandoned, with the infrastructure remaining intact. This suggests a behavioral tendency where people still avoid living near such infrastructure, even if it is no longer in use.

Figure 14: Price Coefficients in Rural Areas at Different Distances



Note: In this regression, both distance and sales price variables are logged. The y-axis shows the results for the interaction terms at different distances with the reference group of 9-10 miles. The x-axis is distance (in miles) between a property and a facility. None of the coefficients are statistically significant. The sample sizes for each distance range are as follows: 0-1 miles (552), 1-2 miles (807), 2-3 miles (669), 3-4 miles (551), 4-5 miles (583), 5-6 miles (663), 6-7 miles (657), 7-8 miles (1,291), and 8-9 miles (774).

Figure 15: Price Coefficients in Rural Areas at Different Distances With Confidence Intervals

Table 11: Closed Facility Results

After	-0.0754
	(0.0633)
Near*After	0.116
	(0.0840)
After*Close	-0.0412
	(0.131)
Near*Close*After	0.0741
	(0.164)
Observations	18,089
R-squared	0.858

Standard errors are clustered at the tract level
*** p<0.01, ** p<0.05, * p<0.1

The results indicate no statistically significant change in property values following a facility closure. This could have implications for policymakers and urban planners who might be considering the economic benefits of closing or re-purposing correctional facilities. Simply closing a facility may not lead to the anticipated increase in property values unless steps are taken to remove or fully re-purpose the infrastructure.

This finding is consistent with the "stigma" effect seen in other contexts, such as environmental hazard cleanups. Similar findings have been reported by McCluskey & Rausser (2003), Messer et al. (2006), and Greenstone & Gallagher (2008), where even after hazards are remediated and no longer pose a threat, local residents are hesitant to live near such sites. In the case of closed facilities, the stigma associated with the physical infrastructure may persist, discouraging nearby residency.

Future research could explore the extent to which different types of facilities (e.g., prisons vs. rehabilitation centers) or regions (urban vs. rural) experience different property value effects when facilities close. Additionally, investigating how the removal or redevelopment of closed facility sites influences the stigma effect could provide more insights into how best to mitigate the negative impact on property values.

6.3 Results Based on Different Security Levels

Table 12 shows the results based on different definitions of security levels. The most significant results come from correctional facilities such as detention, and other larger facilities such as minimum security, medium security, maximum security, and juvenile facilities. These results show that perhaps households respond more to living near larger facilities where individuals cannot come and go. It should be noted that private facilities include halfway homes, charities, and private correctional facilities.

I will next categorize private correctional facilities as a different category to look at the results. Jail facilities could have insignificant results due to the fact that they are often not standalone facilities. They are often place within a courthouse or a police station.

Table 12: Property Values Based on Security Levels

	Detention Facility	Jail Facilities	Min, Med, Max, Juvenile Facilities	Private Facility	Work Camp
After	0.00570	-0.251*	-0.163**	-0.210**	-0.0164
	(0.0470)	(0.135)	(0.0797)	(0.0892)	(0.381)
Near*After	0.232***	0.284**	0.208***	-0.125	-0.0870
	(0.0533)	(0.142)	(0.0539)	(0.123)	(0.0762)
After*Treat	0.238**	0.141	0.172**	0.103	-0.296***
	(0.0995)	(0.162)	(0.0801)	(0.0850)	(0.0481)
Near*After*Treat	-0.344***	-0.275	-0.324**	0.175	0.346**
	(0.0713)	(0.168)	(0.141)	(0.128)	(0.120)
Observations	12,124	10,240	18,838	6,405	550
R-squared	0.785	0.779	0.843	0.692	0.741

Standard errors are clustered at the Tract level
*** p<0.01, ** p<0.05, * p<0.1

The table presents the effects of different security-level facilities on property values, revealing nuanced patterns that warrant further exploration. Properties near detention facilities; jails; and minimum, medium, and maximum security combined with juvenile facilities experience a significant decrease in property values, with declines of approximately 30% in proximity to these facilities.

In contrast, the results for work camps show an unexpected increase in property values of around 35%. Given the smaller sample size (550 observations), this result raises concerns about its validity and generalizability, and thus should be interpreted with caution.

The negative impact of detention and correctional facilities supports the hypothesis that these large, standalone structures have a disamenity effect on nearby homeowners. However, the positive effect observed for work camps is less intuitive. To address this inconsistency, I categorize work camps under minimum security facilities in later analyses. This is also motivated by the negative coefficient for After*Treat where as all other security-level types have a positive coefficient as this is when a facility is not near as opposed to near. The decision to categorize work camps as minimum security facilities stems from the nature of work camps, which typically house lower-risk inmates and have less visible security infrastructure. Their classification as minimum security facilities aligns them more closely with this category, where the disamenity effect is expected to be less pronounced than for higher security facilities.

For detention facilities and jails, the results are consistent with expectations and support my initial hypotheses. Jails are often located within larger complexes, such as courthouses or police stations, which may explain why they do not consistently lead to significant disamenity effects. Their integration into broader public infrastructure could offer some offsetting amenity effects. In contrast, correctional facilities, including detention and juvenile facilities, are large, standalone structures, making their disamenity impact on nearby property values more pronounced.

6.4 Results Based on Different Owners

Table 13 compares property values based on who owns the facility. It should be noted that there were not enough observations in my dataset to include Federal in the results. In this case, facilities that are local or owned by the state have the most number of observations. Not surprisingly then, the standard errors are lower and statistically significant for these categories. State-owned facilities include larger facilities. While local facilities tend to be more minimum-security facilities such as transitional facilities. Local facilities tend to hold people for a temporary amount of time for reasons such as awaiting trial. The difference between a State-owned and Federal-owned facility is whether an offense crosses state lines or international boundaries. State prisons typically hold more violent criminals or people who have committed more violent crimes, such as murder, rape, and assault with a deadly weapon (EisnerGorin LLP, 2024).

Furthermore, I show results on how property values are affected near correctional facilities only (excluding rehabilitation facilities), whether the owner of the facility is a private company that owns a correctional facility, and minimum, medium, and maximum facilities separated into distinct categories.

As also discussed in the appendix, a private prison is a facility that incarcerates offenders for profit. The three private companies that hold the majority of market share are GeoGroup, CoreCivic, and Management & Training Corporation.

Table 14 shows results for just correctional facilities. That is, if it is used to house

Table 13: Comparison of Local, Private, and State

	Local	Private	State
After	-0.0446	-0.152	-0.127
	(0.0653)	(0.132)	(0.138)
Near*After	0.198***	-0.223	0.223***
	(0.0539)	(0.147)	(0.0551)
After*Treat	0.0282	0.0354	0.181**
	(0.0870)	(0.130)	(0.0781)
Near*After*Treat	-0.220***	0.183	-0.345**
	(0.0765)	(0.187)	(0.154)
Observations	22,580	6,557	19,020
R-squared	0.783	0.737	0.815

*** p<0.01, ** p<0.05, * p<0.1.

individuals who have been convicted of a crime formally vs. a rehabilitation center or non-profit where individuals can come and go freely. More specifically, this does not include non-profits and halfway houses which I have categorized as rehabilitation facilities and facilities that are used not for punishment but rather to re-integrate people back into society. The results have similar results even when different definitions of near and not near are used.

Another interesting story to tell is whether or not property values near a facility change based on whether the owner of a correctional facility is private or not. Nearly all new U.S. prisons opened from 2000-2005 were private. Although this is before the scope of my paper, this still provides interesting insight on the dynamics of prison construction and how this could potentially affect property values differently. The results using the first definition of near and not near have property values dropping by 60 percent. However, this could be driven by the fact that most private correctional facilities tend to be larger and also medium security or higher. Private facilities such as CoreCivic, Management Training Corporation, and GeoGroup only own larger and higher security facilities so it makes sense that property values would decline much more.

Lastly, as stated earlier in the data cleaning section, minimum, medium, and maximum security facilities were contained to one category. In the following results, I show that property values do respond to the type of facilities. It is not surprising that property values are the most effected by living near maximum facilities and medium security facilities. Yet, interesting, the results are insignificant for minimum security facilities. This could be due to the fact that many minimum security facilities are transfer facilities or are facilities that have less security features compared to the larger facilities and are designed for the release of inmates.

Table 14: Impact of Correctional Facilities on Property Values

Variables	Distance (1)	Distance (2)
After	-0.105	-0.165***
	(0.0968)	(0.0582)
Near*After	0.148	0.192***
	(0.0970)	(0.0509)
After*Treat	0.0454	0.106
	(0.0995)	(0.0723)
Near*After*Treat	-0.228**	-0.236***
	(0.109)	(0.0748)
Observations	36,658	42,211
R-squared	0.806	0.810

Robust standard errors in parentheses, clustered at the tract-level.
Distance (1): Near (0-1 Miles Urban, 0-5 Miles Rural); Not Near (1-3 Miles Urban, 5-10 Miles Rural).
Distance (2): Near (0-1 Miles Urban, 0-10 Miles Rural); Not Near (1-3 Miles Urban, 10-20 Miles Rural).
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 15: Results for Privately Owned Correctional Facility

Variables	Distance (1)	Distance (2)
After	-0.248*	-0.274**
	(0.127)	(0.126)
Near*After	0.360	0.194
	(0.220)	(0.220)
After*Treat	0.379**	0.286*
	(0.165)	(0.160)
Near*After*Treat	-0.606**	-0.178
	(0.238)	(0.221)
Observations	3,746	4,607
R-squared	0.674	0.687

Robust standard errors in parentheses, clustered at the tract-level.
Distance (1): Near (0-1 Miles Urban, 0-5 Miles Rural); Not Near (1-3 Miles Urban, 5-10 Miles Rural).
Distance (2): Near (0-1 Miles Urban, 0-10 Miles Rural); Not Near (1-3 Miles Urban, 10-20 Miles Rural).
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 16: Maximum, Medium, and Minimum Separated

Variable	Maximum Security Distance (1)	Maximum Security Distance (2)	Medium Security Distance (1)	Medium Security Distance (2)	Minimum Security Distance (1)	Minimum Security Distance (2)
After	0.110	0.0485	-0.283*	0.280	-0.0982	-0.218***
	(0.171)	(0.143)	(0.159)	(0.218)	(0.0713)	(0.0576)
Near*After	0.0408	0.138	0.193	-0.461	0.0918	0.221***
	(0.0847)	(0.150)	(0.177)	(0.302)	(0.0693)	(0.0405)
After*Treat	-0.687**	-1.026***	0.322	-0.219	-0.0747	-0.336
	(0.278)	(0.335)	(0.228)	(0.198)	(0.0917)	(0.269)
Near*After*Treat	-0.926***	0.234	-0.599*	0.179	-0.0634	0.220
	(0.112)	(0.217)	(0.307)	(0.317)	(0.0908)	(0.278)
Observations	1,117	1,407	4,204	4,640	17,200	19,205
R-squared	0.747	0.752	0.764	0.768	0.874	0.875

Robust standard errors in parentheses and are clustered at the Tract level

*** p<0.01, ** p<0.05, * p<0.1

Distance (1): Near (0-1 Miles Urban, 0-5 Miles Rural); Not Near (1-3 Miles Urban, 5-10 Miles Rural).

Distance (2): Near (0-1 Miles Urban, 0-10 Miles Rural); Not Near (1-3 Miles Urban, 10-20 Miles Rural).

I show summary statistics that show how property values change at different distances to the facility before and after the opening in order to first set my hypothesis.

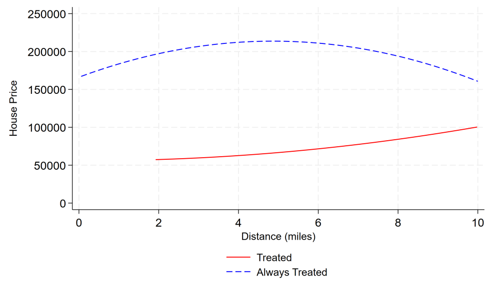
What is the most striking is that there is not enough data near maximum security facilities. Thus, any results with just maximum security facilities will not be accurate.

The results provided support that lumping together all facilities in the main results is the best method. Even so, the summary statistics depicted in the graphs above show interesting patterns based on the type of facility. The graphs above show that jails are not that affected when comparing before vs. after. Minimum, medium, and maximum facilities show an upward trend in property values as a property moves further away from a facility after it opens yet the results are not statistically significant (with the exception of the medium facility results). Though not in the results, facilities that are categorized as rehabilitation facilities (12i and 12j) show that property values are certainly affected after a rehabilitation facility opens in the treated group.

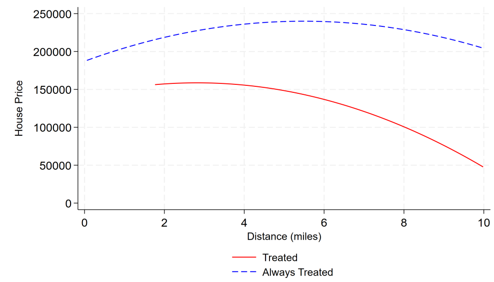
6.5 Defining After in Model Differently

Lastly, I show that defining After differently does not change the results significantly. In the original definition, I define After as 0 if -3 to 0 years (inclusive) and 1 if 1-3 years (inclusive). I provide additional definitions of After to add robustness to these results.

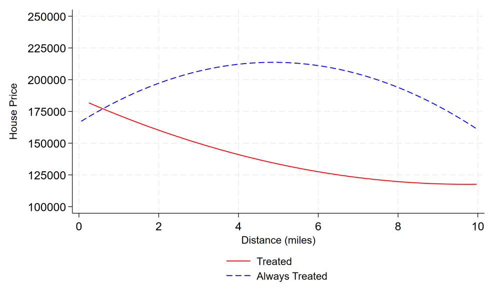
Figure 16: Property Values Based on Type of Facility



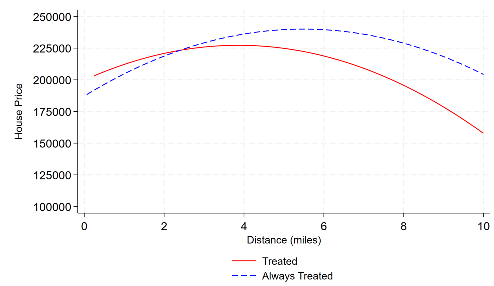
(a) Maximum Before



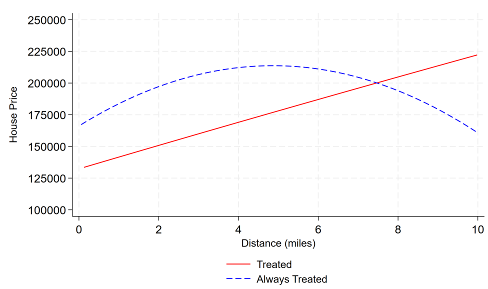
(b) Maximum After



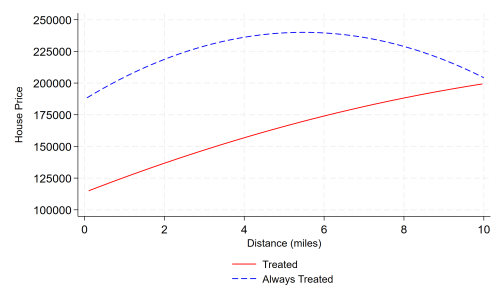
(c) Medium Before



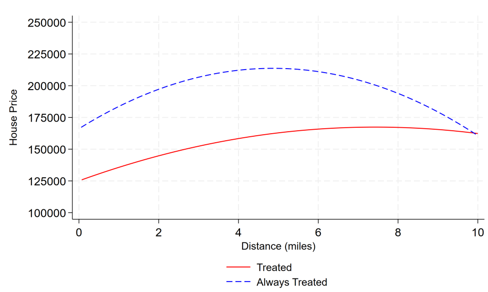
(d) Medium After



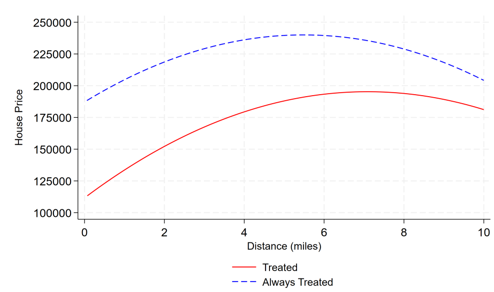
(e) Minimum Before



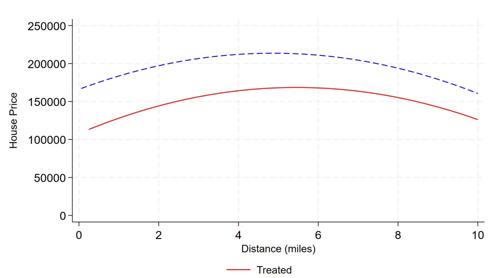
(f) Minimum After



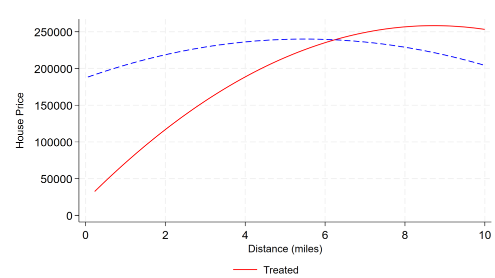
(g) Jail Before



(h) Jail After



(i) Rehab Before



(j) Rehab After

Table 17: Impact of Different Definitions of After on Housing Prices

	After (1 to 3 years)	After (0 to 3 years)
After	-0.259**	0.00591
	(0.112)	(0.0941)
Near*After	0.151**	0.144**
	(0.0635)	(0.0607)
After*Treat	0.0935	0.0599
	(0.0791)	(0.0839)
Near*After*Treat	-0.225**	-0.188**
	(0.0890)	(0.0882)
Observations	37,551	48,157
R-squared	0.790	0.792

Standard errors are clustered at the Tract level. In both columns, the definition of Before remains the same.
*** p<0.01, ** p<0.05, * p<0.1

Changing the definition of After does not have that large of an effect on the results. The definition that is used for the main results are between both of these values.

7 Conclusion

Since the 1980s, there has been a surge in the construction of new correctional facilities, yet little research has focused on their unintended consequences, particularly regarding nearby property values and capitalization effects. To address this gap, I create a novel dataset that combines data on correctional and rehabilitation facilities with micro-level real estate data. I use the hedonic method in the form of a triple-difference with a spatial component to investigate the impact of these facilities on property values.

I find that a newly opened facility significantly decreases nearby property values by an average of 20 percent. Moreover, the magnitude of this effect varies depending on the type of facility and how I define what close proximity means. These findings highlight the complex interplay between criminal justice policies and the unintended repercussions of prison expansion.

The study also highlights the wider impact of a rise in incarceration rates and more correctional facilities being built. When there's a greater need for these facilities, it not only affects those within the criminal justice system, but also has third party effects on nearby households. These findings underscore the intricate relationship between criminal justice policies and the unintended consequences of prison expansion.

One potential policy implication drawn from these findings is the need to reduce the incarceration of non-violent offenders. According to the Federal Bureau of Prisons (2024), approximately 62 percent of those incarcerated in 2024 were imprisoned for non-violent crimes, including offenses such as Banking and Insurance, Counterfeit, Embezzlement, Burglary, Larceny, Property Offenses, Continuing Criminal Enterprise, Courts or Corrections, Drug Offenses, Extortion, Fraud, Bribery, Immigration, Miscellaneous, and Robbery. Addressing the over representation of non-violent offenders in the correctional system could mitigate the adverse effects on property values and alleviate broader social and economic consequences.

For future research, property values could also be affected by a change in zoning laws around correctional facilities. That is, a facility that has recently been built could also change the surrounding zoning to industrial rather than residential which could be a reason for lower property values. This is explored more in Emblom-Hooe (2024), where I proxy industrial zoning laws with distance to power plants and to landfills to see how many facilities are near industrial zoned areas.

8 References

1. How Many Americans Are Unnecessarily Incarcerated? | Brennan Center for Justice. (n.d.). Retrieved February 21, 2024, from https://www.brennancenter.org/our-work/research-reports/how-many-americans-are-unnecessarily-incarcerated

2. Allen Klaiber, H., & Kerry Smith, V. (2013). Quasi Experiments, Hedonic Models, and Estimating Trade-offs for Local Amenities. Land Economics, 89(3), 413–431. https://doi.org/10.3368/le.89.3.413

3. Bakhtiar, M. M., Guiteras, R. P., Levinsohn, J., & Mobarak, A. M. (2023). Social and financial incentives for overcoming a collective action problem. Journal of Development Economics, 162, 103072.

4. Banzhaf, H. S. (2015). Panel Data Hedonics: Rosen's First Stage as a "Sufficient Statistic" (Working Paper 21485). National Bureau of Economic Research. https://doi.org/10.3386/w21485

5. Banzhaf, H. S. (2021). Difference-in-Differences Hedonics. Journal of Political Economy, 129(8), 2385–2414. https://doi.org/10.1086/714442

6. Banzhaf, H. S., & Farooque, O. (2013). Interjurisdictional housing prices and spatial amenities: Which measures of housing prices reflect local public goods? Regional Science and Urban Economics, 43(4), 635–648. https://doi.org/10.1016/j.regsciurbeco.2013.03.006

7. Bellemare, Marc F. and Casey J. Wichman, "Elasticities and the inverse hyperbolic sine transformation," Oxford Bulletin of Economics and Statistics, 2020.

8. Best Practices for Using Hedonic Property Value Models to Measure Willingness to Pay for Environmental Quality. (n.d.). https://doi.org/10.1093/reep/reaa001

9. Bishop, K. C., & Murphy, A. D. (2019). Valuing Time-Varying Attributes Using the Hedonic Model: When Is a Dynamic Approach Necessary? The Review of Economics and Statistics, 101(1), 134–145. https://doi.org/10.1162/rest_a_00722

10. Bishop, K. C., Kuminoff, N. V., Banzhaf, H. S., Boyle, K. J., von Gravenitz, K., Pope, J. C., ... & 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.

11. Boxall, P. C., Chan, W. H., & McMillan, M. L. (2005). The impact of oil and natural gas facilities on rural residential property values: A spatial hedonic analysis. Resource and Energy Economics, 27(3), 248–269. https://doi.org/10.1016/j.reseneeco.2004.11.003

12. Bureau of Justice Statistics, Criminal Victimization in the U.S.: 1973-79 Trends 43 tbl.17 (1981), http://www.bjs.gov/content/pub/pdf/cvus7379.pdf (data for 1979).

13. Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. C. (2024). Event-Studies with a Continuous Treatment (Working Paper 32118). National Bureau of Economic Research. https://doi.org/10.3386/w32118

14. Chay, K. Y., & Greenstone, M. (2005). Does Air Quality Matter? Evidence from the Housing Market. Journal of Political Economy, 113(2), 376–424. https://doi.org/10.1086/427462

15. Cherry, T., & Kunce, M. (2001). Do Policymakers Locate Prisons for Economic Development? Growth and Change, 32(4), 533–547. https://doi.org/10.1111/0017-4815.00174

16. Chen, J., & Roth, J. (2024). Logs with zeros? Some problems and solutions. The Quarterly Journal of Economics, 139(2), 891-936.

17. Chirakijja, J. (2022). The Local Economic Impacts of Prisons. The Review of Economics and Statistics, 1–45. https://doi.org/10.1162/rest_a_01267

18. Chirakijja, J. (2023). The economic effects of prison closures on local communities. Economics Letters, 228, 111155.

19. Criminal Victimization, 2010. (2010).

20. Currie, J., Davis, L., Greenstone, M., & Walker, R. (2015). Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings. American Economic Review, 105(2), 678–709. https://doi.org/10.1257/aer.20121656

21. Davis, L. W. (2011). The Effect of Power Plants on Local Housing Values and Rents. The Review of Economics and Statistics, 93(4), 1391–1402. https://doi.org/10.1162/REST_a_00119

22. Dippel, C., & Poyker, M. (2019). Do Private Prisons Affect Criminal Sentencing? (Working Paper 25715). National Bureau of Economic Research. https://doi.org/10.3386/w25715

23. Engel, M. R. (2007). When a prison comes to town: Siting, location, and perceived impacts of correctional facilities in the Midwest. ETD Collection for University of Nebraska - Lincoln, 1–314.

24. Gamper-Rabindran, S., & Timmins, C. (2013). DOES CLEANUP OF HAZARDOUS WASTE SITES RAISE HOUSING VALUES? EVIDENCE OF SPATIALLY LOCALIZED BENEFITS.

25. Genter, S., Hooks, G., & Mosher, C. (2013). Prisons, jobs and privatization: The impact of prisons on employment growth in rural US counties, 1997–2004. Social Science Research, 42(3), 596–610. https://doi.org/10.1016/j.ssresearch.2012.12.008

26. Greenstone, M., & Gallagher, J. (2008). Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program*. The Quarterly Journal of Economics, 123(3), 951–1003. https://doi.org/10.1162/qjec.2008.123.3.951

27. Greenstone, M., Hornbeck, R., & Moretti, E. (2010). Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. Journal of Political Economy, 118(3), 536–598. https://doi.org/10.1086/653714

28. Hawes. (1985). Cities With Prisons—Do They Have Higher or Lower Crime Rates? | Office of Justice Programs. https://www.ojp.gov/ncjrs/virtual-library/abstracts/cities-prisons-do-they-have-higher-or-lower-crime-rates

29. Hooks, G., Mosher, C., Genter, S., Rotolo, T., & Lobao, L. (2010). Revisiting the Impact of Prison Building on Job Growth: Education, Incarceration, and County-Level Employment, 1976–2004*. Social Science Quarterly, 91(1), 228–244. https://doi.org/10.1111/j.1540-6237.2010.00690.x

30. Kirchhoff, S. M. (n.d.). Economic Impacts of Prison Growth.

31. Kuminoff, N. V., Parmeter, C. F., & Pope, J. C. (2010). Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities?. Journal of environmental economics and management, 60(3), 145-160.

32. Kuminoff, N. V., & Pope, J. C. (2014). Do "capitalization effects" for public goods reveal the public's willingness to pay?. International Economic Review, 55(4), 1227-1250.

33. Krause, J. D. (1991). Community Opposition to Correctional Facility Siting: Beyond the "Nimby" Explanation. Humboldt Journal of Social Relations, 17(1/2), 239–262.

34. Los Angeles Criminal Defense Lawyer | Eisner Gorin LLP. (n.d.). Retrieved February 21, 2024, from `https://www.egattorneys.com/`

35. Messer, K. D., Schulze, W. D., Hackett, K. F., Cameron, T. A., & McClelland, G. H. (2006). Can stigma explain large property value losses? The psychology and economics of Superfund. Environmental and Resource Economics, 33, 299-324

36. Millay, J. R. (1989). From asylum to penitentiary: The social impact of Eastern Oregon Correctional Institution upon Pendleton. Humboldt Journal of Social Relations, 17(1/2), 171-195.

37. McCluskey, J. J., & Rausser, G. C. (2003). Stigmatized asset value: is it temporary or long-term?. Review of Economics and Statistics, 85(2), 276-285.

38. McGeever, K. (2019). Correctional Facility Establishments and Neighborhood Housing Characteristics. Sociological Perspectives, 62(3), 383–401. `https://doi.org/10.1177/0731121418800271`

39. McGee, R. A. (1981). Prisons and Politics. Lexington Books.

40. Muehlenbachs, L., Spiller, E., & Timmins, C. (2015). The Housing Market Impacts of Shale Gas Development. American Economic Review, 105(12), 3633–3659. `https://doi.org/10.1257/aer.20140079`

41. Office of Program Policy Analysis and Government Accountability. (2019). Florida Correctional Facilities.

42. Opsal, T., & Malin, S. A. (2020). Prisons as LULUs: Understanding the Parallels between Prison Proliferation and Environmental Injustices. Sociological Inquiry, 90(3), 579–602. https://doi.org/10.1111/soin.12290

43. Palmquist, R. B. (1992). Valuing localized externalities. Journal of Urban Economics, 31(1), 59–68. `https://doi.org/10.1016/0094-1190(92)90032-G`

44. Perry, S. E. (1987). Communities on the Way (First Edition). State Univ of New York Pr.

45. Pope, D. G., & Pope, J. C. (2015). When Walmart comes to town: Always low housing prices? Always? Journal of Urban Economics, 87, 1–13. https://doi.org/10.1016/j.jue.2014.10.004

46. Pope, J. C. (2008). Fear of crime and housing prices: Household reactions to sex offender registries. Journal of Urban Economics, 64(3), 601–614. https://doi.org/10.1016/j.jue.2008.07.001

47. Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy, 82(1), 34–55.

48. Steiner, B., & Wooldredge, J. (2008). Comparing State- Versus Facility-Level Effects on Crowding in U.S. Correctional Facilities. Crime & Delinquency, 54(2), 259–290. https://doi.org/10.1177/0011128706296963

49. Taylor, L. O. (2017). Hedonics. A primer on nonmarket valuation, 235-292.

50. Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics, 225(2), 175–199. https://doi.org/10.1016/j.jeconom.2020.09.006

51. The Economics of Private Prisons. (n.d.). The Hamilton Project. Retrieved March 27, 2024, from https://www.hamiltonproject.org/publication/post/the-economics-of-private-

52. Wainer, H. (2000). The centercept: An estimable and meaningful regression parameter. Psychological Science, 11(5), 434-436.

53. Yinger, J. (2018, September 4). Hedonic Teaching Material | John M. Yinger. https://joyinger.expressions.syr.edu/hedonic-teaching-material/

Table A1: Breakdown of Facility Type by Treated and Always Treated (AT)

	# of Facilities - Treated	# of Facilities - AT
Correctional Facilities	141	214
Maximum, Medium, and Minimum Security	80	183
Prison Camps	7	12
Detention Centers	23	10
Jails	31	9
Rehabilitation Facilities	18	19

A Appendix A

A.1 Data Collection, Cleaning, & Summary Statistics (continued)

A.1.1 Facility Data Cleaning & Summary Statistics

I excluded any facility in Alaska and Hawaii. There were not many facilities in these states and the information provided on each correctional site was poor. Next, I dropped facilities with missing locations.

Another important step was to double check all the addresses and names of the remaining facilities as some changed their name or moved locations since the surveys were taken. I also checked for duplicate facilities in each year specified above to make sure that a facility is not double counted in my final dataset.

After accounting for detention facilities and jails, my dataset consists of over 2000 facilities across the US that opened after 1980. I chose 1980 as a cutoff date because according to Hooks (2010), after 1980, the growth of these facilities skyrocketed. One reason this was the case was due to the tough-on-crime and war-on-drugs policies initiated during the Reagan administration. Thus, I felt that capturing the open dates after 1980 would be the most relevant to show how demand for correctional facilities shifted as a result.

The following table breaks down the types of facilities in my dataset as well as the owner types.

Table A2: Breakdown of Ownership of Facilities

Facility Owner	# Facilities (Treatment Group)
Federal	3
State	81
Local	50
Private	25
Total	159

Table A3: Summary of Opened and Closed Facilities by State

State	Total Opened	Total Closed
DC	0	4
FL	27	50
GA	28	40
KY	15	17
NC	11	28
SC	0	4
TN	10	11
VA	4	7

Table A1 shows that the majority of facilities in my dataset are correctional facilities. This table and the main results lump together maximum, medium, and minimum security facilities. For robustness checks, I separate these facility types to show the response in property values. While I cannot directly assess how individuals perceive living near different types of facilities, the variations in property value changes offer some insight and pave the way for further research in this area.

Table A2 provides information on what types of owners are in my dataset. The majority of facilities are owned by the state and local government. Many local government-owned facilities in my dataset are jails. Facilities that are privately owned include non-profits and privately owned correctional facilities. Examples of companies that privately own prisons are Management and Training Corporation, GeoGroup, and CoreCivic.

A private prison is a facility that incarcerates offenders for profit. Private prisons are growing in demand as state and federal facilities are forced to operate at or above capacity. The shift from a publicly operated correctional system to one that contains a corporate component has led to concerns about the different interests (The Hamilton Project, 2016). That is, whether cutting costs and maximizing revenue will supersede the interest of justice. Furthermore, there are three private companies that hold the majority of market share: GeoGroup, CoreCivic, and Management & Training Corporation.

The table shows that Florida and Georgia have the highest number of both opened and closed facilities, with 27 and 28 cases opened, and 50 and 40 facilities closed, respectively. On the other hand, Washington, D.C. and South Carolina report the lowest activity, with zero facilities opened but 4 facilities closed in both regions. Kentucky and North Carolina also display a notable number of opened facilities at 15 and 11, with 17 and 28 facilities closed, respectively. Overall, this table provides a clear comparative view of facilities activity by state, highlighting regional differences in facility closure rates relative to openings.

A.1.2 Summary Statistics of Treated and Always Treated Block Groups Before and After a Facility Opens

Table A4 compares the characteristics of block groups before (Group 0) and after (Group 1) the opening of a facility. The "Difference in Means" represents the change in the mean of each variable before and after a facility opens, calculated as Group 0 - Group 1. A positive value suggests that the variable was higher before the facility opened, while a negative value indicates that it increased after the facility opened. The table provides insight into which socio-economic and housing characteristics shift, motivating further research into who is specifically affected by the presence of new correctional or rehabilitation facilities.

2009 was selected as the baseline for the "Before" period, as there is insufficient data available prior to this year to make robust comparisons for the same block groups.

Overall, this analysis provides preliminary evidence of some significant shifts in educational attainment, income levels, and housing characteristics (such as median year built) after the facility opens. However, other variables, such as racial composition and employment, do not show statistically significant changes. This motivates further research into who is specifically affected by the introduction of correctional or rehabilitation facilities and how different socio-economic factors are influenced over time.[34]

A.1.3 Real Estate Data & Summary Statistics

I narrowed the scope of the project to focus on the Southeast by choosing larger counties across the country to test whether or not the data was available. Sending emails to larger and more affluent counties created a good baseline on what states were more likely to have the data and what states do not have the data. For example, after sending emails to major cities in California, many of the counties required substantial payment for the data whereas major cities in Georgia and Florida provided the data for free.

One problem that arose when combing through the data is that even though all counties

[34] There are 58 block groups in this sample with makes up less than 40% of the total sample.

Table A4: T-test Results for Characteristics Before and After

Variable	T-value	P-value	Difference in Means
Population Density	-0.572	0.569	-827.692
White	-0.521	0.603	-2.520
Black	1.100	0.274	5.099
Asian	-0.173	0.863	-0.047
Hispanic	-1.594	0.114	-3.091
Less Than HS	4.387	0.000	10.322
High School	-0.674	0.502	-1.580
College	-1.420	0.158	-2.381
Masters	-1.465	0.146	-1.563
Doctorate	-0.822	0.413	-0.352
Not Enrolled	1.894	0.061	7.071
Employed in Business	-1.363	0.176	-2.873
Employed in Sales	-0.675	0.501	-0.868
Employed in Fishing	-0.173	0.863	-0.105
Employed in Construction	1.057	0.293	1.992
Employed in Transport	-1.283	0.202	-2.561
Median Income 2022	-4.643	0.000	-23758.043
Public Assistance	0.294	0.769	0.292
Housing Occupied	0.123	0.902	0.269
Housing Vacant	-0.123	0.902	-0.269
For Rent	1.101	0.273	5.307
For Sale	0.575	0.567	2.113
Median Year Built	-2.225	0.028	-6.145
Gas	1.501	0.136	6.500
Electricity	-2.264	0.026	-9.946
Coal	0.677	0.500	0.532
No Fuel	-0.654	0.514	-0.217
Median House Value 2022	-1.575	0.118	-48,987

Table A5: Demographic Characteristics at the County-Level

Characteristics	DC	VA	NC	SC	GA	FL	KY	TN
Population Density	11,196	85.5	403.3	155.6	196.7	372.9	54.5	138.3
White Alone (%)	41.0	84.2	64.2	51.4	54.6	74.2	90.9	86.7
Black or African American Alone (%)	46.9	12.1	23.5	42.3	37.0	17.5	4.9	7.6
Asian Alone (%)	3.9	0.7	4.2	1.9	3.7	2.6	0.4	1.7
Hispanic or Latino (%)	10.9	2.5	10.4	5.2	7.8	29.3	3.0	4.0
Less than High School (%)	9.4	14.3	11.0	12.4	13.5	12.9	18.2	13.9
High School Graduate (%)	90.6	85.7	89.1	87.6	86.5	87.1	81.8	86.1
Some College (%)	73.4	52.2	67.7	59.0	58.6	58.8	39.6	55.6
Bachelor's Degree (%)	57.6	23.3	39.3	29.0	31.4	29.8	14.3	26.9
Master's Degree (%)	33.2	8.7	14.4	11.3	12.5	11.1	5.2	11.3
Professional School Degree (%)	12.5	2.2	4.0	3.1	3.8	3.6	1.4	3.5
Doctorate Degree (%)	4.3	0.9	1.8	1.3	1.5	1.2	0.6	1.4
In Labor Force (%)	70.1	57.4	66.1	61.9	62.2	59.5	52.9	58.5
In Armed Forces (%)	0.6	0.2	0.3	2.2	0.5	0.3	0.1	0.1
Employed (%)	64.4	54.5	62.2	55.0	57.4	55.4	49.2	54.9
Unemployed (%)	5.1	2.7	3.7	4.7	4.3	3.8	3.6	3.5
Median Household Income ($)	82,604	52,226	59,252	49,058	52,523	53,027	43,316	46,767
Median Year Structure Built	1953	1960	1990	1975	1974	1976	1977	1980
Living in Poverty (%)	15.1	12.3	12.9	16.4	16.2	14.3	17.5	19.5

provided sales data with the home addresses, several counties did not provide the city or zip code, which is crucial for the next step of geocoding the addresses correctly.

To overcome this issue, I separated all my data that had city and zip code information and data that did not have this information – by state. The files that had city and zip code information could be geocoded directly and tended to have 99 percent or more of the data matching. I used the 2022 Streets data provided by ESRI's Business Analyst to geocode this data directly with ArcGIS Pro version 2.9.5.

The files that did not have all the information were more difficult to handle. For North Carolina, I used the AddressNC geocoder provided from NC OneMap to match the addresses with a location. Around 50 percent of this data was matched and the other 50 percent were tied or unmatched. Any addresses in North Carolina that were not matched or were tied at the end of this process were filtered from this dataset and run through the ArcGIS World Geocoder.

Table A5 provides the same summary statistics for the same variables for counties with real estate data, broken down by state. Overall, there is a large range in state characteristics. The Table shows that most population densities are below 500 people per square mile. The US Census Bureau classifies an area as rural if there are fewer than 1,000 people living per square mile. Thus, for all states except for Washington DC with facilities, the counties are rural.

Table A6: Percentage of Available Observations for Each Characteristic by State

Variable	DC	FL	GA	KY	NC	SC	VA	TN
Acres	0	82	99	86	81	79	57	100
Year Built	0	53	96	57	60	62	20	100
Bedrooms	0	42	4	1	0	53	13	0
Bathrooms	0	42	4	1	21	53	16	0
Square Feet	0	26	27	77	58	27	11	100
Total Rooms	0	0	73	0	21	0	0	100
Stories	0	42	73	51	0	0	0	100
Total Observations	100,968	703,681	241,663	17,641	436,752	8,066	40,125	14,995

Table A7: Summary Statistics for Different Characteristics

Variable	Mean	SD	Min	Max
Year Built	1988.27	26.69	1920	2022
Acres	0.83	1.40	0	20
Square Feet	2,126.45	2,150.49	1	10,000
Bathrooms	2.26	0.92	1	10
Bedrooms	3.43	0.93	1	10
Total Rooms	1.57	2.20	1	10
Stories	0.84	0.72	0	5
Sale Price	397,852.10	2,061,525.00	10,000	10,000,000

Another component of my research project is that I assume that preferences on how close is too close to a facility will also depend on whether the area is urban or rural. The Census Bureau defines rural as having a population density under 1,000 people per square mile and urban as having a population density over 1,000 people per square mile. I use the Spatial Join feature on ArcGIS Pro to join census tract information to the location of each facility. I then merge this data to my real estate dataset. I also join population density for each census tract where each facility is located.

Table A6 shows that there is a considerable range of data availability of what states have certain housing characteristics and what states do not. Washington DC real estate data had no housing characteristics. Whereas, Florida, Georgia, and North Carolina had a decent percentage of Acres and Year Built characteristics but not of Bedrooms and Bathrooms. Tennessee had the best housing characteristic data except for Bedrooms and Bathrooms.

It should be noted that Florida, Georgia, and North Carolina have the greatest number of real estate transactions in my dataset. This is due to the fact that the majority of facilities that opened or closed between 2005 and 2023 were the largest in these states.

Table A8: Facility and Real Estate Sample Sizes By Distance And State

Rural (Population Density < 1000)		
State	0-5 Miles	5-10 Miles
Washington DC	0	0
Virginia	11	11
North Carolina	26	28
South Carolina	12	12
Georgia	105	108
Florida	66	67
Kentucky	26	25
Tennessee	15	15

Rural (Population Density < 1000)		
State	0-5 Miles	5-10 Miles
Washington DC	0	0
Virginia	14423	21625
North Carolina	45381	78869
South Carolina	1631	2242
Georgia	39151	58502
Florida	123765	204379
Kentucky	8124	5531
Tennessee	6079	5001

Urban (Population Density > 1000)		
State	0-1 Miles	1-3 Miles
Washington DC	7	7
Virginia	1	1
North Carolina	16	18
South Carolina	8	7
Georgia	19	26
Florida	22	27
Kentucky	4	4
Tennessee	0	0

Urban (Population Density > 1000)		
State	0-1 Miles	1-3 Miles
Washington DC	45689	112629
Virginia	88	1022
North Carolina	11424	53277
South Carolina	277	1237
Georgia	730	28784
Florida	12722	84026
Kentucky	282	1095
Tennessee	0	0

Table A7 shows summary statistics including the mean, standard deviation, minimum value, and maximum value of my dataset. Acreage here is quite large which could be driving some of the null results for rural areas. Even so, I keep acres between 0 and 20 to keep the sample size of rural areas significant enough to run the models.

Table A8 shows the number of facilities is pared down. Nevertheless, there are still many in my distances of interest. Table A8 still shows that the greatest number of facilities are located in Georgia, Florida, and North Carolina. Table A8 also shows how much of the sample size varies by state. When breaking down by state, the states with the most rural observations include North Carolina, Georgia, and Tennessee and the states with the most urban observations are Washington DC and Florida.

Table A9 drives home the point that this repeated sales sample is very similar to the full sample in terms of what housing characteristics are available and the distribution of the sample size of facilities and transactions. That is, despite the smaller sample size of facilities and real estate data, the distribution of the sample remains the same.

Table A9: Facility and Real Estate Sample Sizes By Distance and State

Rural (Population Density < 1000)		
State	0-5 Miles	5-10 Miles
Washington DC	0	0
Virginia	8	11
North Carolina	23	29
South Carolina	11	17
Georgia	102	112
Florida	64	74
Kentucky	26	27
Tennessee	14	20

Rural (Population Density < 1000)		
State	0-5 Miles	5-10 Miles
Washington DC	0	0
Virginia	944	9,902
North Carolina	18,233	110,308
South Carolina	359	3,014
Georgia	21,473	84,681
Florida	22,310	194,707
Kentucky	5,082	19,060
Tennessee	1,954	9,168

Urban (Population Density > 1000)		
State	0-1 Miles	1-3 Miles
Washington DC	7	7
Virginia	1	1
North Carolina	13	20
South Carolina	4	8
Georgia	17	25
Florida	19	35
Kentucky	4	5
Tennessee	0	1

Urban (Population Density > 1000)		
State	0-1 Miles	1-3 Miles
Washington DC	29,004	88,760
Virginia	56	7,002
North Carolina	6,924	189,286
South Carolina	12	1,426
Georgia	459	207,999
Florida	3,822	195,091
Kentucky	112	2,438
Tennessee	0	2

Table A10: Full Sample Distance Between Facilities and Single Family Homes

Rural (Population Density < 1000)			
Distance To Facility	Always Treated Group	Treatment Group	Total
0 miles - 5 miles	163,783	75,543	239,326
5 miles - 10 miles	270,213	106,030	376,243

Urban (Population Density > 1000)			
Distance To Facility	Always Treated Group	Treatment Group	Total
0 miles - 1 mile	22,582	50,511	73,093
1 mile - 3 miles	130,224	151,861	282,085

A.1.4 Summary Statistics of Distance Between Facilities and Real Estate Data

This subsection provides descriptions and summary statistics of the distances of each real estate transaction location to the facilities in my dataset. The following table shows the sample size of the number of real estate transactions to the facilities at different distances. As discussed earlier, I show the sample sizes divided into a rural and urban category.

Table A10 provides an overview of how many households are near each facility at different distances. The control group or facilities are the facilities that have remained open between 2005 and 2023. The treatment group or facilities are the facilities that have opened or closed between 2005 and 2023.

Because the entire sample is not within 10-miles, I also show how many facilities are captured at different distances. The different cutoffs are arbitrary. My results consider different definitions of "Near" and "Not Near".

As stated above, a fundamental component of the hedonic model is that housing characteristics are a main component in the model. Housing characteristics directly influence the price so if they are not in the model, there is omitted variable bias. The percentage missing of housing characteristics such as year built, acres, bedrooms, bathrooms, and stories is not adequate for a hedonic model. As an alternative, I use house fixed effects, which involves relying on repeat sales, to overcome this issue. In my dataset, there are 557,332 repeated sales. This is adequate for house fixed effects.

Table A12 provides an overview of how many households are near each facility at different distances for the repeated sales sample. The always treated group are the facilities that have remained open between 2005 and 2023. The treatment group are the facilities that have

Table A11: Number of Always Treated and Treatment Groups by Distance and Urban/Rural Classification

Rural (Population Density < 1000)		
Distance To Facility	# of AT Group	# of Treatment Group
0-5 Miles	150	105
5-10 Miles	149	107

Urban (Population Density > 1000)		
Distance To Facility	# of AT Group	# of Treatment Group
0-1 Miles	43	34
1-3 Miles	50	38

opened or closed between 2005 and 2023. Table A13 shows that number of facilities within certain miles ranges which is also similar to the full sample. Table A9 gives a breakdown by state. This breakdown also looks similar to the full sample. A9 shows the sample size now of real estate data for each state.

B Appendix B: Preliminary Results From Machine Learning

This section summarizes attempts to predict house sale prices using available data, such as the year a facility opens, distance from homes to facilities, and geographical indicators like state and tract. Two machine learning models, Random Forest and Extreme Gradient Boosting (XGBoost), were employed to evaluate predictive power. This model uses repeated sales.

Random Forest outperforms XGBoost, achieving a lower mean squared error (MSE) of 0.333 compared to 0.465 for XGBoost, and a higher R^2 score of 0.701 versus 0.583. This indicates that the Random Forest model better captures the relationships between features and sale prices, although both models explain only 58.3% to 70.1% of the variance in sale prices. This suggests that additional features, such as more detailed neighborhood characteristics, could improve model performance.

Figure A1 illustrates the most important variables influencing housing prices, with features like distance to a facility and timing relative to its opening playing a key role. Figure A2 shows the comparison between actual and predicted prices in log form, with reasonable alignment but room for improvement. Based on the R^2, there is improvement that could be

Table A12: Real Estate Sample Sizes for Always Treated and Treatment Groups by Distance and Urban/Rural Classification

Rural (Population Density < 1000)			
Distance To Facility	Always Treated Group	Treatment Group	Total
0 miles - 5 miles	46,748	23,787	70,535
5 miles - 10 miles	76,057	29,295	105,352

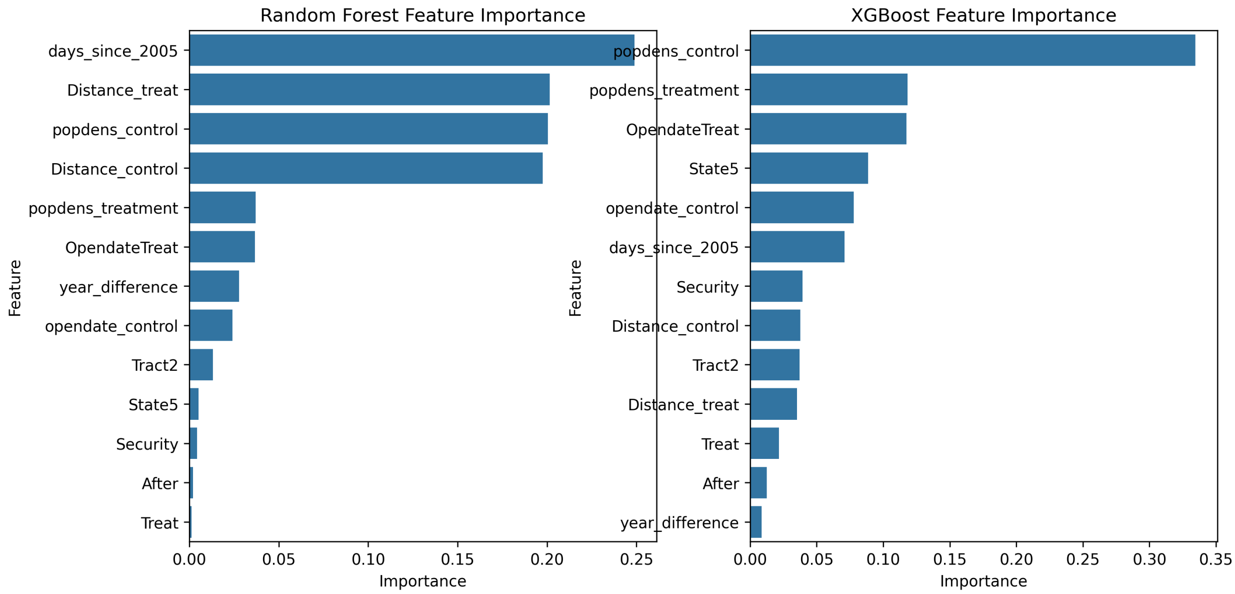
Urban (Population Density > 1000)			
Distance To Facility	Always Treated Group	Treatment Group	Total
0 miles - 1 mile	12,142	28,247	40,389
1 mile - 3 miles	63,769	61,196	124,965

Table A13: Number of Facilities in Each Group Type

Rural (Population Density < 1000)		
Distance To Facility	# of AT Group	# of Treatment Group
0-5 Miles	143	100
5-10 Miles	150	108

Urban (Population Density > 1000)		
Distance To Facility	# of AT Group	# of Treatment Group
0-1 Miles	34	31
1-3 Miles	54	42

Figure A1: Feature Importance



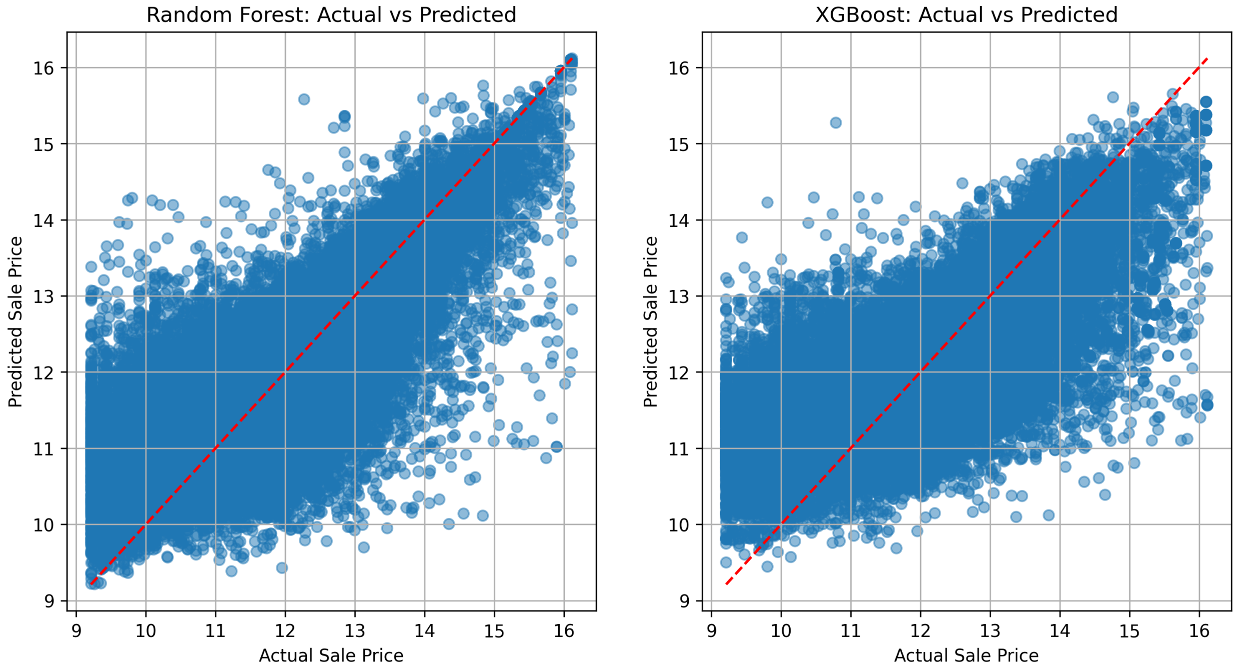
The variables are the following: dayssince2005 are the number of days between the sale of a house and 2005. mindistance is the distance between a facility and a house. popdenscontrol and popdenstreatment are the population densities near each facility at the tract level. OpendateTreat and opendatecontrol is the year a facility in the treatment or control group was opened. yeardifference is the difference between the year a facility opened and the year a house was sold. Tract2 is the specific tract of each house and facility. Security is the type of facility. State5 is the state of the facility. Treat and After are variables from equation (1).

made to make sure the fit is better.

Future iterations of the model could incorporate more neighborhood-level data (e.g., proximity to parks, schools, or other amenities) to further enhance predictive accuracy. The current approach will also inform the analysis in a forthcoming paper called “How Close is Too Close?”, where proximity to amenities and the affect on sale prices is explored in greater detail.

The overall methodology provides a framework for building and evaluating machine learning models in real estate contexts. These results, while indicative, underscore the importance of expanding feature sets to account for a wider range of determinants affecting housing prices.

Figure A2: Actual vs. Predicted Prices in Log Form



C Appendix C: Additional Functional Forms for Continuous Distances

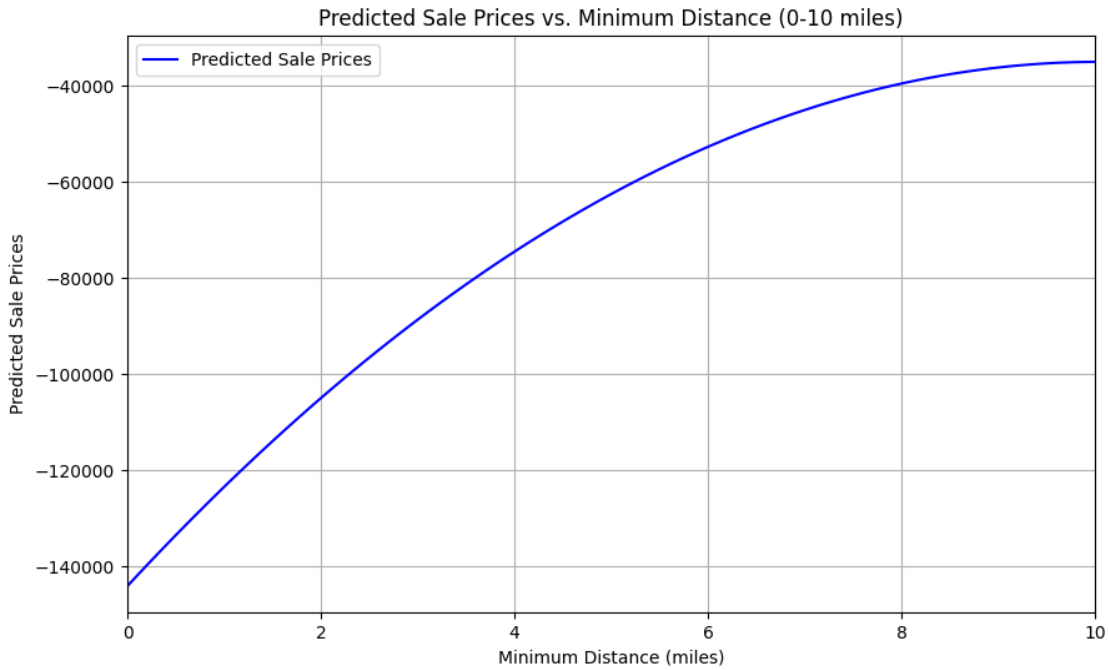
C.1 Quadratic Distances

In this model, I included both linear and quadratic terms for Distance to allow for non-linear relationships between property prices and distance to the facility. By including both the linear interaction term (Distance*Treat*After) and the quadratic interaction term (Distance*Treat*After), the model is able to account for potential differences in the shape of the relationship between property prices and distance. This functional form allows for flexibility in the model.

C.2 Cubic & Quadratic Distance

Here there is both quadratic and cubic terms for the Distance variable. The quadratic term allows for a parabolic response, accommodating scenarios where the effect of distance on sale price may increase or decrease at varying rates. The addition of the cubic term further enhances model flexibility by enabling the possibility of inflection points, where the rate of change in sale prices shifts as distance varies.

Figure A3: Quadratic Distances



Lastly, I show the mean sale price at different bins when Treat = 1 and After = 1 and also when Treat = 0 and After = 1 similar to Figure 6. This also shows that the mean sale price is the lowest near to a facility and steadily increase the further a house is from a facility. Note that this observational data and not results from a regression.

Figure A4: Quadratic & Cubic Distances

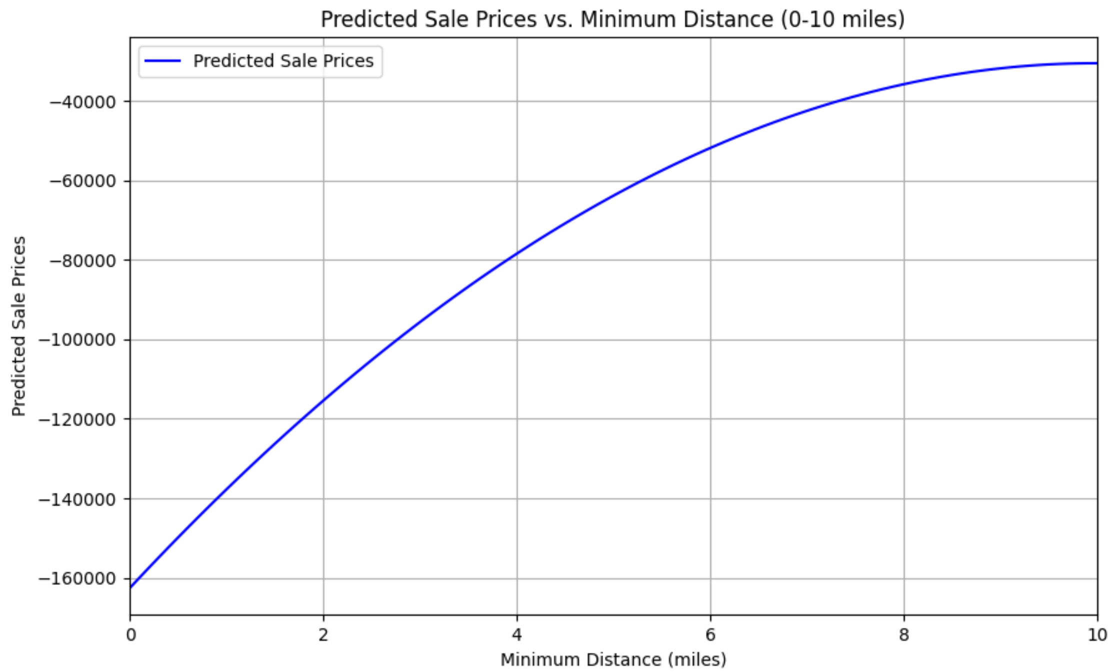


Figure A5: Mean Sale Price at Different Distance

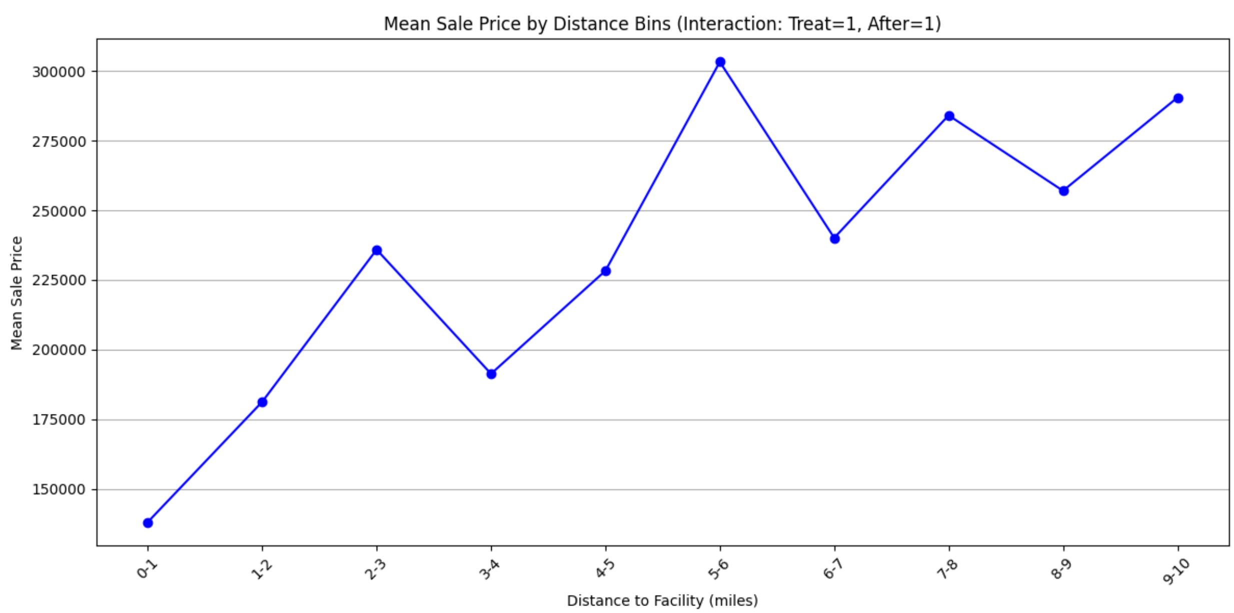


Figure A6: Mean Sale Price at Different Distance