GISC XXXX.XXX – Project Final Report:

Crime and Housing Prices in Dallas County

Tyler Merrill

The University of Texas at Dallas

tjm130330@utdallas.edu

Introduction

Understanding criminal activity has long animated social science research (Fitzgerald 2011), with current interest among researchers not showing any signs of abating: by one estimate, about fifty empirical articles concerning crime are published annually (Pinotti 2025, 208). Indeed, the direct and indirect costs of criminal activity in the United States have been estimated to amount to at least $2.86 trillion per year, or $9,152 per person (Anderson 2021, 876-877). For comparison, total health care costs amount to approximately $3.83 trillion per year (ibid.). Focusing on just some of the costs of crime prevention in the metroplex area, the budget covering the City of Dallas for the 2025 fiscal year allocated the Dallas Police Department $719 million (Bailey Jr. 2025). Any insights that can contribute to policy responses that can help ameliorate current levels of criminal activity thus stand to provide society with significant potential economic benefit. It is in that vein that this paper seeks to shine an empirical spotlight onto the effects of criminal activity in Dallas County.

Investigations into the potential factors of criminal activity are wide-ranging, with studies examining the relationship between crime levels and different dimensions of social life proliferating in the literature. These include studies focusing on family structure (Basto-Pereira and Farrington 2022), inequality (Pazzona 2024), and educational status (Swisher and Dennison 2016), to highlight just a few of them. Some researchers have also turned their attention to teasing out the associations extant between the price of housing and crime levels, trying to map out the interactions between them (e.g., Ihlanfeldt and Mayock 2010, Kortas et al. 2022). Causal findings from such investigations could be incorporated into crime mitigation policies. Econometric work in this vein (De La Paz et al. 2022, Kallberg and Shimizu 2025) has produced mixed results. But if even a predictive association between these societal variables could be established, it might at least bring additional efficacy to data that typically already resides in policy analysts' toolboxes.

Therefore, the purpose of this study is to lay the foundation for a future inferentially sound investigation into any relationship between housing prices and criminal incidents in Dallas County. More specifically, this paper seeks to determine if a higher incidence of police responses in an area to deal with property and violent crime was associated with higher house prices in Dallas County in 2023. Such a positive finding would indicate that levels of crime (as reported in police incident reports) are counterintuitively more common in richer areas rather than in poorer ones and could have ramifications for policing strategies and other dimensions of crime prevention policy.

Background

Ascertaining the association between the price of housing and crime levels has been approached from many empirical angles. De La Paz et al. (2022) used a two-stage approach that involves first calculating relations between house prices and crime types with Los Angeles County data (2022, 11-12). Crime hotspots are estimated as areas where multiple crimes occur in close proximity to each other using a local cluster test (ibid.). These hotspot metrics were then incorporated in an instrumental variable regression model for estimating housing prices (12-13). Ihlanfeldt and Mayock (2010) used an instrumental variables regression that incorporates aggregate neighborhood-level spatial occurrences of crime to estimate different types of crime impacts on housing prices in Florida (2010, 162, 165-168).

Crime hotspot estimates were also generated in Ceccato and Wihelmsson's 2020 study using Swedish data (2020, 92, 94, 96). This was then incorporated into a hedonic regression model to estimate the effect of crime hotspots and other demographic variables on housing prices (92-94). Kallberg and Shimizu (2025) built a hedonic regression model to account for the effects of distances from crime hot spots on housing prices with Seattle data (2025, 3, 16-19). Akinsomi et al. (2025) also built a hedonic pricing model on Johannesburg, South Africa, data to investigate the relations between different types of crime and housing prices.

De La Paz et al. (2022) conclude that while higher crime rates do affect housing prices, how much they do so depends on what types of crime occur (2022, 23, 26). Akinsomi et al. (2025) also found evidence that certain crimes correlated with housing prices differently: robberies were positively associated with higher house prices, while burglaries were negatively related (55, 61-62). This is in contrast with Ihlanfeldt and Mayock's 2010 study, which concluded that both robbery and assault crimes were negatively related to house values (2010, 168-169). Criminal hot spot effects were found to be pertinent by Ceccato and Wihelmsson (2020, 96), who noted that distance away from a crime hotspot was positively associated with housing prices. Kallberg and Shimizu (2025) found that while crime rates were negatively associated with housing prices, being closer to a crime hotspot was counterintuitively positively related to house prices (2025, 32-33).

While it is clear that a correlation between crime and housing prices can often be found, the cardinality of this relationship is not always consistent across locales. Furthermore, the fact that not all types of crime are similarly associated with housing prices suggests that not all crime is impactfully the same across geo-economic strata: Potentially, not all crimes affect those who can afford different types of housing stock equally at different places. The most salient takeaway is that spatial density effects, such as hot spots, should be part of any attempt to capture the interplay of these two phenomena as they are instantiated in Dallas County.

Methods and Data

A bespoke dataset was developed from three separate sources. One is the appraisal data for all accounts in Dallas County for the year 2023 from the Dallas Central Appraisal District (DCAP). The other is ‘Police Incidents’, a record of all incidents involving the Dallas Police Department, hosted by the City of Dallas on the Dallas Open Data web board. The final dataset is zip-code level demographic and economic data extracted from the U.S. Census Bureau’s one-year American Community Survey (ACS) estimates and hosted by the IPUMS database’s National Historical Geographic Information System (NHGIS) repository.

All residential property data is individual-level data denoting owner names, property locations, and additional administrative filing information. Locational data is recorded as both street address data and GIS parcel data. The DCAP data records 7,686,698 properties in total. Police data is incident-level data. Specifically, all reported criminal incidents as recorded in the Dallas Police Department’s Records Management System. It breaks down each incident by the type of criminal activity encountered or reported, demographics of the suspects involved, name and badge numbers of the officer(s) involved, as well as the location, date, and time of the incident. Locational data is given as both street address and coordinate data. The PI data records 144,439 incidents. ACS data from NHGIS was initially composed of dozens of variables for every zip code in the nation. This was pared down to those zip codes of the households that comprised the final dataset. In the R software environment, the elemental data frames were merged on a cleaned and re-formatted convenience sample of matched addresses, resulting in a finalized dataset of 142 observations of Dallas County households with crime incident report information.

The unit of analysis for this study is addresses residing within Dallas County. The dependent variable is the estimated market value of homes located at those addresses. A spatial-error model (SER) of housing price is estimated off select demographic and crime variables of the form: $Price = \beta 0 + \beta X1 + BX2 + \delta,\ \delta = \lambda W\delta + \epsilon$. Where $'\lambda'$ in the residual term captures how much spatial autocorrelation exists between a location's residuals and those of its neighbors (Sun 2025). Spatial autocorrelation being a measure of how often similar values tend to cluster together in space (ibid.). By incorporating this information in this manner, such a model should allow an estimation of house prices based on a set of variables with the impacts of spatial autocorrelation between observations controlled for (Brazil 2024). Taking Kallberg and Shimizu (2025) and Ceccato and Wihelmsson (2020) as guides, the following independent demographic ($X1$) variables were selected: the percentage of males residing in the zip-code of the address, the percentage of females, the percentage of African-Americans, the percentage that identify as Hispanic, the percentage that identify as Asian, the percentage of residents that have received SNAP benefits in an address's zip-code within the last twelve months, and the median income in an address's zip-code over the past 12 months. Additionally, the following crime-related ($X2$) independent variables were incorporated: a proportion variable specifying the precent of crimes committed within the zip code of each house out of all zip codes in the dataset, a factor variable ranking the criminal incident reported at an address as either a 'property' or 'violent' crime, and a variable capturing the Euclidian distance between an address and a custom hotspot: the address with the most reported number of incidents in the dataset.

A table of summary statistics for the eight variables is presented in ‘Figure 1’, and a map of the 142 Dallas County houses made with the ‘folium’ package and an ‘OpenStreetMap’ underlay is presented in ‘Figure 2’ (with the hotspot household marked with a red tab), below:

	tot_val	%male	%female	%white	%black
count	1.420000e+02	142.000000	142.000000	142.000000	142.000000
mean	4.499480e+05	0.491589	0.508411	0.401924	0.263033
std	7.233468e+05	0.023161	0.023161	0.207955	0.185794
min	1.000000e+03	0.447034	0.437866	0.058739	0.011080
25%	1.800625e+05	0.478337	0.503216	0.261039	0.109951
50%	2.360200e+05	0.488898	0.511102	0.331632	0.201873
75%	4.681100e+05	0.496784	0.521663	0.596981	0.439795
max	7.473300e+06	0.562134	0.552966	0.912702	0.742051

	%hispanic	%asian	%SNAP	%crime	median_income
count	142.000000	142.000000	142.000000	142.000000	142.000000
mean	0.808212	0.024662	0.024463	0.043526	69531.521127
std	0.809871	0.027086	0.018228	0.029283	30726.721072
min	0.000000	0.001274	0.000000	0.006452	28739.000000
25%	0.193311	0.004823	0.007267	0.019355	52987.000000
50%	0.531243	0.009714	0.021675	0.032258	57315.000000
75%	1.133978	0.040232	0.031832	0.064516	88464.000000
max	4.350024	0.106976	0.076397	0.103226	189985.000000

	crime_type	ht_dist
count	142.000000	142.000000
mean	0.091549	0.116718
std	0.289410	0.066327
min	0.000000	0.000000
25%	0.000000	0.062169
50%	0.000000	0.118342
75%	0.000000	0.171111
max	1.000000	0.262544

Figure 1

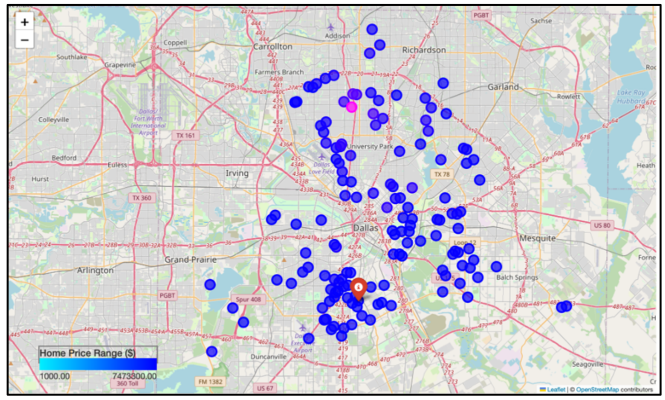

Figure 2

First, evidence of global spatial autocorrelation as regards home prices in the dataset was checked to ascertain if an SER model (which controls for this) was appropriate. Moran’s I is a metric that provides an estimate of how much spatial autocorrelation is evident in a dataset (Sun 2025). Calculating and plotting (‘Figure 3’) Moran’s I[1] from the ‘esda’ package revealed there was evidence to reject a null hypothesis of random spatial lag as regards home prices in the dataset (Moran’s I = 0.4458, Expected-I = -0.007, P-value = 0.001), meaning that spatial autocorrelation was present. Therefore, it was deemed sensible to attempt to use a spatial model to control for the presence of spatial autocorrelation (Tenkanen 2022).

[1] The optimal number of neighbors (K) to build the weights matrix was selected by picking among the lowest K= 1 to 20 possible neighbor numbers that resulted in the smallest AIC score *and* a fully connected weights matrix (K = 5).

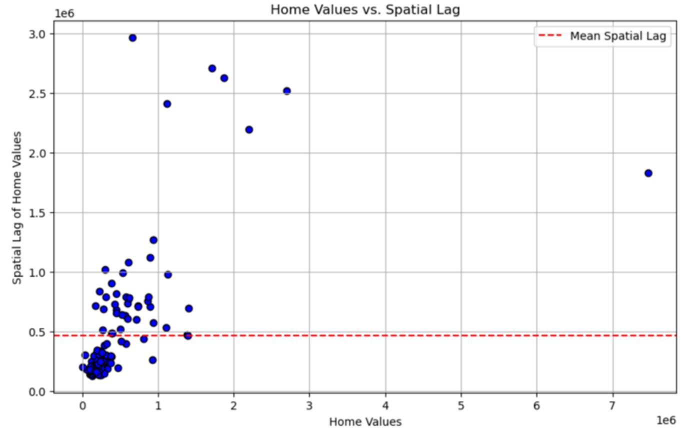

Figure 2

Examining the dependent variable visually ('Figure 4, left') and with a symmetry test from the 'scipy.stats' package revealed this variable to be unsymmetrically distributed and positively skewed (Skew = 6.966 | > 0). Logging the sales price variable results in a more symmetrical distribution ('Figure 4, right'; Skew = -0.887 | ~ 0), so the analysis proceeded with such a transformation.

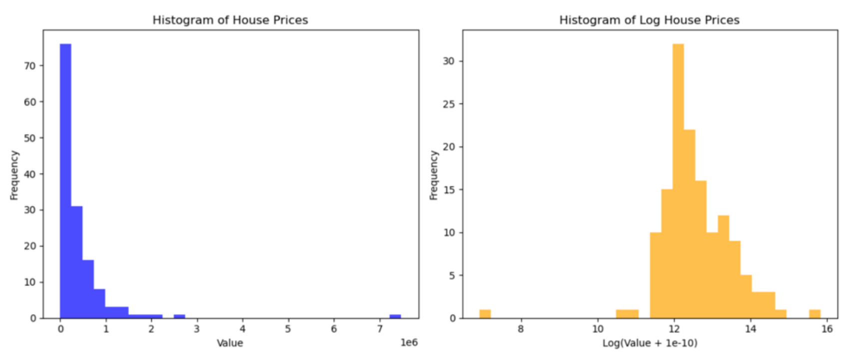

Figure 3

Variable Inflation Factors (VIFs) can be used to estimate how much coefficient intercorrelation and hence multicollinearity might be present in a model (Pardoe et al. 2018). Using the variable inflation factor test function from the 'statsmodels' package ('Figure 5') revealed that the variance of some of the variables was heavily influenced by intercorrelations with other variables (VIF > 10).[2]

	Feature	VIF
1	%female	287.815422
0	%male	193.773542
2	%white	16.815849
8	median_income	9.434925
3	%black	9.134332
6	%SNAP	4.697349
5	%asian	2.536945
9	ht_dist	2.435450
7	%crime	1.680577
4	%hispanic	1.221064
10	crime_type	1.061784

Figure 4

Looking at a scatterplot matrix of the variables (‘Figure 6’) revealed some correlation, again mainly between males and females.

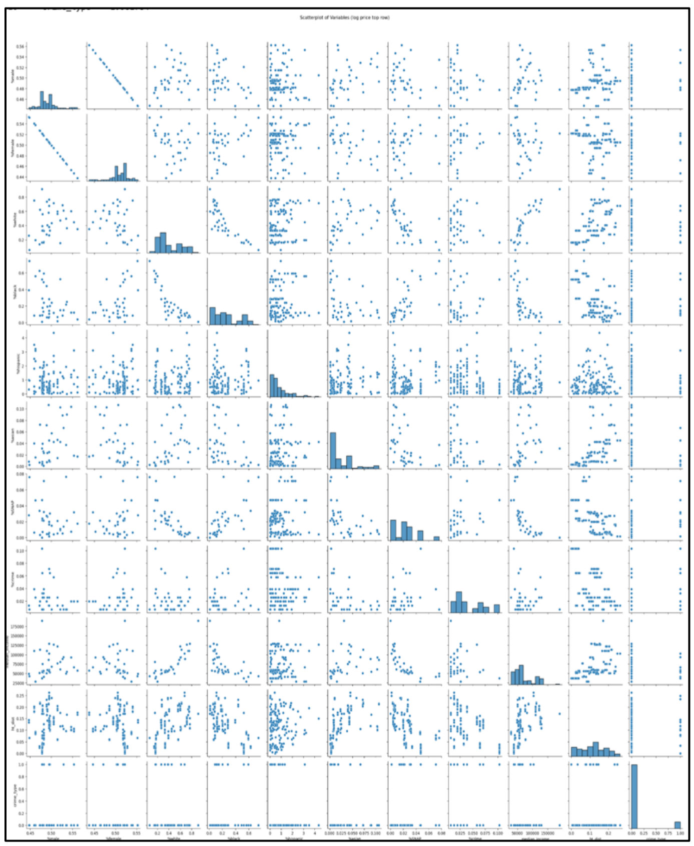

Figure 5

[2] Heuristic from Pardoe et al. (2018).

To account for this and the different types of independent variables in the model, the independent variables[3] were standardized with the 'StandardScaler' function from the 'sklearn' package. Additionally, given the high multicollinearity evident between some of the variables, the variables '%female' and '%white' were dropped from the model. The results of this, as reported in 'Figure 7', display a much less intercorrelated (VIF < 10) set of independent variables to use in the SER model.

	Feature	VIF
4	%SNAP	4.200509
1	%black	3.847964
6	median_income	3.075630
3	%asian	1.912992
0	%male	1.485906
5	%crime	1.484188
7	ht_dist	1.303393
8	crime_type	1.168746
2	%hispanic	1.164626

Figure 7

Building a geodata frame with 'GeoPandas' and constructing a KNN = 5 weights matrix[4] with the 'libpysal' package, the 'spreg' package was used to build a spatial-error model with the 'ML_Error' function, the results of which are presented below ('Figure 8'):

```
REGRESSION RESULTS
------------------

SUMMARY OF OUTPUT: ML SPATIAL ERROR (METHOD = full)
---------------------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :Log of House Price           Number of Observations:         142
Mean dependent var  :     12.5569                Number of Variables   :          10
S.D. dependent var  :      0.9382                Degrees of Freedom    :         132
Pseudo R-squared    :      0.4366
Log likelihood      :   -145.5457
Sigma-square ML     :      0.4423                Akaike info criterion :     311.091
S.E of regression   :      0.6650                Schwarz criterion     :     340.650

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT        12.25885         0.23509        52.14607         0.00000
               var_1        -0.00270         0.08215        -0.03283         0.97381
               var_2        -0.31841         0.13735        -2.31817         0.02044
               var_3         0.01922         0.06101         0.31501         0.75275
               var_4         0.14321         0.09820         1.45846         0.14471
               var_5         0.18299         0.13379         1.36778         0.17138
               var_6         0.14309         0.09728         1.47092         0.14131
               var_7         0.31231         0.11264         2.77272         0.00556
               var_8         2.59553         1.82948         1.41872         0.15598
               var_9        -0.12619         0.19303        -0.65374         0.51328
              lambda         0.39221         0.10472         3.74540         0.00018
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

Figure 8

---

[3] Excepting 'ht_dist', which being measured in Euclidean distance, is already standardized, and 'crime_type', because it is a factor variable.

[4] As above, the number of neighbors for each observation used in the weights matrix (K) was chosen by picking the lowest among the K= 1 to 20 possible neighbor numbers that resulted in the smallest AIC score of the model *and* a fully connected weights matrix (K = 5).

The positive lambda variable (z-score = 3.745, $p < 0.05$) again indicated that spatial autocorrelation was present to be controlled for: similar values were found to cluster together in space. Only the negatively related 'var_2' (z-score = -2.318, $p < 0.05$), the coefficient on '%black', and the positively related 'var_7' (z-score = 2.773, $p < 0.05$), the coefficient on 'median income', were statistically significant among the independent variables. The same variable set was then fed into the 'LazyRegressor' function of the 'lazypredict' package, which allowed a quick derivation of potential comparative models that could be built using the variables. This was done to tease out the relative effectiveness of utilizing spatial autocorrelation in the SER model compared to models that did not utilize it. A truncated selection of the function output is given in 'Figure 9' below:

(Model	Adjusted R-Squared	R-Squared	RMSE	Time Taken
NuSVR	0.32	0.54	0.50	0.00
SVR	0.31	0.53	0.50	0.00
LassoLarsIC	0.29	0.52	0.51	0.00
OrthogonalMatchingPursuitCV	0.28	0.51	0.51	0.01
OrthogonalMatchingPursuit	0.28	0.51	0.51	0.00
LassoCV	0.27	0.51	0.51	0.03
LassoLarsCV	0.26	0.50	0.51	0.01
LarsCV	0.26	0.50	0.51	0.01

Figure 9

Two non-spatial models suggested by the package with comparable RMSEs, Ordinary Least Squares (OLS) and LarsCV,[5] were then directly compared with the SER model according to their Akaike Information Criterion (AIC). This metric allows the ranking of models according to an estimate of which model is most likely (has the highest log-likelihood), given a dataset (Brazil 2024). This metric is also useful for comparing spatial and non-spatial models (Dumelle et al. 2023, 8-9). Running both models with the 'sklearn' package allowed an extraction of the AIC of each model. The summary generated from the OLS model is given below ('Figure 10'). It generated an AIC of '321.1'. The LarsCV model produced an AIC of '314.51'. 'Figure 8' above shows that the SER model produced an AIC of '311.091'.

[5] RMSE = 0.51 and 0.53, respectively. K = 5-fold CV was used with LarsCV.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              log_price   R-squared:                       0.442
Model:                            OLS   Adj. R-squared:                  0.404
Method:                 Least Squares   F-statistic:                     11.60
Date:                Sun, 16 Nov 2025   Prob (F-statistic):           2.71e-13
Time:                        19:34:29   Log-Likelihood:                -150.55
No. Observations:                 142   AIC:                             321.1
Df Residuals:                     132   BIC:                             350.7
Df Model:                           9
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
const            12.3240      0.175     70.254      0.000      11.977      12.671
%male            -0.0261      0.075     -0.346      0.730      -0.175       0.123
%black           -0.2809      0.123     -2.284      0.024      -0.524      -0.038
%hispanic         0.0140      0.066      0.213      0.832      -0.116       0.144
%asian            0.1630      0.087      1.882      0.062      -0.008       0.334
%SNAP             0.1831      0.125      1.465      0.145      -0.064       0.430
%crime            0.1263      0.079      1.604      0.111      -0.029       0.282
median_income     0.3961      0.107      3.713      0.000       0.185       0.607
ht_dist           2.1388      1.420      1.506      0.134      -0.670       4.948
crime_type       -0.1823      0.217     -0.839      0.403      -0.612       0.247
==============================================================================
Omnibus:                       81.132   Durbin-Watson:                   2.132
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              894.992
Skew:                          -1.702   Prob(JB):                    4.52e-195
Kurtosis:                      14.819   Cond. No.                         43.5
==============================================================================
```

Figure 10

The SER had a lower AIC score than either the OLS or LarsCV models, indicating that a greater fit with the data is achieved with the SER model than the others by this metric. However it should be noted that the same (standardized) variables were found to be statistically significant with the same cardinality via the OLS approach as in the SER one. The coefficients output by the LarsCV function indicated that, in addition to the aforementioned relations, the proportion of an address's zip code population that was Asian was positively related to the log of the house's price. But these comparisons do demonstrate it is possible to create a model predictive of house prices in Dallas County that explicitly accounts for the effects of spatial autocorrelation with a (slightly) greater estimated probability than models that do not.

Concluding Discussion

The foregoing analysis did produce a spatial model of house prices in Dallas County of comparable estimation quality to two other models that did not utilize the information contained in the spatial autocorrelations of observations directly. But it did not find a statistically significant association[6] between proxies for crime and house prices. It did find a significant negative association between housing prices and the proportion of an address's zip-code population that is African American, and a significant positive association of housing prices with regard to the median income of an address's zip code.

Ceccato and Wihelmsson (2020) found that a further distance from a crime hot spot resulted in a practically significant increase in estimated housing prices (2020, 98). The variable measure of distance to the singular hotspot utilized in this paper did not rise to such significance. However, a positive relation between hotspots and house prices was noted by Kallberg and Shimizu, at least when spatial autocorrelation was accounted for (2025, 40-41). It is possible that the incorporation of additional hotspots in the SER model might have found significant correlations of either cardinality with house prices. Although ideally, due to the fact that the crime cases incorporated into this study's dataset were pulled as a convenience sample, a greater number of crime records would be desirable to produce a more unbiased sample of crimes committed to better ascertain what parts of Dallas County are exposed to relatively high levels of crime.

Different relations between the types of crime and housing prices were noted by De La Paz et al. (2022, 26) and Akinsomi et al. (2025, 62) and Ihlanfeldt and Mayock (2010, 169-170). However, this paper did not find any difference between 'violent' and 'property' crime to be significant. With the addition of a more representative crime sample, it could become more feasible to contrast more granular categories of crime with housing prices.

It is also important to emphasize that the model specification adopted here is also not inferentially unobjectionable. Many other potential demographic and economic variables could be

[6] At the α = 95% significance level.

incorporated in the SER model presented here. Such additional measures could help overcome the confounding bias inherent in any approach that attempts to account for all the relevant factors of any such relation as that considered in this paper, either *in situ* or in other locational contexts. A clearer picture of the importance of the variable set utilized here could then be made. Additionally, this study, using cross-sectional data, did not examine changes in any of these variables over time. Such a multi-period perspective may reveal seasonal or other temporal patterns in the relation between crime and house prices in Dallas County.

References

Akinsomi, Omokolade, Kola Ijasan, and Peterson Owusu Junior. 2025. "Are Residential Crimes the Same? Evidence from the Link between House Prices and Crime Levels in South Africa." *International Journal of Housing Markets and Analysis* 18(7): 46–65. doi:10.1108/IJHMA-03-2025-0068.

Anderson, David A. 2021. "The Aggregate Cost of Crime in the United States." *The Journal of Law and Economics* 64(4): 857–85. doi:10.1086/715713.

Bailey Jr., Everton. 2025. "Dallas Police, Fire Come out on Top in Proposed Budget. Here's How." *The Dallas Morning News*. https://www.dallasnews.com/news/politics/2025/08/11/dallas-police-fire-come-out-on-top-in-proposed-budget-heres-how/.

Basto-Pereira, Miguel, and David P. Farrington. 2022. "Developmental Predictors of Offending and Persistence in Crime: A Systematic Review of Meta-Analyses." *Aggression and Violent Behavior* 65: 101761. doi:10.1016/j.avb.2022.101761.

Brazil, Noli. 2024. "Spatial Regression.", University of California, Davis. https://geo200cn.github.io/spatialreg.html#Model_Selection.

Ceccato, Vania, and Mats Wilhelmsson. 2020. "Do Crime Hot Spots Affect Housing Prices?" *Nordic Journal of Criminology* 21(1): 84–102. doi:10.1080/2578983X.2019.1662595.

"Census Bureau Data." *United States Census Bureau*. https://data.census.gov/ (September 9, 2025).

"Data Products." *Dallas Central Appraisal District*. https://www.dallascad.org/DataProducts.aspx (September 9, 2025).

De La Paz, Paloma Taltavull, Jim Berry, David McIlhatton, David Chapman, and Katja Bergonzoli. 2022. “The Impact of Crimes on House Prices in LA County.” *Journal of European Real Estate Research* 15(1): 88–111. doi:10.1108/JERER-01-2021-0002.

Dumelle, Michael, Matt Higham, and Jay M. Ver Hoef. 2023. “Spmodel: Spatial Statistical Modeling and Prediction in R” ed. A. K. M. Anisur Rahman. *PLOS ONE* 18(3): e0282524. doi:10.1371/journal.pone.0282524.

Fitzgerald, Charity Samantha. 2011. “Historical Theories of Crime and Delinquency.” *Journal of Human Behavior in the Social Environment* 21(3): 297–311. doi:10.1080/10911359.2011.564954.

Ihlanfeldt, Keith, and Tom Mayock. 2010. “Panel Data Estimates of the Effects of Different Types of Crime on Housing Prices.” *Regional Science and Urban Economics* 40(2–3): 161–72. doi:10.1016/j.regsciurbeco.2010.02.005.

Kallberg, Jarl G., and Yoshiki Shimizu. 2025. “Crime Measures and Housing Prices: An Analysis Using Quantile Regression and Spatial Autocorrelation.” *The Journal of Real Estate Finance and Economics*. doi:10.1007/s11146-024-09997-w.

Kortas, Fabiënne, Alexander Grigoriev, and Giulia Piccillo. 2022. “Exploring Multi-Scale Variability in Hotspot Mapping: A Case Study on Housing Prices and Crime Occurrences in Heerlen.” *Cities* 128: 103814. doi:10.1016/j.cities.2022.103814.

Pardoe, Iain. 2018. “10.7 - Detecting Multicollinearity Using Variance Inflation Factors.” Pennsylvania State University. https://online.stat.psu.edu/stat462/node/180/.

Pardoe, Iain, Laura Simon, and Derek Young. 2018. “10.7 - Detecting Multicollinearity Using Variance Inflation Factors.” https://online.stat.psu.edu/stat462/node/180/.

Pinotti, Paolo. 2025. "The Credibility Revolution in the Empirical Analysis of Crime." *Italian Economic Journal* 6(2): 207–20.

"Police Incidents | Dallas OpenData." https://www.dallasopendata.com/Public-Safety/Police-Incidents/qv6i-rri7/about_data (September 10, 2025).

Schroeder, Jonathan, David Van Riper, Steven Manson, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. "IPUMS National Historical Geographic Information System: Version 20.0." doi:http://doi.org/10.18128/D050.V20.0.

Sun, Shipeng. 2025. "Chapter 4: Spatial Regression in R." In *Data Analysis and Visualization with R: Spatial*, Hunter College. http://www.geo.hunter.cuny.edu/~ssun/R-Spatial/spregression.html.

Swisher, Raymond R., and Christopher R. Dennison. 2016. "Educational Pathways and Change in Crime Between Adolescence and Early Adulthood." *Journal of Research in Crime and Delinquency* 53(6): 840–71. doi:10.1177/0022427816645380.

Tenkanen, Henrikki. 2022. "Tutorial 4 - Spatial Regression in Python.", Aalto University. https://sustainability-gis.readthedocs.io/en/2022/lessons/L4/spatial_regression.html.