Austin Coffelt
Tanisha Gauns

The Effect of Walkability on Rent

I. Introduction

The US has a very car-focused economy. Most major American economic hubs are designed around personal vehicles, in stark contrast to the rest of the world, which explicitly design their cities and centers of capital around trains and subways. This car-centric infrastructure has its benefits, the biggest being privacy, but its drawbacks are becoming impossible to ignore. According to the US Energy Information Administration (2024), motor fuel consumption accounted for 22 percent of US emissions in 2023. With the increase in climate related disasters year over year, US city planners and developers must start shifting to a new model of urban development, one that deemphasizes the car and focuses on walkability.

US cities in recent years have taken strides to promote walkability, like the removal of parking minimums in Seattle and congestion pricing in New York, but a big component of this problem is the demand side, as the American public must value walkability for it to be a fix for car emissions. We decided to look closely into the demand for walkability, asking the question: What is the impact of walkability on the median rent in a neighborhood? Using the 2019 National Walkability Index from the Environmental Protection Agency (EPA), which gives walkability estimates at the block group level, the US Census' estimates of median block group rent, and controls from the EPA and the US Census, we look for a causal relationship. We also investigate the impact crime has on this relationship, using a high crime binary and an interaction term.

This is a relationship that has been studied extensively in prior research. Two recent papers on this relationship are by Boyle, Barrilleaux, and Scheller (2014), who find walkability has no impact on housing values in Miami after controlling for relevant covariates, and by Chernobai and Ma (2022), who find no general impact in Orange County, though they did find conditional effect that were significant. There is also literature on the impact of walkability on crime from Lee and Contreras (2021), who find a 22 percent increase in robberies in walkable neighborhoods.

II. Data Summary

Our walkability data comes from the EPA's 2019 National Walkability Index. This is a cross-sectional dataset, with 220427 observations for each block group in the US (including Puerto Rico and other non-state territories). We get our outcome, median contract rent, and many controls from the US Census, which has missing values and does not report values outside a certain range for numeric observations. This causes erosion in our sample, taking us from 220427 to 146247 observations. This is the sample of data we use for our model without any crime interactions. Appendix A Figure 1 shows summary statistics for our 220427 observation population as well as our 146247 observation sample.

Our crime data comes from cities and counties that release geolocations of crimes so we can match them up with block groups. It is the three year average crime rate for the given block group for the years 2017 through 2019. This limits our sample to just the block groups in cities that release incident reports, as well as to block groups that had a crime committed in them since they must be in the incident report. The crimes that we are focusing on are violent crimes, property crimes, motor vehicle thefts, and all three added together (referred to as all crimes throughout this paper). Appendix A Figure 3 shows the walkability and rent summary statistics for each category of crime. The threshold for our high crime binary variable is the 75th percentile crime rate for Metropolitan Statistical Areas in 2019 (shown in Appendix A Figure 2).

III. Model Selection and Limitations

To understand the impact of block group walkability on rent, we use a simple OLS regression, with controls accounting for bias to give us a causal estimate. We standardize walkability to give a better interpretation of results, and log rent to account for the skewness of the metric. Our controls were chosen to represent three areas that affect rent and walkability: neighborhood demographic characteristics, characteristics of the commute, and characteristics of the rental properties and household. We used a lasso regression to find relevant controls, and then removed any that did not covary with walkability. A full list of controls can be found in Appendix C. We also added fixed effects at the county level to account for fixed differences in the counties. We use heteroskedasticity robust standard errors to account for non-constant distribution of errors. Our estimating equation with county fixed effects and our matrix of controls and their coefficients is given by:

$$Rent\ (log\ scale) = \beta_0 + \beta_1\, SD(Walkability) + X\,\Pi + \gamma_{county} + u$$

Our estimating equation with county fixed effects and our matrix of controls and their coefficients with our crime binary and interaction term is given by:

$$Rent\ (log\ scale) = \beta_0 + \beta_1\, SD(Walkability) + \delta_0\, High\ Crime + \delta_1 (SD(Walkability) * High\ Crime) + X\,\Pi + \gamma_{county} + u$$

The interpretation of effects for β_1 are in the form of a 1 standard deviation increase in Walkability causes a $100\beta_1$ percent increase in rent.

There are major limitations in our argument for causality that come from omitted variable bias, selection bias, and measurement error. Due to lack of robust data, there are many unobserved variables in our model that cause bias in our estimate. Examples of omitted variables in our model are terrain of the neighborhood (hilliness), average temperature, transit frequency, and median square footage. Expanding on our average temperature variable, we would expect both rent and walkability to decrease as average temperature in an area increases, and by multiplying these effects we get a positive omitted variable bias. This positive bias we expect for all of the listed omitted variables, and thus, we expect our coefficients to be more positive than the real causal effect.

Selection into sample is evident in our crime data. We were only able to get crime data from the cities and counties listed in Appendix C, most of which are large, expensive, and very walkable. We can see the result of our selection bias in Appendix A, where the 25th percentile, median, and 75th percentile observations of walkability and rent are significantly larger for the crime sample. This selection greatly lowers our variance, increasing the standard errors and shrinking the coefficients because walkability is standardized. We also expect walkability to be valued more highly in our crime data due to the politics of the people in those cities and the high walkability baseline, so we expect these coefficients to be positive overestimates. Another form of selection in our crime sample is that to be included in the sample, there must be a crime committed in the block group. This lowers our variance by removing needed observations, though we are unsure whether this has any significant biasing effect on our coefficients.

Measurement error is also an issue with our model, as Census data is inherently an estimate with a margin of error. Our crime data can also be located incorrectly as this process has only improved with time, though these two issues are difficult to account for in our interpretations.

IV. Results and Discussion

Appendix B Figure 1 shows the results of our first model. This model gives us a 1.1 percent increase in median rent for a 1 standard deviation increase in walkability, statistically significant at the 1 percent level. The sharp decrease in the coefficient as controls and fixed effects are added suggest that our omitted variables have a great confounding effect, possibly creating a null result if observed in our model, which would be in line with previous research. With walkability having a definite range (from 1 to 20), we can find the total effect over that range by multiplying our estimate by the range over the standard deviation. Doing this we get a total effect over the range of walkability of 4.9 percent, which shows that this effect is not practically relevant.

Appendix B Figure 2 compares the results of our first model with the results of our second model (using all crime). Here, we see significantly higher standard errors for our estimate, caused by both smaller sample size and less variance. We also see smaller coefficients for the crime data, though because of the smaller standard deviation in this sample, the total and per unit effect is actually larger than the crime. Using the same technique as before, we can see that the total effect over the range of walkability for our crime data estimates is 5.2 percent. This result seemingly confirms our expectation of positive bias due to selection into sample for our crime data.

Appendix B Figure 3 shows the results for our second model with all the different types of crime, reporting the walkability coefficient, the effect on the intercept (high crime), and the interaction term. Again we see that for each type of crime, the per standard deviation change is smaller than our first model, but the total effect over the range is greater. This is the most true for motor vehicle thefts, with the total effect over the range estimated as 7.6 percent, suggesting that the selection bias is most present in motor vehicle thefts. Our interaction terms are all statistically insignificant, though because they are all negative, our biases are pulling them towards zero. The closest coefficient to statistical significance is motor vehicle theft, with a coefficient of -0.007 and a standard error of 0.005, giving a t-stat around 1.3. With the knowledge of our limitations and biases, this is a very interesting outcome, showing that with more precise and expansive data we may be able to uncover some statistically significant effects.

V. **Conclusion**

In this project, we used EPA, Census, and individual city data to investigate the relationship between a block group's median rent and walkability, looking closely at the impact different types of crime have on said relationship. We found through use of OLS with controls a 1.1 percent increase in median rent from a 1 standard deviation increase in walkability, though we expect this effect to be overestimated because of omitted variables. We find no significant effects on our interactions with crime, though again we expect these estimates to be biased positively due to sample selection and omitted variable bias, so we cannot rule out significant interaction effects with crime.

Further research is needed to remove the limitations that hinder our ability to draw conclusions in this paper. We would recommend for further research to follow in the methods of previous published papers, using Walk Score or some other address level walkability score for analysis. To investigate our crime hypothesis, we recommend using a certain radius around a given address for crime rate. Most of all, though, further research into this topic and interaction must find a dataset representative of the population of interest, which this paper is unable to do.

References

Boyle, Austin, et al. “Does Walkability Influence Housing Prices?” *Social Science Quarterly*, vol. 95, no. 3, 2014, pp. 852–67. *JSTOR*, https://www.jstor.org/stable/26612197. Accessed 8 May 2025.

Chernobai, Ekaterina, and Zhongming Ma. “The effect of walkability on house prices.” *Journal of Housing Research*, vol. 31, no. 1, 2 Jan. 2022, pp. 53–73, https://doi.org/10.1080/10527001.2021.2007583.

“Frequently Asked Questions (Faqs) - U.S. Energy Information Administration (EIA).” *Frequently Asked Questions (FAQs) - U.S. Energy Information Administration (EIA)*, U.S. Energy Information Administration (EIA), 2024, www.eia.gov/tools/faqs/faq.php?id=307&t=10.

Lee, Narae, and Christopher Contreras. “Neighborhood Walkability and Crime: Does the Relationship Vary by Crime Type?” *Environment and Behavior*, vol. 53, no. 7, Aug. 2021, pp. 753–86. *EBSCOhost*, https://doi.org/10.1177/0013916520921843.

Appendix A: Summary Statistics

Figure 1: Summary statistics for Walkability for our full dataset and for our data after attrition through merging with our controls, as well as Rent. Standardized Walkability and log scale Rent are used for the regressions

	Walkability (Full Data)	Walkability (Data with Controls)	Rent (Data with Controls)
count	220740	146247	146247
mean	9.54	9.78	947
SD	4.37	4.26	500.2
min	1.00	1.00	101
25%	5.83	6.17	587
50%	9.17	9.50	822
75%	13.17	13.33	1188
max	20.00	20.00	3500

Figure 2: Summary Statistics for Crime. Chart shows the 2019 percentile rankings of crime rate per 100,000 people as well as percent of observations that are above the 75th percentile rate (observations with a 1 in our regression)

	25th Percentile	50th Percentile	75th Percentile (Cutoff for our Binary Variable)	Percent of observations above the 75th percentile
All Crime	1818.85	2437.5	3020.4	30.99%
Violent	222.3	326.8	453.4	38.88%
Property	1560.2	2124.2	2639.7	29.47%
Motor Vehicle Thefts	94.3	162.75	254.55	54.76%

Figure 3: Summary Statistics for Walkability and Rent by different crime categories. Counts are the same for the same categories. Standardized Walkability and log scale Rent are used for the regressions

	Walkability (All Crime)	Rent (All Crime)	Walkability (Violent)	Rent (Violent)	Walkability (Property)	Rent (Property)	Walkability (Motor Vehicle)	Rent (Motor Vehicle)
count	11573		10801		11124		8628	
mean	13.81	1235	13.91	1221	13.85	1228	13.96	1188
SD	2.82	561.03	2.74	553.53	2.81	562.46	2.74	551.32
min	1.83	172	1.83	172	2.00	172	2.67	172
25%	12.17	816	12.33	808	12.33	807	12.33	781
50%	14.17	1151	14.17	1139	14.17	1140	14.17	1077
75%	15.83	1546	15.83	1527	15.83	1535	15.83	1484
max	20.00	3500	20.00	3500	20.00	3500	20.00	3488

Appendix B: Regression Tables

Figure 1: Regression of Walkability (SD) on Rent (log scale) with controls, state fixed effects, and county fixed effects:

Regression Results (All Data)

	Dependent variable: Median Rent (log scale)			
	OLS	OLS	OLS	OLS
Walkability (SD)	0.191^{***}	0.118^{***}	0.058^{***}	0.011^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Controls	No	Yes	Yes	Yes
State FE	No	No	Yes	No
County FE	No	No	No	Yes
Observations	146247	146247	146247	146247
R^2	0.146	0.670	0.744	0.809
Residual Std. Error	0.463	0.288	0.254	0.221
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ Standard errors in parentheses Heteroskedasticity-robust standard errors			

Figure 2: Comparing the regressions of our full dataset versus just the observations with crime data:

No Crime vs. Crime

	Dependent variable: Median Rent (log scale)							
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Walkability (SD)	0.191^{***}	0.118^{***}	0.058^{***}	0.011^{***}	0.072^{***}	0.039^{***}	0.013^{***}	0.008^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.003)	(0.003)	(0.003)
Crime Data	No	No	No	No	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
State FE	No	No	Yes	No	No	No	Yes	No
County FE	No	No	No	Yes	No	No	No	Yes
Observations	146247	146247	146247	146247	11573	11573	11573	11573
R^2	0.146	0.670	0.744	0.809	0.024	0.687	0.757	0.769
Residual Std. Error	0.463	0.288	0.254	0.221	0.460	0.261	0.230	0.225
Note:								$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
								Standard errors in parentheses
								Heteroskedasticity-robust standard errors

Figure 3: Regression of Walkability (SD) and Rent (log scale) with high crime interaction term:

Regression Results (Crime Data)

	Dependent variable: Median Rent (log scale)			
	OLS	OLS	OLS	OLS
Walkability (SD)	0.008^{***}	0.009^{***}	0.008^{**}	0.011^{***}
	(0.003)	(0.003)	(0.003)	(0.004)
High Crime	-0.009	-0.022^{***}	-0.001	-0.020^{***}
	(0.006)	(0.006)	(0.006)	(0.006)
Interaction ($\delta 1$)	-0.005	-0.005	-0.004	-0.007
	(0.005)	(0.005)	(0.005)	(0.005)
Crime Type	All	Violent	Property	Motor Vehicle Theft
Controls	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Observations	11573	10801	11124	8628
R^2	0.769	0.766	0.771	0.773
Residual Std. Error	0.225	0.225	0.224	0.222
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
				Standard errors in parentheses
				Heteroskedasticity-robust standard errors

Appendix C: Controls and Cities for Crime Data

Controls:
Neighborhood Characteristics:

- Percentage of population that are non-white (Census)
- Median household income (log scale) (Census)
- Population density (EPA)
- Housing density (EPA)
- Percentage of households that own no cars (EPA)
- Percentage of households that own two or more cars (EPA)
- Population count that works for low wages (EPA)
- Population count that works for high wages (EPA)
- Percentage of population that works for high wages (EPA)

Characteristics of the Commute:

- Jobs within 45 minute drive (EPA)
- Regional centrality (EPA)
- Percentage of population that drive to work (Census)
- Percentage of population that take public transportation to work (Census)
- Percentage of population who walk to work (Census)
- Percentage of population who start their commute before 6AM (Census)
- Percentage of population with a less than 15 minute commute (Census)
- Percentage of of population with more than 1 hour commute (Census)

Characteristics of the rental property and the people in them:

- Average household size for renters (Census)
- Percentage of households with gas heating (Census)
- Percentage of households with electric heating (Census)
- Percentage of households with no heating (Census)
- Percentage of households with no internet connection (Census)
- Percentage of population living alone (Census)
- Median number of rooms for rental properties (Census)
- Median year property was built (Census)
- Median year tenant moved into the house (Census)
- Percentage of households that pay extra for utilities (Census)

Note: The specific Census resource we are using is the American Community Survey

Cities/Counties with crime data:

- Alameda County, CA
- Austin, TX
- Boston, MA
- Buffalo, NY

- Cambridge, MA (only all crime data)
- Chattanooga, TN
- Chicago, IL
- Cincinnati, OH
- District of Columbia
- Detroit, MI
- Gainesville, FL
- Houston, TX
- Johns Creek, GA
- Los Angeles, CA
- Montgomery County, MD
- New Orleans, LA
- New York City, NY
- Oakland, CA
- Omaha, NA
- Philadelphia, PA
- Raleigh, NC
- San Francisco, CA
- Tempe, AZ

Note: This data was collected from individual cities publishing the data on their websites or data portals