What we can learn from small units of analysis

by

Andrew Palmer Wheeler

A dissertation submitted to

The University at Albany, State University of New York

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Criminal Justice

2015

Abstract

The dissertation is aimed at advancing knowledge of the correlates of crime at small geographic units of analysis. I begin by detailing what motivates examining crime at small places, and show how aggregation creates confounds that limit causal inference. Local and spatial effects are confounded when using aggregate units, so when the researcher wishes to distinguish between these two types of effects it should guide what unit of analysis is chosen. To illustrate these differences, I generate simulations of what happens to effect estimates when you aggregate a micro level spatial effects model or presume a neighborhood effects model.

I provide further examples in case studies that examine local, spatial and contextual effects for bars, broken windows and crime using publicly available data from Washington, D.C. Using negative binomial regression models, I estimate that adding a bar to a street unit (street midpoints and intersections) increases the number of Part 1 crimes per year on the local street by around 1 on average, but increase the sum of crime on neighboring streets by 2. I also provide estimates of the selection of bars into criminogenic neighborhoods using a non-equivalent dependent variable design, and estimate the selection effect is 25 percent.

Using 311 calls for service as a proxy for physical disorder, I estimate their effects on crime using fixed effects for omitted neighborhood level variables. 311 calls for service have a consistent positive effect on crime using several different neighborhood boundaries, but the effects are very small. I also show the fixed effects model is not a good fit to the data, and that it potentially introduces artefacts at the boundaries of neighborhoods.

I end the dissertation by building a general model of crime including a variety of variables as well as non-linear terms to account for spatial trends. I show how this model corrects for poor predictions and spatial autocorrelation that was apparent in prior models. In these models prior estimates of the effect of bars on crime are slightly moderated, but the effects of 311 calls for service are still similar in size.

Acknowledgements

While I am proud of this work, this dissertation is not the one I set out to write. I had originally set out to write basically an entirely descriptive, but quantitative, analysis. Harassment by nearly all of the committee members that this is unacceptable I am sure was in my best interest, but I still remained unconvinced of the utility of kitchen sink regression models over simple scatterplots and bivariate correlations.

Despite these disagreements, I have much to owe and I thank all of the committee members. Colin influenced my perspective likely more than he realizes. His class on measurement and crime was the base from which I realized all aggregation bias is the result of simply adding random variables, and the rules for what happens to variances and covariances. Glenn and Shawn both clearly spent much time reading and understanding the draft and giving critical feedback. Glenn I've managed to hound relentlessly for letters of recommendation, and on numerous times I've bothered Shawn with questions via email or dropping into his office uninvited. Graeme was the only voice of consistent, positive support at multiple times throughout the whole process.

I thank Rob for not only advice and listening to me complain, but giving me a job for all of these strung out years in Albany. I can not imagine having a better advisor or friend to help me along the way.

Finally, I thank my wife and son. To you this work is dedicated. I will make sure to build you some more giant canvas frames Mandy. I would surely be the most boring person in the world without you and Lincoln.

Contents

Chapter 1

Introduction

This dissertation is aimed to advance the knowledge of the correlates of crime at small places. With recent motivating examples from the successful work of David Weisburd and colleagues (Weisburd et al., 2012) I presume this is and will continue to be an ongoing popular area of criminological research[1]. The aim of the dissertation is to contribute to knowledge in two specific ways: One, through providing empirical reasoning as to why we want to examine small units of analysis at all, and what type of information we are missing by only examining larger units. The second reason is to advance knowledge about the correlates of crime at small places, with a particular emphasis on bars, measures of physical disorder, and measures of the built environment.

I begin the dissertation by focusing on *why we want to look at small units of analysis at all?* Chapter 2, *What Can We Learn from Small Units of Analysis* introduces examples of how using larger aggregate units of analysis obfuscates the research objective of finding the

[1]Of course this builds on much historical work of crime at small places, such as that of Ralph Taylor and Dennis Roncek, and also builds on the work examining neighborhood effects (Sampson, 2012).

causes of crime. In this chapter, I formulate reasoning why we are interested in examining smaller units of analysis based on causal logic. Through this demonstration, I hope to unify reasoning why we are interested in using smaller units of analysis, as well as how using larger units limits particular types of inferences.

Examples of *what happens* when using ecological units of analysis are already discussed and documented in the social sciences, mainly under the guise of aggregation bias and the ecological fallacy. But, I hope my recapping of the relevant material, along with a mathematical demonstration of how aggregation bias occurs, provides further motivation as to why we would rather examine small(er) units of analysis. In addition to this, I demonstrate how aggregation *creates confounds* in the analysis, and with smaller units of analysis we can identify and test a richer array of place based criminological hypotheses, in particular those concerning spatial spill-over effects.

Chapter 3, *Arguments for Specific Units of Analysis*, provides a review of different motivations researchers have given for using either small or large units of analysis. While it may be in vogue to use small units of analysis, I am critical of prior work that vaguely asserts heterogeneity in smaller units as a reason to justify small units. Although less frequent, I also assess several arguments for using larger units of analysis, and detail how these arguments are misguided by confusion with practical constraints versus logical reasons to prefer larger units of analysis. This hinges on a demonstration of how you can always see the aggregate effects with smaller units of analysis, and only practical constraints are an acceptable reason to choose a larger unit of analysis over a smaller one. The chapter then ends with my reasoning to use the combination of street segments and intersections (what I refer to as street units) as the unit of analysis.

In Chapter 4, *Place Based Theories of Crime*, I suggest a novel way of categorizing the criminological literature which makes it clear how one measures the correlates of crime at places. The measures historically used are the characteristics of people that live at places, the characteristics of the places themselves (e.g. the built environment), and the number of opportunities to commit crime in any particular slice of time and space. In this description I provide a clearer, actionable plan on how we measure the correlates of crime.

Chapter 5 defines spatial effects, present theoretical instances where one might expect to observe spatial effects in criminological research, and briefly describe how one can use spatial regression models to estimate spatial effects. Spatial effects are where a characteristic of a location can affect crime not just nearby, but also further away. One example is that bars can increase crime not only inside the bar, but diffuse offenders and victims into adjacent areas. This chapter is presented because part of the motivation for examining smaller units of analysis is to identify the differences between local and spatial effects which are confounded when using aggregate units.

The subsequent chapters in the dissertation discuss the intended analysis and go into more detail on how the units of analysis and the measures are constructed. Chapter 6, *Data Management of Small Geographies*, is a detailed description of the creation of units of analysis and the process of creating the independent variables. This is very important, as there are a multitude of arbitrary decisions one can make in this step of the analysis, and these are often difficult to reconstruct given the limited descriptions of others' work in journal articles. This chapter also describes the independent variables that will be used in the subsequent analyses, as well as how the spatial weights matrix is constructed.

Chapter 7, *Simulation of a Spatial Process*, is a brief simulated example showing how aggregation bias occurs, why you can not identify a neighborhood level process with only the neighborhood level data, and how a spatial process can look like a neighborhood level process when aggregated up. The Chapters 8 *Local and spatial effects of bars on crime* and 9 *311 Calls for Service and Crime* then present some initial in-depth analysis of the relationship between bars, 311 calls for service (as a proxy measure for physical disorder) and crime.

In Chapter 8, *Local and Spatial Effects of Bars on Crime*, extensive attention is given to accurately describing the size of both the local and spatial effects on crime. This research estimates slightly larger effects of bars on crime than it appears in previous research, with an additional bar causing typically two or more Part 1 crimes per year in this sample. Effects in past research tend to be around an additional one crime per year, although most prior research only evaluates violent crime. Spatial diffusion effects are *larger* than local effects, e.g. adding a bar on Street A causes more crime in total on the neighboring streets than just on Street A, but one would not make that conclusion by just examining the model coefficients. To end the chapter, a non-equivalent variables design is presented to determine if bars self-select into criminogenic locations to begin with. This test presumes that bars have no effect on burglaries, and so the partial correlation between bars and burglaries is an estimate of the selection effect. It is found that bars do have an effect on burglaries, but it is small. So likely *some* of the estimated effect of bars on Part 1 crimes is confounded with selection effects, but not entirely.

In Chapter 9, *311 Calls for Service and Crime*, an alternative neighborhood model of crime is estimated to attempt to see if physical disorder is confounded with unobserved neighborhood effects. The effects of disorder on crime are estimated to be statistically significant but are incredibly small. The small disorder effects appear to not be confounded with neighborhood effects, but the neighborhood effects model of crime does not appear to be a reasonable description of the observed data. In this chapter I also discuss how one can use residual spatial auto-correlation plots to assess if a neighborhood model of crime is reasonable, and these show clear evidence that crime does not conform to discrete neighborhood boundaries.

In Chapter 10, *A General Model of Crime*, I incorporate a series of additional covariates that include place measures of human activity (e.g. shopping malls, bus stops) and non-linear spatial trend terms to attempt to control for spatial auto-correlation. Findings from the previous chapter are largely unchanged, and so the effects of bars and 311 calls for service on crime are not confounded by the omission of the additional covariates. The fit of the model compared to prior iterations appears to be much more reasonable, and outlier locations in previous models are not outliers in the general model. A slight amount of auto-correlation still persists in the model even with the non-linear spatial terms though.

Chapter 11, *Conclusion and Future Goals*, I summarize the findings of the dissertation and discuss future endeavours that will move the field forward in modelling crime at micro places.

Chapter 2

What Can We Learn from Small Units of Analysis

As opposed to beginning the dissertation with a review of the relevant theoretical literature, I start here, and address the question, *why should we examine small units of analysis?* This is not because theory is irrelevant or less important, but because this methodological question is pertinent to ask regardless of theory. Who cares if one uses a street segment, a census block, a census tract, a city, county or even a country? Chapters 2 and 3 are aimed at demonstrating *why* we want to examine small units of analysis (Chapter 2), and a review of past justifications to use small units of analysis (Chapter 3). This knowledge will then be used to intelligently choose a unit of analysis.

2.1 How Aggregation Bias Occurs

Prior literature in many different fields has described problems when analyzing aggregate data. While several different problems have explicit names attached to them, such as the ecological fallacy (Robinson, 1950), change of support (Gotway and Young, 2002), and the modifiable areal unit problem (Oppenshaw, 1984), these problems have been demonstrated to be essentially the same issue (or the same potential problems, mainly different types of confounds, occur in all of these instances) (Cressie, 1996; Diez-Roux, 1998; King, 1997). Hence from this point forward, as a goal of this research is to estimate causal relationships, I will refer to the problem as solely aggregation bias; how different units of analysis impact the ability to uncover the true empirical relationship between variables.

One could avoid mathematical formalism and use a fairly simple anecdote to describe *when we want to examine smaller units of analysis.* Imagine a city that has multiple bars. Also pretend that crime never occurs on any street that a bar is located on, but crime is more prevalent on streets nearby those bars. If one aggregates up to the neighborhood level, one would lose the ability to identify that bars aren't directly associated with crime on the local street, but are potentially associated with crime on nearby neighboring streets. *Aggregation creates confounds*, and does not allow one to identify the difference between local, spatial and neighborhood (i.e. contextual) level effects. **To the extent that we care about such differences, we should want to use smaller units of analysis.** Clarification on what these particular effects are and examples will be given later in this and future chapters, but a brief explanation is: local effects are that a characteristic of that place affect crime specifically at that same location, spatial effects are when a location can impact crime not only locally but can affect crime in nearby areas, and a contextual (neighborhood) effect is when crime at a particular location is affected by a particular context and all locations within that context are all affected equally.

It is not always necessary for a research design to be able to distinguish between local, spatial and contextual effects, e.g. in citywide policy evaluations *the aggregate* effect is of interest, to determine the overall benefit or harm of a particular policy. However, this will provide the framework for choosing a smaller unit of analysis, what particular research questions *need* smaller units of analysis, and what we might learn from examining smaller units of analysis that we can not learn from examining larger ones. Distinguishing the differences between local, spatial and contextual effects is useful for both furthering criminological theory, and is pertinent to particular policy evaluations that might have different contributions to local and spatial effects. For example, a police camera may decrease crime locally, but increase crime in other nearby locations (i.e. displace crime). If one aggregates to a too large of unit, it may be that the two effects cancel out, although the camera may be considered a partial success evaluating at the local level. Distinguishing between the local and spatial effects provides a more in depth understanding of what phenomena is under study. Even for the city wide policy implementations it would almost always be useful to see if there were differential effects in different neighborhoods, e.g. if a gun buy back program not only decreased gun violence citywide, but if it decreased gun violence in some neighborhoods more than others. This effect heterogeneity is not identified when working with the macro level city as the unit of analysis.

A useful point of departure is to demonstrate under what circumstances aggregation bias does not occur, and this shows why aggregation bias regularly occurs in practice. Past simulation studies have demonstrated that random aggregation (i.e. randomly designating different units to be aggregated, as opposed to aggregating based on spatial contiguity) does not result in bias of the correlation coefficient between two variables (Blalock, 1972; Oppenshaw, 1984). What follows is an algebraic explanation of why when one randomly aggregates units this bias does not occur[1].

2.2 Proof that Random Aggregation is Unbiased

One can represent the disaggregated correlation coefficient as;

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \tag{2.1}$$

Which can subsequently be represented in terms of the units of analysis as;

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2.2}$$

For simplicity in notation I assume we are working with mean centered data, hence equation (2.2) can be reduced to;

$$\rho = \frac{\sum_{i=1}^{n}(x_i)(y_i)}{\sqrt{\sum_{i=1}^{n}(x_i)^2}\sqrt{\sum_{i=1}^{n}(y_i)^2}} \tag{2.3}$$

[1] This contribution is not unique, and several other works have gone through similar expositions (Cressie, 1996; Firebaugh, 1978; Grunfeld and Griliches, 1960; Hammond, 1973; Hannan and Burstein, 1974; King, 1997; Langbein and Lichtman, 1978; Theil, 1965). In my own unique formulation I hope to present a clear and concise representation, as well as to reiterate old knowledge that still does not appear to be well known or understood.

Here the subscript i is used to represent the disaggregated items, and n to represent the total number of disaggregated items. Also $\bar{x}$ and $\bar{y}$ represent the means of the disaggregated items. When referring to observed disaggregated units of analysis, I use lower case characters, and upper case for aggregate units.

But typically we are working with aggregated items, so all we can observe is;

$$\rho = \frac{Cov(X_j, Y_j)}{\sqrt{Var(X_j)Var(Y_j)}} \tag{2.4}$$

Where the subscript j represents the aggregated unit. In equation (2.5) the aggregated form of the correlation coefficient is represented in terms of the aggregated units;

$$\rho = \frac{\sum_{j=1}^{m}(X_j - \bar{X_j})(Y_j - \bar{Y_j})}{\sqrt{\sum_{j=1}^{m}(X_j - \bar{X_j})^2}\sqrt{\sum_{j=1}^{m}(Y_j - \bar{Y_j})^2}} \tag{2.5}$$

Where the j subscript denotes the aggregated unit and m denotes the number of aggregated units. Since all the aggregated units are sums of random variables, I will replace X_j with the sum of the random variables within X_j, and represent this as the sum of all x_i elements where the individual unit i is nested within the aggregated unit j;

$$\rho = \frac{\sum_{j=1}^{m}(\sum_{i=1\in j}^{k} x_i)(\sum_{i=1\in j}^{k} y_i)}{\sqrt{\sum_{j=1}^{m}(\sum_{i=1\in j}^{k} x_i)^2}\sqrt{\sum_{j=1}^{m}(\sum_{i=1\in j}^{k} y_i)^2}} \tag{2.6}$$

Here k is the total number of units summed within each aggregated unit. For simplicity it is assumed the aggregated units in the above equation are mean centered, and that each $y_i \in j$ has a corresponding $x_i \in j$ to which it discretely maps. Subsequently, the inner sums in the numerator in equation (2.6) can be represented as;

$$(\sum_{i=1\in j}^{k} x_i)(\sum_{i=1\in j}^{k} y_i) = \sum_{i=1\in j}^{k}\sum_{i=1\in j}^{k}(x_i)(y_i) \tag{2.7}$$

(i.e. the product of the sums equals the sum of the product of all pairwise combinations). If one is assuming *random* aggregation, the mean of products not indexed as the same element (i.e. when $Y_i \in j$ and $X_i \in j$ do not map to the same disaggregated element) would be expected to be zero by definition, else it is not random aggregation! As such, equation (2.7) can be represented as a matrix;

$$\sum_{i=1\in j}^{k}\sum_{i=1\in j}^{k}(x_i)(y_i) = \begin{bmatrix} (x_1y_1) & \cdots & \cdots & (x_1y_k) \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ (x_ky_1) & \cdots & \cdots & (x_ky_k) \end{bmatrix} \tag{2.8}$$

Because of random aggregation, one would only expect the cross products of elements indexed at the same unit to be non-zero. If one thinks of a univariate distribution, and all the observations are independent and identically distributed (e.g. no autocorrelation of the observations), knowing one observation gives you no information about the other observations in the distribution. The same logic extends to a multivariate distribution. Even if the correlation between x and y is well defined, knowing the observed value of x and/or y for one realization of the process (i.e. one unit) gives you no information for other units, unless of course the units themselves have some type of autocorrelation. Random aggregation inherently suggests they do not; a unit has no auto-correlation with another unit chosen at random. This is the most important point to make, as it demonstrates why aggregation bias occurs so regularly. When one is not aggregating randomly, assuming the inter-item covariances is zero is untenable in most circumstances, as the majority of spatial data have positive spatial autocorrelation, e.g. Tobler's law of geography "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p. 236).[2].

For all of the products in the matrix in equation (2.8), only on the diagonal would we expect on average non-zero values. Given this assumption equation (2.8) reduces to just one sum;

$$\sum_{i=1\in j}^{k}(x_i)(y_i) = \begin{bmatrix} (x_1y_1) & 0 & 0 & 0 \\ 0 & (x_2y_2) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & (x_ky_k) \end{bmatrix} = \begin{bmatrix} (x_1y_1) \\ (x_2y_2) \\ \vdots \\ (x_ky_k) \end{bmatrix} \tag{2.9}$$

Inserting this back into the numerator for equation (2.6), one then has;

$$\sum_{j=1}^{m}(\sum_{i=1\in j}^{k}(x_i)(y_i)) \tag{2.10}$$

Which is just the sum of a sum, hence we end up with equation (2.11) as the sum of all k items within m units which encompasses all of the n disaggregated elements;

$$\sum_{i=1}^{n}(x_i)(y_i) \tag{2.11}$$

[2]The sum of the off-diagonal covariances can be thought of as Firebaugh's rule (Firebaugh, 1978).

Which is the same as we started with. The same logic functionally applies to the denominator of equation (2.6), and so one starts with;

$$\sqrt{\sum_{j=1}^{m}(\sum_{i=1 \in j}^{k} x_i)^2}\sqrt{\sum_{j=1}^{m}(\sum_{i=1 \in j}^{k} y_i)^2} \tag{2.12}$$

The variance of the sum of random variables is equal to the variance of each individual unit plus two times the inter-item covariances. Again in this instance when we are assuming random aggregation, the inter-item covariances would have an expected mean of zero. The variance of the sum of the disaggregated items will simply equal the sum of their individual variances. Equation (2.13) can then be rewritten as the sum of the sums in the same manner as was done for equation (2.11);

$$\sqrt{\sum_{i=1}^{n}(x_i)^2}\sqrt{\sum_{i=1}^{n}(y_i)^2} \tag{2.13}$$

Inserting that back into the denominator one ends with the original equation for the correlation;

$$\rho = \frac{\sum_{i=1}^{n}(x_i)(y_i)}{\sqrt{\sum_{i=1}^{n}(x_i)^2}\sqrt{\sum_{i=1}^{n}(y_i)^2}} \tag{2.14}$$

The important point to note here is that when non-random aggregation occurs, the covariances between variables and within variables will likely not be zero on average. The majority of past simulations aggregating actual data demonstrate that with random zoning (i.e. grouping units based on spatial proximity) the correlation coefficient frequently increases (Blalock, 1972; Gehlke and Biehl, 1934; Robinson, 1950). This would suggest that the between variables covariance within groupings is positive, and that neighboring spaces had a cross correlation within the units that were being aggregated[3].

This has logical connections to the way we perceive and model spatial relationships. If the between item covariances were positive, we may interpret that as a spatial effect. Hence with the aggregated units, we can not appropriately decompose the effects of the within item units and the neighboring units. This is an inherent confound when analyzing aggregated units. Note here that one can not even make inferences about aggregate (i.e. contextual) effects with the aggregate level data, as the contextual effects are confounded with the individual level effects. This seems to be a lost notion in much contemporary research, in particular those studies purportedly testing the differences between different units of analysis (Hipp, 2007). Such differences between different units of analysis can be decomposed into the relationships I noted above, so arguing that in some instances a larger unit of analysis is appropriate is misguided.

What proceeds are some examples illustrating the differences between local, spatial and contextual level effects.

2.3 Confounded Effects with Aggregated Units

The above mathematical demonstration pertains to solely the bivariate case, but the same principles apply in general to the case in which one is examining the coefficients in a multiple regression equation. In general, the aggregation bias listed above is just a special case of

[3]Although it is likely that the within variables covariances is non-zero as well the Cauchy-Schwarz inequality prevents the ratio of the variance of the numerator and the denominator from ever being greater than 1 (Cressie, 1996).

omitted variables bias.

Consider the following set of equations;

$$y_1 = \beta_1(x_1) + e_1 \tag{2.15}$$

$$y_2 = \beta_2(x_2) + e_2 \tag{2.16}$$

If one were to aggregate y_1 and y_2, one would then have the equation;

$$(y_1 + y_2) = \beta_1(x_1) + \beta_2(x_2) + (e_1 + e_2) \tag{2.17}$$

Considering that with aggregation one would not have the individual x's either, depending on the context it might be reasonable to assume that $\beta_1 = \beta_2$. If we are thinking in terms of individual level units (that is y_1 and y_2 are two individual level observations), without this assumption one has an incidental parameter problem. Although it is not inherently reasonable (e.g. treatment heterogeneity would suggest this is not true), but assuming such an equality one can then write the aggregate level equation as;

$$(y_1 + y_2) = (\beta_1)(x_1 + x_2) + (e_1 + e_2) \tag{2.18}$$

Or one could replace β_1 with β_2 (as they are equal). So here we can see that there is no inherent problem with aggregating the two individual level equations, and one will recover the same information from the aggregate level equation as the individual level equation *if* x_2 is not correlated with y_1 (and vice-versa if x_1 is not correlated with y_2). If those particular correlations were non-zero, say y_1 and y_2 are crime on neighboring streets, and x_1 and x_2 represent the number of coffee shops on a particular street. It is possible that the benefits of those coffee shops (a sign of neighborhood gentrification) spill over onto the neighboring street (Papachristos et al., 2011). This would be considered a spatial effect.

If one is aggregating spatial effects one will not be able to identify the average local effect, as one will have additional confounded terms in the aggregation. A spatial effect is observed when some phenomenon at a particular location is caused by attributes at other (nearby) locations. A theoretical example in criminology might be bars, especially when they close for the night. The bar attracts a lot of people (some deviant, and some being intoxicated become more vulnerable to being victimized), and when the bar closes, many people all at once are forced to leave the bar. The neighboring streets may have an elevated amount of crime at that time, in places spatially proximate but not directly outside of the bar. Thus the bar has a spatial effect on the amount of crime in the neighboring areas.

To be explicit, consider the aggregation of an equation with neighboring effects, where x_1 and x_2 are neighbors in space;

$$y_1 = \beta_1(x_1) + \beta_2(x_2) + e_1 \tag{2.19}$$

$$y_2 = \beta_1(x_2) + \beta_2(x_1) + e_2 \tag{2.20}$$

$$(y_1 + y_2) = \beta_1(x_1 + x_2) + \beta_2(x_2 + x_1) + (e_1 + e_2) \tag{2.21}$$

Again, one only has the aggregate of the X, so the estimated effect (in terms of the original

Table 1: Disaggregated Broken Windows and Crime

Street	Neighborhood	BW	Crime
1	1	5	35
2	1	10	25
3	2	6	30
4	2	8	26

units of analysis) will simply be;

$$(y_1 + y_2) = (\beta_1 + \beta_2)(x_1 + x_2) + (e_1 + e_2) \tag{2.22}$$

In this example one has absorbed both the local effect and the spatial effect into the same parameter estimate. When one only has the aggregated items, this effect can never be *decomposed* into the individual effects. This is not so problematic in the case that no spatial effects occur (e.g. random aggregation), as the average treatment effect will be the same. This is unlikely to be the case with most data in criminology though, and so by using aggregate data these unique effects are inherently *confounded* in the research design. Say in this example x_1 is the number of bars on the street, and x_2 is the number of bars on the neighboring street. One can not identify whether the aggregate effect is due entirely to local bars or partially to both local and neighboring bars.

To put a face on this example, say there are four streets nested within two neighborhoods. In Table 1 BW pertains to the number of broken windows on that street, and Crime pertains to the count of crime on that street. Street pertains to the unique identifier of a street, and neighborhood pertains to what neighborhood that street is nested within. In Table 1 *I* *defined* the count of crime according to the arbitrary function;

$$Crime = 1 \cdot (BW_{Local}) + 3 \cdot (BW_{Neighbor})$$

Where BW_{Local} is the number of broken windows on the local street, and $BW_{Neighbor}$ is the number of broken windows on the neighboring street (the opposite street within the same neighborhood is the neighboring street). Here the neighboring street could be considered a spatial effect, and would have obvious theoretical interest. Broken windows theory as originally conceptualized may be amenable to spatial effects, but it certainly would not make any sense for the spatial effects to be stronger than the local effects!

Now what happens to the estimated effect of broken windows on crime if you aggregate to the neighborhood? Here by aggregate I just mean sum the items within each neighborhood. One would then have Table 2. And hence upon estimating the effect of broken windows on crime, one would then have the resulting equation of;

Table 2: Aggregated Broken Windows and Crime

Neighborhood	BW	Crime
1	15	60
2	14	56

$$Crime = 4 \cdot (BW)$$

The estimated effect is just the local and the neighboring street effects added together.

This does not necessarily create inherent problems when we are trying to represent relationships at the aggregate level. If the theory in which we are modelling relationships suggests that all attributes within the region (regardless of their co-location within that region) can affect other attributes within the region then representing the model with the aggregate units is not disingenuous. Policy level implementations for some larger polity would generally fit this definition. Also if one simply wants to estimate the impacts of some effect in general (such as the estimated effect some policy might have), the overall effect will be of interest (Jargowsky, 2005; Tcherni, 2011) (albeit largely unenlightening as to the processes that formed that effect). Going with the same previous example, one would have the total effect of broken windows on crime, which could be of interest in and of itself.

2.4 How They Are Formed Matters

There are other problems with the aggregation process that will have deleterious effects on the effects estimated. If one aggregates according to the y values in the above equations, the independent variables will be correlated with the error term, and hence the estimated effects will be biased (Blalock, 1972; Hammond, 1973; Langbein and Lichtman, 1978).

Fortunately this is likely not the case with most of the research done in criminology, as frequently aggregated units used are defined by the Census Bureau to be homogeneous according to some demographic characteristics and not crime. An instance in which this is problematic in the criminological literature is when the units of analysis are defined by some type of clustering criterion or by the level of crime itself. An example of this is in McCord and Ratcliffe (2007) (or the suggestion in Groff (2011)). In these cases, if one chooses to use these particular units of analysis, the estimated effects of independent variables will be biased as the independent variables are correlated with the error term.

2.5 How You Should Choose Your Unit of Analysis

Hopefully this gives some indication where to start in selecting an empirically and theoretically appropriate unit of analysis. If one wants to identify the effect of X on Y at some logical level of aggregation, and they want that effect to be independent (to the extent possible) of co-grouping effects caused by the aggregation, one needs to find the smallest unit of analysis at which random aggregation can be assumed (i.e. no spatial effects), or live without being able to distinguish between local and neighboring effects. Another possible alternative is to define the unit of analysis based on an *independent* variable of interest (Ratcliffe and Taniguchi, 2008; Taniguchi et al., 2011; Wang, 2005). Although this does not solve the inability to decompose between local and neighboring effects, this type of aggregation does not bias the aggregate affect between the independent variable of interest, X, and Y. Co-grouping (e.g. spatial or contextual effects) can be theoretically plausible at *any* level of aggregation, and so one should strive for small enough units of analysis where the distinction is not material in how one interprets the information (in addition to practical constraints).

As an example to illustrate these points I will use the affect of alley gating on burglary (Bowers et al., 2004). For an ecological analysis using some type of spatial unit there are likely several choices, such as an individual house, a street block, or some larger census aggregation or representation of a neighborhood. With an individual house as the unit of analysis, one would be implicitly limited in the inferences one can make about the relationship between alley gating houses and burglary imposed by the aggregation. For a hypothetical example, if you used individual parcels and found that alley gating had no effects, you would not be able to distinguish if burglaries decreased at the rear of the residence but increased at the front of the residence (i.e. the method of entry was displaced). That is unless you *assume* such spatial effects do not exist, which you can not determine from the aggregated data at hand.

If one were to use a neighborhood as the unit of analysis the same logic holds, although at larger units of aggregation one has various other effects that need to be accounted for to prevent spurious associations. Although this is true for the smaller units of analysis, it can become more problematic in larger units of analysis. The same problem occurs when one is trying to identify contextual (or neighborhood) effects (Hauser, 1970, 1974; Manski, 1993; Oakes, 2004). The individual level equation needs to be perfectly specified, else the higher level effect could simply be a spurious artifact of an aggregated individual level effect. This is referred to as the atomistic fallacy (Oakes, 2004)[4]. It is very likely omitted spatial effects would result in spurious contextual level effects. This I feel is not well articulated in prior research, and the plethora of research that utilizes multi-level models for causal inference are suspect to this critique (Gelman et al., 2001).

2.6 Conclusion

Aggregation creates an inherent confound in identifying causal relationships. TO further understanding of criminological theory it is necessary to examine the correlates of crime at micro places. Such a refinement in our understanding of the processes (local, spatial and contextual) that cause crime can only occur by examining smaller units of analysis, as such

[4]Oakes (2004) suggests a solution is to use exogenous community interventions (i.e. experimental or quasi-experimental research designs) aimed at neighborhoods. This does not allow one to identify the difference between individual level (or local) effects and neighborhood level effects though. The difference between the two is still confounded when using neighborhood level data.

distinctions are confounded when using larger areal units. Hubert Blalock had a concept that one could continuously zoom in to causal effects, and continuously be more specific about causal processes (Blalock, 1972). This zooming down to smaller geographies is simply an extension of this reductionism to geographic units of analysis.

In the next chapter I present a review of prior units of analysis used in criminological research justifying different units of analysis based on the presentation of aggregation bias in this chapter.

Chapter 3

Arguments for Specific Units of Analysis

Given knowledge of how aggregation bias occurs, it will be useful to evaluate the claims made by prior research to use one unit of analysis over another. This section will consider both arguments for the use of smaller units of analysis (Andresen and Malleson, 2010; Groff et al., 2010; McCord and Ratcliffe, 2007; Oberwittler and Wikstrom, 2009; Roncek, 1975, 1981; Smith et al., 2000) and arguments for the use of larger units of analysis (Raudenbush and Sampson, 1999). The section will then conclude with situations in which smaller units of analysis are likely preferable, extending prior arguments for smaller units of analysis based on theoretical relevance of micro level processes, the fact that neighborhood level processes can not be validated by using neighborhood level data, and that aggregation bias is a realistic threat to the validity of the study of crime when using aggregate units.

The chapter will end with reasoning for using street segments and intersections as a unit of analysis in this particular study. This will be based on the empirical discussion about how aggregation bias occurs, that street segments offer a reasonably small unit of analysis in which local and spatial effects are immaterial, and for other practical reasons (which includes using pre-aggregated data).

3.1 Arguments for Smaller Units

The majority of prior literature suggesting the use of smaller units of analysis makes amorphous reference to heterogeneity as a reason to prefer smaller units (Andresen and Malleson, 2010; Groff et al., 2010; Roncek, 1975; Smith et al., 2000). The following sections will detail two different interpretations of heterogeneity, between units and within units, and explain how that heterogeneity in and of itself is an insufficient argument for the use of smaller units.

3.1.1 Within unit heterogeneity

While most prior research cited is ambiguous about what exactly they mean when they refer to heterogeneity, it seems most research is referring to within unit heterogeneity. That is to say, when comparing an observation at a larger level to the lower level, the variable of interest shows variation between the observations at the lower level that are *within* the larger level. An example of this is in Roncek (1975, p. 846);

> The basic problem is that, as the degree of aggregation increases, so does the magnitude of the correlations . . . This is due to the heterogeneity of land uses and densities within each area which are then combined to obtain an average density. It may be that crimes are committed or criminals reside in the low-density or uncrowded parts of these areas or the area may contain high-density, high crime pockets, but this cannot be discovered

by using large areas.

Given the definition of aggregation bias in the prior section, the above argument is insufficient to demonstrate that bias would occur between density and crime at larger levels of aggregation. One would have to further this argument by stating *what exactly is correlated with density at the lower unit of analysis that also causes crime.* Then if the confounded variable is not omitted at the aggregate level, one has to go on and explain how the variable is correlated with density within the aggregated units of analysis in different areas (e.g. spatial/contextual effects) and how those effects cause crime. This would be a much more detailed and nuanced argument than is presented in either the past quote or the majority of past research.

It should be also be obvious that the statement, within unit heterogeneity exists, is insufficient to choose a specific unit of analysis, because it could be applied *Ad infinitum.* Why do we not choose to examine the differences in correlates of crime between the front stoop and the back porch, as they inevitably show some variation between them for some correlate that could be theoretically related to crime?

Some work goes beyond just stating heterogeneity within units exists, and then goes on to note that clustering of observations within units is a reason to prefer smaller units of analysis (Andresen and Malleson, 2010; Lum, 2011). This also fails to meet the conditions necessary for aggregation bias to occur. Again, one needs to relate how the particular independent variable is correlated with another variable in space and how each cause crime. For example, say that crime is a linear function of social disorganization (where *sd* stands for a measure of social disorganization);

$$Crime = \beta{\cdot}sd + e$$

Then go on to state that social disorganization itself has some type of spatial organization (where d is the distance from the city center);

$$sd = \beta{\cdot}d + e$$

In this scenario, both crime and social disorganization would have a specific spatial organization (and exhibit spatial autocorrelation or clustering of crime incidents). Still no aggregation bias would occur for the estimate of the effect of social disorganization on crime in the aggregate units (assuming units weren't aggregated according to the amount of crime in that unit), and one had an appropriate aggregate measure of social disorganization (the mean in this example), because there are no omitted effects from the aggregate model[1].

This note is particularly salient for theories of crime which argue that it is inappropriate to represent their causal constructs at larger units of analysis. In the future chapter on place based theories of crime, this is given as a reason why some empirical examinations of defensible space theory have returned null findings (Taylor and Gottfredson, 1986). This theoretical variance within units is not sufficient to explain the null findings (in essence the whole is a function of its parts). One would need to bolster the argument for null findings

[1]Note that although an unbiased relationship is recoverable from the aggregate level data, the nature of the relationship is not! That is, as I have specified the model it wouldn't be considered a neighborhood/contextual effects model, as social disorganization is a continuous gradient (there are no discrete neighborhoods). The aggregate level data would not be able to discern whether the relationship was a neighborhood effect, a spatial effect, or a local effect.

based on other factors (reasonable ones in the examples given in Taylor and Gottfredson (1986) could be measurement error, under-powered study, and spatial spill-over effects).

3.1.2 Between unit heterogeneity

Another possible interpretation of heterogeneity is between units (Groff et al., 2010; Lum, 2008; Smith et al., 2000). While this argument is encountered less frequently, it can be related to the topic of stationarity. Stationarity for temporal processes is frequently defined as the time series having a constant mean and variance, regardless of the location of the observations within the series. A series can be empirically demonstrated to be weakly stationary by examining the covariance between any two points in time (Enders, 2010). The same is true for spatial processes, and the stationarity of a spatial process is defined in terms of the covariances between any two observations in the spatial distribution (Cressie, 1993). When one is talking about between unit heterogeneity, one is making a statement about the *relationships* between units (i.e. the autocorrelation), and saying that the *process* is not constant across the study space.

While many of the prior works cited have noted interesting distributions of crime, in particular that some streets may have large volumes of crime while nearby streets have little (Groff et al., 2010), this between unit heterogeneity is not a complete argument for using smaller units of analysis (nor does it inherently establish that the spatial distribution of crime is non-stationary). Take for instance the statement by Groff et al. (2010, p. 23), "temporal crime trajectory pattern membership often varies from street segment to street segment", which does not justify the conclusion "the processes which create crime free streets are most likely very different from those which result in streets with significant amounts of crime." in the following paragraph. The fact that the distribution of consistently high crime streets co-locating next to no crime streets does not in and of itself say anything about the *process* through which various exogenous phenomenon cause crime as those phenomenon undoubtedly vary between streets.

For an example, using the typical temporal trends found in the previous work of Weisburd and colleagues for streets in Seattle (Groff et al., 2010; Weisburd et al., 2004, 2009), that the sole cause of a street being in a high crime trajectory is that it has an elevated amount of foot traffic (such as due to commercial shops or bus stops that attract walkers to that area). Nearby streets may have practically no foot traffic, and hence they have practically no crime (and are encompassed by one of the other trajectory groups). This would not inherently suggest the relationship between foot traffic and crime varies in space (e.g. foot traffic has a larger effect in some neighborhoods than others). In this example, the co-location of streets with a low crime trajectory and a high crime trajectory say nothing about the causal process. The process (or the causal effect some variable has on crime) may still be constant throughout space, and hence the estimate of the aggregate relationship will be the same as the disaggregate one given the same constraints listed prior.

In equations, for simplicity just imagine there are two locations in a study, place 1 and place 2, and that crime at each place is a function of foot traffic.

$$\text{Crime}_1 = \beta_1(\text{Foot Traffic}_1) \tag{3.1}$$

$$\text{Crime}_2 = \beta_2(\text{Foot Traffic}_2) \tag{3.2}$$

If one aggregates the two locations and $\beta_1 \neq \beta_2$ then the aggregate estimate of foot traffic on crime will not be equal to the micro level equations, but will be some mixture of them.

Articles within the criminological field that have attempted to take into account spatially varying parameter estimates are papers using geographically weighted regression or spatially varying coefficient models (Britt et al., 2005; Messner et al., 2007; Cahill and Mulligan, 2007; Waller et al., 2007; Graif and Sampson, 2009; Wheeler and Waller, 2009; Zhu et al., 2006). Another example is a set of recent articles by Lum (2008, 2011) that have used a exploratory approach to analyze the spatial distance between drug and violent crimes in different census tracts, repeating the same analysis in each census tract to see if any differences occur across the study space.

Another way work in criminology has addressed this question is through examining interaction effects between variables (Smith et al., 2000; Taniguchi and Salvatore, 2012). This is saying that the causal effect between some variable and crime is conditional on the value of a third variable. Using the same street crime trajectory example, one may have a theory that foot traffic has a stronger effect on crime within areas that have greater amounts of physical decay (e.g. broken windows theory). As long as the conditioning variable (physical decay) varies across the study space, the effect of foot traffic on crime would vary as well.

$$\text{Crime}_1 = \beta_1(\text{Foot Traffic}_1) + \gamma(\text{Phys. Decay}_1) + \delta(\text{Foot Traffic}_1 \cdot \text{Phys. Decay}_1) \quad (3.3)$$

$$\text{Crime}_2 = \beta_1(\text{Foot Traffic}_2) + \gamma(\text{Phys. Decay}_2) + \delta(\text{Foot Traffic}_2 \cdot \text{Phys. Decay}_2) \quad (3.4)$$

In this example if $\text{Phys. Decay}_1 = 0$, then the γ and δ terms for Crime_1 are zero. If $\text{Phys. Decay}_2 = 1$, then the equation can be rewritten as:

$$\text{Crime}_1 = \beta_1(\text{Foot Traffic}_1) \quad (3.5)$$

$$\text{Crime}_2 = (\beta_1 + \delta)(\text{Foot Traffic}_2) + \gamma \quad (3.6)$$

So this shows how the additional variable Phys. Decay can allow the effect of foot traffic to vary between the spatial units. The relationship between between unit heterogeneity and aggregation is more difficult to conceptualize than within unit heterogeneity. If between unit heterogeneity exists on a specific scale, aggregation will undoubtedly diminish the ability to identify those varying parameters. But, the fact that some phenomenon varies at smaller units of analysis in some unexpected ways does not demonstrate between unit (or treatment) heterogeneity in causal effects. It also makes little sense to argue that one should use smaller units of analysis to identify between unit heterogeneity if one does not estimate models that identify the between unit heterogeneity.

3.2 Arguments for Larger Units

Most of the contemporary literature pertinent to this discussion encourages using smaller units of analysis (Weisburd et al. (2009) is an entire book essentially devoted to that sentiment). But besides using aggregate level data out of necessity or convenience, occasionally one will come across arguments for the use of larger units of analysis, and frequently these hinge on whether or not the independent variable of interest is reliably measured at that unit (Oberwittler and Wikstrom, 2009; Raudenbush and Sampson, 1999; Sampson and Raudenbush, 1999; Savitz and Raudenbush, 2009) or whether the construct is only applicable to larger units (Chamlin and Cochran, 1995). What follows are two reasons why this sentiment is likely misguided. One, reliability of neighborhood constructs is not an attribute of the neighborhood size, but of the number of samples within the neighborhood. Two, the aggregate process is always observable with the disaggregated data.

3.2.1 Reliability of perceived neighborhood characteristics

Robert Sampson and Stephen Raudenbush introduced the concept of ecometrics as a descriptor for the quantitative measurement of neighborhood processes in social ecology (Raudenbush and Sampson, 1999; Sampson and Raudenbush, 1999). This suggestion was largely in response to the mass of data that had been collected for the project on human development in Chicago neighborhoods (PHDCN). The PHDCN project had collected a massive amount of data from multi-wave panel survey and videotapes of streets. Within this work, Sampson and Raudenbush both noted the need for accurate measures of neighborhood social processes was a key part of the project, as well as assessing the validity of those metrics (in particular the discriminant and convergent validity of those indicators).

For illustration, assume neighborhoods have an underlying construct of collective efficacy (Sampson et al., 1997), and we attempt to measure this construct through a resident survey. It is likely that a single resident perception does not accurately capture the social milieu of the entire neighborhood (either through what we typically define as measurement error, or that the perceptions of one individual are inherently not reflective of the range of potential perceptions in a neighborhood)[2]. So one needs to ask multiple respondents their perception of the neighborhood to be able to reliably measure the true latent construct of the neighborhood affecting the resident perceptions. The more respondents are surveyed within the neighborhood, the more accurately the underlying latent construct should be captured.

Accounting for measurement error in indicators is important when one is trying to estimate causal effects in a multiple regression setting. For bivariate associations, measurement error in either of the variables will attenuate the correlation between the two variables (and for a simple linear regression will make the regression slope flatter (Kutner et al., 2004)). For multiple regression, if certain independent variables are mis-measured this can lead to spurious relationships between other variables that would be zero if the mis-measured variable(s) were properly accounted for in the model.

While measurement error in the independent variables creates a similar situation to that

[2]This brings up further questions about what exactly is a neighborhood process, or what are latent variables at the neighborhood level. For instance in the example given, the PHDCN data collected neighborhood perceptions through survey questions constructed as Likert items (e.g. a five point scale asking if an action was very likely, likely, neither likely or unlikely, etc.) Considering different individuals can have different conceptions of what likely and unlikely are, everyone could answer the survey perfectly as they intended (e.g. if someone retested the respondents, and the respondents answered the survey questions with the exact same responses every time). But this does not mean the respondents perception of the neighborhood is inherently reflective of all individuals within that same neighborhood. This is even *assuming* a neighborhood process exists, which it isn't obvious how one can empirically distinguish between a neighborhood process and one that has a continuous gradient throughout the study space.

of omitted variables in that it results in biased regression coefficients, this doesn't inherently justify using larger units of analysis (nor does a high ecological reliability inherently suggest that the causal process operates at that particular support). One can see this if you evaluate the metric Sampson and Raudenbush developed to assess the reliability of ecological constructs (Sampson and Raudenbush, 1999, p. 646). The internal reliability of a measured neighborhood process will always increase with more measurements within a neighborhood. Hence, the metric itself can not identify if a neighborhood exists, and is susceptible to all the same critiques of aggregation bias that were presented for other units of analysis.

Sampson and Raudenbush note that the units of analysis used to represent neighborhoods in their studies (neighborhood clusters of census tracts) were chosen *A priori* based on homogeneity of demographic characteristics (Sampson and Raudenbush, 1999). They assume the neighborhoods exist, and then measure their theoretical influences on various human behavior. There is nothing to suggest that what they detail as neighborhood effects could not be operating at much smaller levels, or that neighborhoods in reality could be organized into different spatial units. Using the results of the studies on the PHDCN data to justify larger units of analysis (or to even justify the fact that neighborhood processes exist) is tautological. Larger units of analysis can only be justified based on *practical limitations* of reliably measuring constructs within those units. One could reliably measure characteristics at smaller units if one took enough samples.

Oberwittler and Wikstrom (2009) also suggests that the trade-off in the number of units with the power to identify relationships between constructs (i.e. the power goes up because of more units of analysis at lower levels) is worth using smaller units of analysis. One should only choose larger units of analysis if practical constraints prevent them from examining smaller units. While in their infancy, current statistical techniques such as small area estimation (Kang et al., 2009; Savitz and Raudenbush, 2009) and techniques estimating attributes at different spatial supports may help alleviate such needs in the future (Foster et al., 2012; Young et al., 2009).

The other argument for using aggregate constructs I presented was that the construct is only applicable to larger units of analysis (Chamlin and Cochran, 1995). What follows is why this is a poor justification to use aggregate level units, as the aggregate level process is always observable at the disaggregated level units.

3.2.2 The neighborhood can always be seen at the individual level

Past research has justified using larger units of analysis based on either theoretical relevance (Chamlin and Cochran, 1995) or practical limitations in measurement (Raudenbush and Sampson, 1999; Tita and Ridgeway, 2007). When one is identifying relationships at the individual level, it is possible to identify neighborhood or spatial processes of interest in the exact same way one can with the aggregate data itself. Similar to how earlier aggregation bias was documented to be a case where effects were not decomposable into their unique parts, one can examine the unique parts of the model to determine if neighborhood effects exist.

For an example, say the causal models exists of the form;

$$y_i = \beta_1 x_i + \beta_2 \bar{X} + e \tag{3.7}$$

In this formula, $\bar{X}$ is the neighborhood average and so β_2 would be interpreted as a contextual effect. If one were to estimate the effect of X in the aggregate level equation, one would simply observe the effect as the sum of β_1 and β_2. Not only can one not identify the distinction between individual and contextual effects, but when using the aggregated units ones conception of neighborhood effects are imposed by the aggregated units. If one were to reduce this equation to say the neighboring observations that share the same neighborhood, one might then estimate the equation;

$$y_i = \beta_1 x_i + \beta_2 x_1 + \beta_3 x_2 + \cdots \beta_k x_k + e \tag{3.8}$$

Where $x_1 \cdots x_k$ are the neighbors of x_i. While it is infeasible to estimate all of the potential parameters (and is referred to as the incidental parameter problem (Anselin, 1988)), one at least has an option to explore different potential specifications at the smaller unit of analysis. Realistic examples in criminology might be the bars on the adjacent street or the number of criminals living in a certain proximity of the location. If one were using aggregated units (i.e. they could only observe $\bar{Y}$ and $\bar{X}$) one would essentially be stuck with the equation;

$$\bar{Y} = \beta \cdot \bar{X} + e \tag{3.9}$$

The main limitation of this aggregate equation is that many different processes at the disaggregated level could result in the aggregate equation. Hence it makes it difficult to make any sort of rigorous theoretical claim based on the aggregate data. To assume that the above equation is indicative of neighborhood or an individual level process, one has to make a series of unprovable (and likely unjustified) assumptions.

One potential advantage of the aggregate model over the disaggregated model is that only one parameter needs to be estimated. The only obvious benefit of this is that the statistical power of the single β estimate is larger than many separate β_k estimates. In practice, the argument for increased power may be mitigated by the larger sample size when using smaller units of analysis (Oberwittler and Wikstrom, 2009), so one may still prefer to use the smaller units even in light of this. Regardless, the aggregate form of the data can not justify the pooling of the separate effects (i.e. one can not empirically ascertain if aggregation bias is occurring when only observing the aggregated units), and so one should evaluate the micro level model before one decides if the aggregate model is appropriate.

One does have limits to the potential effects they can explore between a unit and its neighbors (even using the disaggregated units), but through appropriate exploratory data analysis one would hope to identify the correct functional form of the posited relationship(s) between variables (if those relationships do exist). The ultimate nature of the functional form to be assumed can be much more flexible with the small units, although can be potentially restricted to be equivalent to the aggregated model if the data permit. This is synonymous with building a general statistical model and trimming it to be more parsimonious. If one starts with the specific model, one can not identify if a more general model is appropriate. With the disaggregated equation one can always build up (such as by specifying a multi-level model), but the obverse (getting the disaggregated relationship from the aggregate data) is not possible.

3.3 My Arguments for Smaller

Given the above criticisms of past work justifying smaller or larger units of analysis, what follows is a presentation of a more complete argument for the use of a specific unit of analysis. One should take the lessons presented here in choosing a suitable unit of analysis given the particular questions of the study design. To follow I present some examples where smaller units would allow one to examine more theoretically interesting causal relationships. This argument will ultimately not discriminate between many different potential smaller units of analysis (e.g. parcel addresses, census blocks, an arbitrary lattice). Hence the chapter will conclude with an argument for the use of street segments and intersections based on theoretical and practical relevance of streets as a unit of analysis, as well as practical constraints.

3.3.1 Micro level relationships are more interesting

Imagine a scenario in which the prevalence of crime (Y_i) in a certain space is a function of several attributes of that space;

$$Y_i = \beta_1 \text{po}_1 + \beta_2 \text{po}_2 + \beta_3 \text{bars}_1 + \beta_4 \text{bars}_2 + e \tag{3.10}$$

Say that po_1 is the number of potential offenders who live on that street, po_2 is the number of potential offenders that live on neighboring streets, bars_1 is the number of bars on that street, bars_2 is the number of bars on physically adjacent streets. Let's also say that all of the preceding variables have a positive correlation with one another (bars self selecting into neighborhoods with a higher number of deviants would produce such a correlation (Gruenewald, 2007; Treno et al., 2008)). This would then imply an aggregate level equation;

$$\bar{Y} = \beta_5 \bar{\text{po}} + \beta_6 \bar{\text{bar}} + \mu \tag{3.11}$$

If one were working with the aggregated data, β_1 and β_2 would be inherently confounded, and would be both absorbed by β_5 in the aggregate level equation. The same is true for β_3 and β_4. A neighborhood model might imply that β_1 and β_2 are equal. A micro level model might imply β_2 is zero or very small compared to β_1. If one starts with the disaggregated equation, one can check these assumptions to see if they are true.

For the effects related to bars, the difference between β_3 and β_4 would have both theoretical and practical implications. If β_4 was positive and non-negligible, this might suggest limiting licensing of bars within close proximity to crime hot spots may be a reasonable strategy to reduce crime (even if the bar is not located at the exact location of a hot spot). This also has pertinence to the theory of how bars would impact crime (it would suggest a diffusion effect into neighboring areas). Imagine also if one found an interaction effect of bars on the local street and bars on neighboring streets. This would suggest that licensing of bars nearby other bars should be limited to prevent increases in crime. These types of interesting relationships between micro level and neighborhood process essentially extend to *every* general theory of crime with place based relevance. One can not use the neighborhood level data to justify the existence of a neighborhood level process!

In the same example above, if either bars or potential offenders were omitted (from either the disaggregated or the aggregated equation) the other variable's effect would be upwardly biased (as I initially specified that all of the variables had a positive correlation with one another). Hence the disaggregated equation does not necessarily prevent biased estimates, but allows one to potentially understand the causal processes that influence crime more accurately. In the case that crime has some type of endogenous spatial effect (i.e. when crime on one street causes crime on another street (Johnson et al., 2009; Mohler et al., 2011; Short et al., 2009)) this information is always lost in the aggregation process. In the cases these endogenous spatial effects are also correlated with other independent variables, the effect of the other independent variables on crime will never be identifiable. Given that such diffusion processes are realistic, one should choose a small unit of analysis in which such effects could be identified.

3.3.2 Why street segments?

Ultimately, the above arguments do not justify using street segments and intersections over other small units of analysis. One has to choose some unit though, and here I will argue for the use of street blocks (i.e. street segments) and street intersections. I will justify the use of these units for four reasons; 1) theoretical relevance as a contextual unit (Taylor, 1997), 2) consistency with other contemporary research (Groff et al., 2010; Weisburd et al., 2004, 2009), 3) practical relevance for crime prevention (Ratcliffe et al., 2011; Sherman and Weisburd, 1995; Weisburd et al., 2008), and 4) practicality of measurement and model estimation.

The strongest proponent of street segments as interesting and relevant contextual units is Ralph Taylor. In his work, he argues that street blocks create physical areas in which neighbors interact on a daily basis, and create zones in which behavioral settings are defined (Taylor, 1997). The street block (unlike entire neighborhoods) creates a specific visual arena, in which potential offenders can be seen by vigilant neighbors and potential offenders can perceive different territorial boundaries (Newman, 1972). The social process that influences behavior outside of the immediate visual perception of an offender necessarily operate through distinct processes. Take for example the presence of visible decay. Broken windows theory would suggest that the visible presence of physical decay (such as actual broken windows) gives implicit cues that deviant behavior is acceptable in particular settings. If broken windows have an effect on crime outside of the immediate visual perception of that physical decay, it suggests the theory is incomplete as stated above. It may be that individuals attach signs of physical decay in adjacent areas and infer those same behavioral cues for different (but nearby) areas, or it could be that there are other neighborhood processes afoot that explain the neighboring effects.

Consistency with other contemporary research is another reason that street segments will be the unit of analysis chosen in this study. The work of David Weisburd and colleagues (Groff et al., 2010; Weisburd et al., 2004, 2009) on identifying the temporal trajectories of street segments and crime being one of the strongest influences. In particular, the identification of high crime and low crime trajectory streets, and the fact that these streets can be located next to each other in space (Groff et al., 2010) suggests that interesting patterns could emerge within neighborhoods that have causal influences on crime.

The causes of crime at the street segment level have practicable implications for policing crime. Hot spots policing has been demonstrated to be one of the most effective crime reduction techniques currently implemented within police departments (Braga, 2001; Lum et al., 2010; Sherman and Weisburd, 1995). These targeted interventions are frequently implemented at very small areas (or specific locations) in which a large number of crimes occur. Police interventions for larger units are not as successful as those at smaller places (Weisburd et al., 2008; Lum et al., 2010).

Using street segments at the onset of the project makes some of the data analysis more practical. When using smaller units of analysis (such as specific addresses) ultimately many of the units will have zero occurrences of crimes at those locations. This point mass of zero crimes creates difficulties when trying to create statistical models that is partially evaded when using street segments. Also using fewer street segments somewhat prevents improper geocoding of street addresses. The smaller number of units also makes certain tasks, such as estimating characteristics at the street level that are not disseminated at that support, slightly less arduous due to the smaller number of units. Although within this dataset, Washington D.C., still contains nearly 25,000 street segments, allowing great flexibility to examine multivariate and spatial relationships.

The main reason for using street segments and intersections with this particular research project is that the public crime data that will be used (from dc.gov) is already aggregated to the street segment and intersection. I could (technically) utilize smaller units of analysis by utilizing crime data I have at the address level, either through the Finn Institute, e.g. (Wheeler, 2012). But D.C. (at least at the time of beginning this prospectus) had one the most varied sets of potential covariates related to the built environment. If I wanted to use the D.C. crime data, and all the aspects of the built environment that come along with it, street segments and intersections is as small as it gets.

As it is becoming more common for police departments to disseminate geocoded crime data, a frequent unit of analysis it is disseminated at is the street segment. A few examples (besides D.C.) are Houston, Philadelphia, Baltimore, Chicago, Omaha, and all of England[3]. If criminologists want to use this wealth of public data, they will need to learn to live this level of obfuscation. As was related previously in the chapter on aggregation bias, not being able to differentiate among spatial effects within the street segment is a small consolation to pay, and there are still plenty of interesting causal questions that can be evaluated at the street segment level.

3.4 Conclusion

The particular unit of analysis should be dictated by the question a researcher is interested in investigating. The reductionist argument to examine small units of analysis is thus an extension of arguing that the distinction between local, spatial and neighborhood effects are theoretically interesting. Of note, for those examining neighborhood effects, one needs data that is not aggregated to the neighborhood. Also to rule out other data generating process (in particular spatial ones) one needs more refined spatial labels for the data observations than are typically recorded (or disseminated) in neighborhood level data collections.

This leaves the applied researcher with no specific guidance on how small is reasonably small. I provide some arguments here why street segments are a useful end, although different research designs may call for different units. The hope is here to bring mathematical

[3]URL's for where each PD is disseminating the data (as of 3/29/2012) are Houston (`http://www.houstontx.gov/police/cs/stats2.htm`), Philadelphia (`http://citymaps.phila.gov/CrimeMap/`), Baltimore (`https://data.baltimorecity.gov/`), and Chicago (`https://data.cityofchicago.org/`). Baltimore and Chicago links point to a more general data portal built by an independent firm named Socrata. Omaha publishes geocoded crime data to the online web mapping service, `crimemapping.com` published by a independent firm named the Omega Group. At that same site there are many other police jurisdictions publishing data. The website, `police.uk` is the centralized location where online crime maps for all of England can be found.

discussions on how aggregation bias occurs to guide one to an appropriate unit of analysis, as opposed to relying on incomplete arguments. The mathematical arguments also provide a frame of reference for how we should interpret data that are already aggregated.

Chapter 4

Place Based Theories of Crime

This next chapter in the dissertation attempts to organize theoretical reasons given in the literature as to what causes crime at micro places in the urban environment. The potential breadth of theories for such a review is vast. Many theories could be potentially concocted to be a spatial theory of crime (in that the theory explains spatial variations of crime in addition to individual person motivations).

To make the task more manageable I have organized the review into three dominant themes for the causes of crime *at places*. These themes are; 1) the characteristics of people who live in certain places, 2) the characteristics of places themselves (e.g. the built environment) that influence behavior, and 3) the number of *potential* crime incidents at any particular place. I do this for two main reasons. One is that frequent popular theories of crime have such large theoretical overlaps that it makes them difficult to distinguish. Another reason is that this perspective makes it very clear *how we measure* these characteristics at places.

These general themes will largely subsume several widely accepted (more general) theories of crime, and so each section will review those popular general theories of crime. I conclude the chapter with reasoning why I think it is important to understand and represent the theories in empirical models as clearly as possible for future work.

4.1 People at Places that Cause Crime

Criminologists have been predominately interested in explaining geographic patterns of crime through how demographic and economic characteristics of *people that live in certain places* explain crime. That is to say, certain people have characteristics that pre-dispose them to commit crime at a higher frequency than other types of people. Another mechanism through which the characteristics of people that live at certain places cause crime is that certain people (or groups of people) are better at controlling deviant behavior than other people, and the distribution of those people (or groups) impacts the locations of crime. This is otherwise known as informal social control.

What makes these theories explain crime at places is that the aforementioned people (or groups of people) in turn are more likely to live in certain urban areas than others, hence this differential in the *types of people* that live in places can at least partially explain the geographic distribution of crime[1].

This historical tradition of using such theories to explain crime at places can be mainly traced back to social disorganization theory, and so I devote a section to an explicit review of the theory[2]. I then talk about a current extension of social disorganization theory, collective

[1]It is frequently clichè to state that most crime occurs nearby the home to extend how where certain types of people live explains *where people commit crime*. This is grossly over-simplifying the journey to crime literature, and does not justify such conclusions.

[2]Or at least social disorganization theory is the oldest theory that is still contemporary

efficacy.

4.1.1 Social disorganization theory

I start the review of place based theories of criminal behavior with a review of social disorganization theory. I do this for two reasons. One, of the theories I will talk about it has the oldest lineage. Two, some subsequent theories I will talk about, collective efficacy in particular, could arguably be extensions or elaborations of the original social disorganization theory. It will also have aspects directly pertinent to all of the other theories discussed, and to a certain extent subsequent theories had to discriminate themselves from social disorganization to be seen as unique. The popularity of social disorganization is likely due to its consistency as a strong predictor for crime at places, both for large aggregated areas (Osgood and Chambers, 2000; Pratt and Cullen, 2005) and for small areas (Lum, 2011; Shaw and McKay, 1969)).

Social disorganization theory originated in the work of Shaw and McKay (1969). Figure 1 provides a graphical representation of the theory, as it pertains to the causal antecedents of delinquency, through a path diagram.

The main thing to note in Figure 1 are first the structural variables on the left of residential turnover, economic disadvantage, and ethnic heterogeneity. These then cause social disorganization, and then social disorganization causes crime. The concept of social disorganization is important to criminological theory as it identifies the presence of *informal social controls* that can affect deviant behavior of individuals within a community. Specifically, they focus on how individuals within communities can limit deviant behavior, and

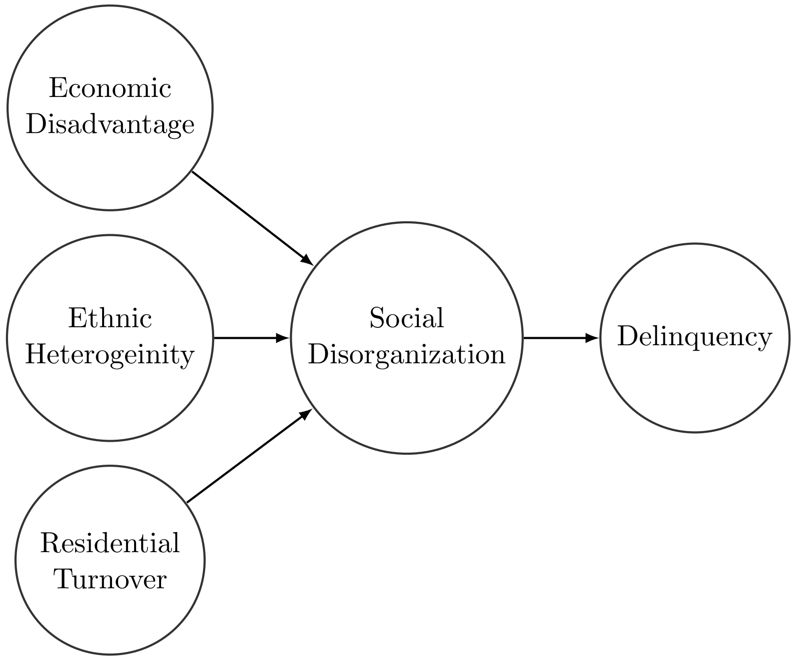


Figure 1: Path diagram depicting theoretical processes for social disorganization theory.

how socially disorganized places have a limited capacity to prevent deviant behavior. This is opposed to formal social controls, like the criminal justice system (e.g. police, courts, probation). The idea of informal social control is pervasive throughout subsequent criminological theories presented in this chapter. Shaw and McKay never identify direct measures of social disorganization in their research, and only use the concept of social disorganization to theoretically link the structural covariates of residential turnover, economic disadvantage, and ethnic heterogeneity to crime and delinquency.

Notably absent from the diagram in Figure 1 is any mention of the concentric zonal rings, perhaps the most well-known part of their theory. It is notable because it is the notion that makes it a spatial theory, but this omission is intentional. The idea of concentric zonal rings defining different ecological areas of the city is a description of the ecological growth and organization of the city. This organization of the city then in turn affected *who lived where*, and it is the characteristics of the people in particular places that had effects on delinquency (and mental illness, infant mortality, and cases of tuberculosis). They even mention that industrial land use, which is the main identifying factor for the first zone, is not directly related to delinquency. Only through its effects on residential turnover is it correlated with crime (Shaw and McKay, 1969, p. 143);

> The location of major industrial and commercial developments, the distribution of buildings condemned for demolition or repair, and the percentage increase or decrease in population by square mile areas were presented in chapter ii as indications of the physical differentiation of areas within the city. Quantitative measures of the first two are not available, but inspection of the distribution maps shows clearly that the highest rates of delinquents are most frequently found in, or adjacent to, areas of heavy industry and commerce. These same neighborhoods have the largest number of condemned buildings. The only notable exception to this generalization, for Chicago, appears in some of the areas south of the central business district.
>
> There is, of course, little reason to postulate a direct relationship between living in proximity to industrial developments and becoming delinquent. While railroads and industrial properties may offer a field for delinquent behavior, they can hardly be regarded as a cause of such activities. Industrial invasion and physical deterioration do, however, make an area less desirable for residential purposes. As a consequence, in time there is found a movement from this area of those people able to afford more attractive surroundings. Further, the decrease in the number of buildings available for residential purposes leads to a decrease in the population of the area.

There have been many measures of social disorganization, but part of the popularity of the theory has been the ability to use readily available measures from the census to approximate the structural covariates that cause disorder. Shaw and McKay (1969) used measures such as median rental price, percentage of home ownership, and percentage of families on relief to represent economic disadvantage. They use changes in population between two time periods to measure residential turnover, and percentage of foreign born to represent ethnic heterogeneity. More contemporary literature has used some different measures, such as measuring residential turnover by US census items such as the proportion of residents that have moved in the past 5 years, or for economic disadvantage the number of unemployed or below a census defined poverty line (Hipp, 2007; Land et al., 1990). Many measures of ethnic heterogeneity have since been developed (Massey and Denton, 1988).

Other *direct* measures of social disorganization are measures of family disorder (e.g. the proportion of female headed households, proportion married) and measures used to proxy the strength of social networks (Hipp, 2007; Kubrin and Weitzer, 2003; Pratt and Cullen, 2005). In the original conception of social disorganization, these variables would be considered effects of the structural variables already listed. The subsequent section on collective efficacy will examine these measures in more detail, as it is more contemporary literature that has attempted to directly measure social disorganization.

Another frequent measure associated with social disorganization is non-residential land use. Although, as stated previously, Shaw and McKay only stated a link between *industrial* land use and crime theoretically was formed through residential turnover. Thus land use may be considered more similar to the other structural covariates listed in 1, in that it is causally before social disorganization (and theoretically causally before residential turnover)[3].

[3]More contemporary work has further identified the correlation between land use and crime, and in particular has identified how *certain types* of non-residential land use can actually directly cause crime (Kinney et al., 2008; Stucky and Ottensman, 2009; Wilcox et al., 2004). Such correlations will be brought up in this chapter in a review of crime pattern theory.

4.1.2 Collective efficacy

Referring back to figure 1, Shaw and McKay linked the structural characteristics of residential turnover, economic disadvantage, and ethnic heterogeneity to crime by causing what they called social disorganization. They fully admitted in their work that they did not have a direct measure of social disorganization obtainable from any observable data source, but made the link on epistematic logic. Contemporary research has identified a construct that is the obverse of social disorganization, collective efficacy.

Robert Sampson, and mainly his work on the PHDCN have advanced the construct of collective efficacy. Sampson et al. (1997) defines collective efficacy as 'social cohesion among neighbors combined with their willingness to intervene on behalf of the common good'. This construct would be considered the ability of neighborhood social groups to exert informal social control over its constituents. This is the opposite of social disorganization, as socially disorganized areas were considered unable to exert informal social control to prevent deviant behavior. This is a relevant spatial theory of crime as one's neighbors have a direct impact on an individual's willingness to act on behalf of that group. It is quite natural to define group here as spatially proximate neighbors (often nested within a series of mutually exclusive areas), although it is certainly possible to define it through other means (see Hipp et al. (2011) for an alternate example using social network statistics).

Measures of collective efficacy are most frequently taken from resident surveys. For example, Sampson et al. (1997) defined collective efficacy as the combination of two pre-determined likert scales relating to informal social control (e.g. would you intervene if a child was skipping school) and social cohesion (e.g. people in this neighborhood can be trusted).

Bellair and Browning (2010) utilize similar survey measures asking about participation in various social groups to directly measure social cohesion.

4.2 Places that Cause Crime

Another mechanism, besides the demographic characteristics of people at places that may cause crime are *the characteristics of the places themselves*, which may predispose particular individuals to commit crime when those individuals are exposed to the conditions at a particular place. To use an example, imagine an individual leaves a window on their car door down and they have a portable GPS device on the front dashboard. An individual may take advantage of the opportunity to steal the GPS device because the window is down and the individual can simply reach in and grab it. The car window being down isn't a description of people, but of the place, and similar examples can be extended to other (more permanent) places than car windows.

This section details two popular theories of crime that involve how the characteristics of particular places influence behavior of individuals exposed to those places. Those two theories are broken windows theory and defensible space as well as crime prevention through environmental design.

4.2.1 Broken windows

Broken windows theory has been a highly cited theory of criminal behavior, traced back to Wilson and Kelling (1982) who based their interpretations on an experiment conducted by

Zimbardo (1973))[4]. Visible signs of physical disorder give off cues to individuals that "no one cares", and encourage deviant behavior. These visual cues then lead to minor incivilities, like vandalism, which then escalate to more serious deviant behavior. Figure 2 displays the path diagram representing the theoretical causal processes in Broken Windows Theory[5].

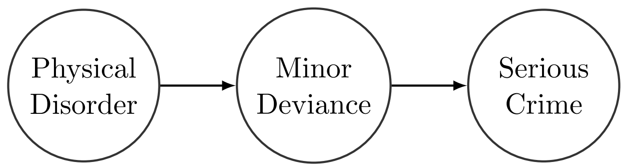


Figure 2: Path diagram depicting theoretical processes for broken windows theory.

Motivated by seeing many cars vandalized and stripped of parts in the neighborhoods of New York, Zimbardo developed an experiment to investigate a hypothesis that once the car showed visual cues of abandonment, individuals would steal from the car. Zimbardo conducted his experiment in two locations, across the street from the NYU campus in the Bronx, and in Palo Alto near the Stanford campus. In each location Zimbardo placed a car with the hood up and license plate removed to provide visual signals of abandonment. In the Bronx neighborhood, it only took 10 minutes before an individual stole parts from the vehicle. In three days the vehicle was completely stripped of all valuable parts and destroyed. The car in Palo Alto on the other hand was left untouched, besides someone putting the hood down during a rainstorm! Not until Zimbardo got a few of his graduate students to start

[4]There are other scholars who had similar sentiments about the link between physical urban decay and deviance before Wilson and Kelling (1982), in particular Jane Jacobs (Harcourt and Ludwig, 2006), although the popularity of the contemporary theory is undoubtedly due to Wilson and Kelling (1982). Shaw and McKay (1969) also mentioned some of the same processes of physical decay in relation to residential turnover, in particular how industrial locations tended to have a negative impact on the upkeep of nearby residential homes (Taylor et al., 1995).

[5]Ralph Taylor in his book, *Breaking Away from Broken Windows*, has an entire chapter devoted to what he calls the "Incivilities Thesis", although it is equivalent to what I am calling broken windows theory (Taylor, 2001). Taylor goes into much further detail than I do here, and has similar path diagrams that highlight more specific processes, such as including fear of crime.. The minimal path diagram in 2 is not misleading though, and emphasizes the main points of the theory, both by Wilson and Kelling (1982) and subsequent scholars.

to damage the vehicle with a sledge hammer did other individuals partake in the deviant behavior of damaging the car.

Another experiment that replicated similar findings was conducted by Keizer et al. (2008). In the series of experiments, Keizer et al. (2008) demonstrate that when one deviant behavior is witnessed, individuals are more likely to participate in that same deviant behavior. While many of the examples would be considered minor deviant acts, such as littering, it is further evidence that individual behavior is influenced by visual cues of disorder and deviance.

These experiments illustrate the process by which visual stimuli of disorder influence human behavior. They are also enlightening as to the potential greater complexity through which individual behavior is influenced by contextual factors. For Zimbardo's experiment, while the same stimulus was presented in New York as in Palo Alto, the behavior that followed was different. This suggests that not only the visual stimulus of physical decay is necessary to generate deviant behavior, but greater contextual effects are necessary for "broken windows" to encourage deviant behavior. Zimbardo attributed the difference in behavior to anonymity afforded by living in New York City as opposed to Palo Alto. There are many other differences between New York City and Palo Alto, and it is plausible to attribute the difference in behavior to many other factors (either in differences of the place or the people). Although given other experiments (Keizer et al., 2008) it may be reasonable to expect the deviant effects of witnessing physical disorder are generalizable to other situations and places.

Although the causal mechanisms explicated by broken windows theory are fairly simple, there is a fair amount of debate around whether the theory has been empirically verified (Cerd et al., 2009; Gau and Pratt, 2008; Harcourt and Ludwig, 2006; Kelling and Sousa, 2001; Rosenfeld et al., 2007; Sampson and Raudenbush, 2004), at least in observational contexts outside of some of the smaller experimental examples previously mentioned (Keizer et al., 2008; Zimbardo, 1973). The theory is so strongly tied to order maintenance policing initiatives that evidence for or against such initiatives is taken to be direct evidence for or against broken windows theory itself (Kelling and Sousa, 2001; Rosenfeld et al., 2007). It is probably simpler to consider the theory *in absentia* of the order maintenance policing policies with which it is associated, as evaluations of such policies in many observational settings are fairly weak research designs (Bushway and McDowall, 2006; Berk, 2005), and whether or not order maintenance policing was in part responsible for the great crime decline in the 1990's is still under debate (Eck and Maguire, 2000; Kelling and Sousa, 2001; Levitt, 2004; Rosenfeld et al., 2005)[6].

Other direct aspects of the the broken windows theory have come under attack besides the efficacy of order maintenance policing. In particular, the connection between perceived disorder and subsequent behavior has been questioned, and whether perceived disorder can be reasonably differentiated from other perceptions of criminal behavior (Gau and Pratt, 2008) or general prejudices about certain people or places (Sampson and Raudenbush, 2004). Other research findings of a direct correlation between physical disorder and crime are mixed, with some having moderate evidence of a correlation (Cerd et al., 2009; Skogan, 1990; Taylor, 2001) to others having little to no evidence of correlation between the two (Harcourt and Ludwig, 2006; Sampson and Raudenbush, 1999)[7].

[6] Also the NYPD's order maintenance policing initiatives have been criticized as directing police attention to minority populations, as opposed to actual areas of disorder (Fagan and Davies, 2000; Rosenfeld et al., 2007). To the extent this is true, it would preclude one from evaluating such policing initiatives as direct evidence for the broken windows theory, regardless of whether they were effective or not.

[7] For a different interpretation of Sampson and Raudenbush (1999) in favor of the broken windows theory see Gault and Silver (2008).

The measurement of broken windows theory typically involves identifying two distinct sources, sources of physical disorder and sources of social disorder. One is again referred to the path diagram in figure 2. Physical disorder usually pertains to measures directly related to the built environment. Skogan (1990) characterizes physical decay related to dilapidation and abandonment of buildings, and (Taylor, 2001) identifies buildings that are vacant as well as buildings that are vacant and boarded up. Cerd et al. (2009) also use a measure of street sidewalk condition collected by a New York City agency, and Cohen et al. (2000) use measures culled from a survey on physical school conditions which include items pertaining to physical disorder such as broken floor tiles. All of those examples previously cited should be considered indicators of disorder that change more slowly. Other measures of physical disorder include vandalism, unkempt lawns, and graffiti or litter. These measures of physical disorder are less stable in time than the other items related to physical decay of permanent structures.

Social disorder pertains to public acts that are bothersome or only considered minor crimes. Examples might be individuals pan handling for money, groups of youth hanging out, or other more directly criminal actions like selling/taking drugs or prostitution. Measures of social disorder are frequently derived through resident surveys, such as asking residents whether they view the above conditions as problematic in their neighborhood. Measures from systematic social observation are less frequently used to represent social disorder, and pertinent measures are difficult to relate to other data collected by administrative agencies. It is generally considered inappropriate to use measures of police arrests for such actions to prevent confounding police enforcement practices with actual different criminal behavior between places. Also reported crime (or calls for service) is likely a gross undercount for such victimless crimes.

Because measures for broken windows theory are typically done via a special data collection, as opposed to using administratively collected data, the units of analysis for which they are collected are typically small areas, such as street segments[8], although some articles aggregate up the disorder measures to larger units (Wilcox et al., 2004), or at least suggest it is reasonable to do so (Raudenbush and Sampson, 1999).

4.2.2 Defensible space and CPTED

Some of the previous theories discussed, such as social disorganization, collective efficacy, and routine activities, focus on broad demographic and structural characteristics of *the people* who live in particular places, and how those structural characteristics in turn affect crime. The two particular theories I aim to describe here, defensible space and crime prevention through environmental design (CPTED) have motivations based on how aspects of the built environment, not aspects of people, affect crime. CPTED focuses on how the built environment shapes behavior, and can prevent or promote criminal acts. Thus how the built environment shapes behavior is distinct from demographic characteristics of the residential population, and structural characteristics may be considered distal impacts for crime patterns at any particular place (Tilley, 1997). Defensible space and CPTED do not focus on the characteristics of people that are in any particular space, but on how any particular space can equally affect all of the individuals exposed to that particular space[9]. I begin my review

[8] An interesting application that uses street view imagery (available via Google Maps) to attempt to replace systematic social observation is in Clarke et al. (2010).

[9] Because my focus here is on the measurement of crime, I do not discuss the theory of situational crime prevention directly, although it has been considered by Ronald Clarke a superset of defensible space and CPTED (Clarke, 1992). This is mainly because situational crime prevention may be too broad to be only characterized as the organization or manipulation of the built environment. An example would be limiting

here with the work of Oscar Newman in his book, *Defensible Space*, and then attempt to expand on how the built environment affects criminal behavior by detailing a variety of case studies.

The theory of defensible space was first explicated by Oscar Newman in his book, *Defensible Space* (Newman, 1972). Newman focused on how places organized human interactions, and how subsequent to the organization of the built environment different patterns of behavior and use of the space could evolve. He did this through an extensive case study of public housing projects, focused in New York City, which mainly intended to examine why in particular high rise public housing buildings had elevated levels of crime. Here I will arbitrarily choose three concepts which Newman introduced in his book that have seen subsequent traction in criminological literature and are directly related to the built environment; territoriality, natural surveillance, and access control.

Territoriality

Territoriality is a visual or physical cue that either an object or a place has an owner. For offenders, visual cues provide signs that an area has an owner, and hence the individual is (or potentially will be) under the owner's surveillance (Wortley and McFarlane, 2011). Such visual cues also present barriers (albeit permeable ones) to travelling through particular areas (Brown and Altman, 1991; Rengert, 2004). In particular the concept is frequently applied to crimes with instrumental monetary motivation (e.g. burglary, robbery) (Taylor and Gottfredson, 1986), and the added barriers increase perceptions of risk. Or in the case the barrier is a physical one (such as a lock on a door), it will take further effort to traverse

the sale of spray paint to juveniles, which has no intrinsic relationship to the built environment.

and reduce the utility of committing any crime that involves crossing that barrier.

It is useful to distinguish between different types of spaces, those of public spaces, semi-public spaces, and private spaces. Private spaces are within ones one home, and public spaces are those that are accessible by everyone. Semi-public lie in-between, being proximally close and connected to a private space, but typically being visible or accessible by a wider array of individuals.

In *Defensible Space* Newman identifies architectural characteristics that help local residents identify and take more stock in semi-public spaces. For example, some smaller residential public housing spaces had local shared corridors between a few apartments, and local residents tended to decorate and keep clean those local spaces. Newman's work suggested that certain architectural designs of public housing can facilitate community integration that in turn strengthens local informal social controls. This is in contrast to high rise public housing facilities, where long rows of apartments created many more neighbors, but did not foster interaction between them.

Visual cues that a space is within the private domain of an individual are frequently obvious, but of interest here are likely cues of territoriality in public areas. Examples of territorial cues in public spaces are visual signs of ownership, such as a plaque on a side of a building or a community garden (Branas et al., 2011). Examples of intrinsically physical territorial barriers in public spaces are a fence or a row of bushes separating the sidewalk from the building. An example of architecture for delineating a boundary in a public space Newman used was the raised stoop in front of row homes. Although they are visibly public spaces, the physical marker of the stoop is an obvious sign one has crossed from the public sidewalk into the domain of one's home.

A related concept in defensible space is what Newman calls the *milieu* of a place (Reynald and Elffers, 2009). This is related to an appearance of a place, and an image of a well kept area suggests that an area has owners who control that particular area. In this sense cues of public ownership are the obverse of signs of physical decay central to broken windows theory. If a place is well kept and has physical signs of use, it has an owner, and someone cares what you do in that particular place (Wortley and McFarlane, 2011). Again, as is the case for broken windows, one can consider some aspects of this concept fairly time stable while others are potentially more variable over short time periods (Perkins et al., 1993). A fence is a permanent structure, but other visual cues, such as a well kept lawn or garbage on the street, can vary considerably in time or based on other external forces independent of local resident actions.

Natural surveillance

Another concept central to Newman's original work was that certain architectural designs could facilitate covert operations by concealing individuals and their deviant acts (Wood, 1991). Newman's best example of this was the notoriously dangerous stairwells in high rise public housing buildings. The twisting stairwells were highly concealed from public view; you could not see anything besides half a floor below you at any place in the stairwell. In turn the stairwells were effective places to sell and take drugs, and were largely avoided by local residents (even further reducing the potential surveillance of them). The concept of visibility has been subsequently used to explain why parks in urban areas have increased levels of deviant acts (Demotto and Davies, 2006; Groff and McCord, 2012) as well as theoretically suggest large and dense shrubbery may help conceal crime. But empirical evidence suggests the exact opposite is true (Kuo and Sullivan, 2001; Donovan and Prestemon, 2012), and this might be expected given that well kept shrubbery is a sign of territorial ownership and would operate to reduce crime. Another example in which the built environment can effect natural surveillance is through street lighting at night time (Farrington and Welsh, 2004).

One thing to note in the section about natural surveillance is that it seems it is more important how the offender perceives being subject to surveillance than the actual surveillance and potential intervention of witnesses (Mayhew, 1991). This is one explanation for the initial deterrent effect that decays after a few months following installation of security cameras (Ratcliffe et al., 2009).

Access control

One last central concept originally introduced by Newman in *Defensible Space* is access control. Access control is meant to signify the ability of individuals to gain access to a particular location. For instance, when in a public park one has the ability to freely roam about the grounds, whereas when one is trying to access a private park one frequently needs to go through one particular entranceway. Newman again brings this up in the context of the dangerous stairwells, and the fact that frequently the doors would be propped open and available to be entered by the general public makes it trivially easy for anyone, including deviant non-residents, to access the stairwells.

Forms of access control limit the number of places at which one can enter or exit a certain space, and with that sometimes restrict individuals who can enter a particular space and/or subject individuals to an inspection before allowing entrance. A good example related to general crime and manipulating the built environment is that of 'Alley-Gating' in British towns, which has shown to be successful in reducing burglaries (Bowers et al., 2004; Haywood et al., 2009). Alley-gating is the act of securing the alleys behind row homes, which only allow local homeowners access. As the alleys were restricted from public view, they were a popular location to break into homes. Another example is street closings, which deterred Johns from having easy access to prostitutes in one case study (Matthews, 1992).

Access control can deter crime in several ways. The most obvious being that it prevents exposure of vulnerable targets to potential offenders. To the extent that offenders are impulsive and only commit crime when presented with the opportunity, simply preventing exposure of targets to motivated offenders should prevent crime. To the extent offenders are more motivated and intentionally seek out opportunities, such measures would simply deflect offenders to search for other targets (Jacobs, 2010)[10].

One might also consider aspects of the built environment that act in the opposite way to access control, making them more permeable and have easier access to the general public. Easier access has been suggested to increase crime through making local residents less aware of who is local and who is a foreigner to the space, and hence reduces the ability of local residents to exert informal social control (Browning et al., 2010; Stucky and Ottensman, 2009).

One example of the configuration of the built environment that is linked to the permeability of an area, and has been linked to a reduced amount of burglary incidents, is whether a house is located within a cul-de-sac (Johnson and Bowers, 2010). Another aspect of the built environment that makes a particular location more permeable to public entrance is

[10]There is evidence that suggests diffusion of benefits is just as common as that of displacement for situational crime prevention techniques (Guerette and Bowers, 2009). Note not all SCP techniques would be considered changes to the built environment, but also include police interventions (Weisburd et al., 2006).

via public transportation, and several examples can be found of elevated crime either after public transportation has been developed (Jackson and Owens, 2011; Clare et al., 2009) or around public transportation in comparison to other local places (Levine et al., 1986; Loukaitou-Sideris, 1999; McCord and Ratcliffe, 2009). Another aspect of the built environment that increases the amount of human traffic of non-residents are commercial land uses (e.g. gas station, shopping mall, parking garage) (Wilcox et al., 2004).

4.3 The Number of Potential Crime Incidents

The previous two sections described two causal mechanisms that can describe why certain places have elevated levels of crime incidents compared to other places. Those mechanisms were that the people that live in a particular place have a higher prevalence of crime, or that some places can actually directly affect an exposed individual's propensity to commit crime. This last section details a different mechanism, through the distribution of potential offenders and victims in space. Figure 3 displays the difference between these two concepts.

So exactly what is a potential incident? We may consider any interaction between a potential victim (or an inanimate target, such as a house) and a potential offender as a potential incident[11]. For simplicity in further examples I will consider any person a potential victim and offender (although whether a person could be a victim and/or offender could change when the definition of the crime changes). Restating the theoretically different mechanisms in terms of agents, one would say that theories of crime that explain behavior explain how

[11] Other work has attempted to define what an incident (or a situation) is more explicitly. See Newman (1997) and several of the subsequent chapters in Newman et al. (1997) for some discussion. I largely avoid the discussion here because one could choose different definitions but they are immaterial to the description presented.

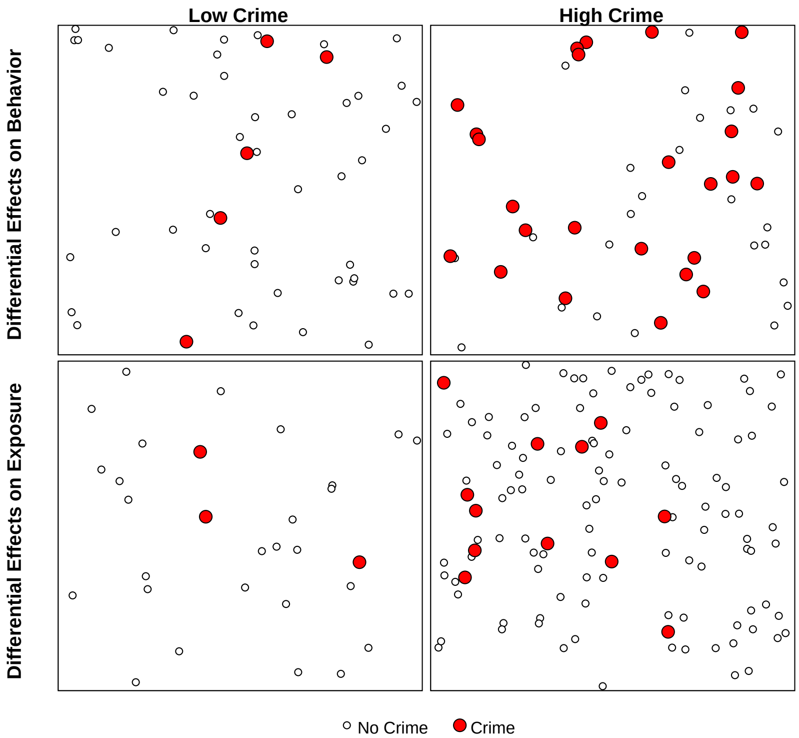


Figure 3: High Incidence versus High Prevalence. The top panels display that the higher crime place has different behavior between the agents in the incidents. Each upper plot contains 50 incidents, but the prevalence of crime is much higher in the upper right panel (50%) than the upper left panel (10%). In the lower panel the higher crime place has more crime events simply due to greater exposure. The plot on the lower left contains 30 incidents, while the lower right contains 120. The prevalence of crime incidents is the same in each of the lower plots though (10% of the incidents).

the propensity to either be victimized or become an offender *differ between agents*. This extends to places as well, an agent, in a specific place, has a different propensity of becoming a victim or offender, *ceteris paribus* compared to a different place (i.e. that place has some attribute that has a criminogenic effect on behavior). Whereas the other mechanism is simply that certain places have a greater number of agents, and so even if the propensity of becoming a victim or offender is equal among agents and places, some places have more crime incidents due to a higher number of potential incidents, due to higher levels of agents (and hence a greater number of interactions) within that space.

Routine activities theory when initially proposed was intended to explain the increasing crime trend nationwide (Cohen and Felson, 1979), and used broad changes in human organization post World War II. It is conceptually simple to extend that theory to more micro units though, either in space or time (Felson and Poulsen, 2003; Groff, 2008). In particular, since the interaction of a victim and a suitable offender is necessary for a crime to occur, variations in the spatial locations where the two agents meet at a higher incidence, all else equal, should result in higher incidents of crime. But where are the places (and times) that more offenders meet victims? That question is within the domain of crime pattern theory.

4.3.1 Crime pattern theory

Most criminals do not spend the majority of their time committing criminal acts, but are engaged in routine activities of everyday life. This, in combination with the fact that offenders rarely go to areas they are unfamiliar with to commit crime, makes the patterns of human activity highly predictive of where crime will occur (Brantingham and Brantingham, 1993).

For a trivial example, think of the pond in Washington Park in Albany. Certainly there is not anything by definition to prevent a crime from occurring in, on or below the pond, but there are likely no crimes that have occurred there. For another example, think of Washington Park when an event is occurring (such as the Tulip Festival). We may expect there to be an increase in criminal events during that time period, not due to the characteristic of the place nor of the people in the place, but simply due to the increased interaction of many more people in that particular location for a short period of time.

Unfortunately to account for this potential interaction is very difficult in causal models of crime at places for two reasons. One is that dynamic measures of the potential incidents at small places do not exist, and this is described further in the section named 'the unknown baseline'. Two is that certain places can have characteristics that actually attract deviant individuals, confounding the characteristics of places with the number of potential incidents. This is described in the section titled 'crime attractors and generators'.

The unknown baseline

The potential interaction of humans is very dynamic, and can vary greatly within a time period at one place. For instance, the street outside of a row of bars at 7 am on a Sunday is likely to be largely devoid of human interactions, whereas that same street just a few hours before at midnight may have many interactions. Even ignoring short term temporal variation, long term measures of human traffic rarely exist for small places either. While using the residential census population estimates may be a reasonable approximation for larger areas, this is unlikely the case for smaller areas. Places where many people interact on a regular basis do not tend to be places where many residences are located. The use of the ambient population may help alleviate this issue some, but even that is likely a larger unit that some may be interested in, approximately a square kilometer (Andresen, 2011).

To what extent is this problematic when measuring and modelling the causes of crime? To the extent that the number of potential incidents is not correlated with other purported measures of crime, our ability to identify the effect of particular causes on crime is not altered. It is largely speculative though the extent to which such a correlation exists. One may wish to not only consider typical demographic measures obtained from the census, but also more local measures of the number of residences obtained from parcel datasets. Also one may want to include other aspects of the built environment that are predictive of increased human interaction (such as the location of shopping malls for instance). Such generator locations will be expanded upon in the next section.

Another effect of the lack of reasonable small place measures of human activity is that it makes no sense to speak in terms of crime rates when we are describing crime at small places. The best we can hope for is to attempt to approximately account for the differing distribution through a series of imperfect and indirect measures. One can not calculate a rate if one has no estimate of the baseline.

Crime generators and attractors

Crime generators and crime attractors were suggested by the Brantinghams to describe human activity in different areas in the city in regards to its impacts on crime (Brantingham and Brantingham, 1995). A crime generator is a place that simply has a high amount of human traffic, most often related to activities totally unrelated to crime. An example is a shopping mall. As stated in the previous section, the increased number of human interactions will often increase the number of crime incidents, although the prevalence of crime in proportion to the number of potential incidents may stay the same. This is the situation Figure 3 is displaying in the lower right panel. Another example is when closing time occurs at a location where many bars are coincident, and so many individuals crowd the streets all at once (Rossmo, 1995).

Crime attractors are places that, due to their nature, *deviant individuals are deliberately drawn to.* They can either be places that attract deviant people for reasons unrelated to crime, or they can be places that attract deviant individuals who have a specific criminal intention. For an example some bars tend to attract more deviant individuals than others (Block and Block, 1995; Briscoe and Donnell, 2003; Cohen et al., 2003; Gruenewald, 2007). Subsequently, these deviant bars tend to have more problematic criminal behavior, such as inter-personal assaults. Another example are areas in which offenders go to buy drugs, and they then go and commit other crimes to generate money to buy the drugs in the local area.

Places that contain a large number of vulnerable targets may attract deviant offenders as well. For instance, Bernasco and Block (2009) indicate that robbers are more likely to choose areas that have higher counts of drug and prostitution arrests, suggesting that such individuals provide easier targets to rob. The obverse of this is situations that repel offenders, such as offenders tend to be uncomfortable travelling to areas that are composed of different racial groups or opposing gang territories (Bernasco and Block, 2009).

Crime generators and attractors are frequently represented in causal models through either nominal coding of different land use categories (Kinney et al., 2008) or identifying whether a place is nearby a noxious institution (McCord and Ratcliffe, 2009). When using aggregated units the measure may be the number of institutions in that particular area or the number of institutions normalized by some other measure, such as per square mile or per road area (Treno et al., 2008). Since land use categories are mutually exclusive, measures of them may include the proportion of a particular use within the aggregated unit (Block and Block, 1995). This is also a good example of the change of support problem when working with different sources of data. Land use data is frequently obtained from parcel datasets, or sometimes areal sensed imagery in the form of a raster dataset for land cover data (Kent and Leitner, 2009).

Some examples of suggested specific noxious institutions are bars (Murray and Roncek, 2008), schools (Roncek and Lobosco, 1983), public transportation wait areas (Levine et al., 1986; McCord and Ratcliffe, 2009) or alcohol treatment facilities (Taniguchi and Salvatore, 2012). Frequently it is admitted a place can simultaneously be a crime generator and a crime attractor, so empirically differentiating between the two is difficult in practice (Kurland et al., 2014).

A further elaboration of the Brantinghams' crime pattern theory identifies path and edges between nodes of high activity (Brantingham and Brantingham, 1993), and so it is not unreasonable to suggest that many of these noxious nodes diffuse crime into other areas of the city. A direct example of this would be implementation of public transportation systems that resulted in an increase of crimes located nearby the new origins and destinations (Clare et al., 2009; Jackson and Owens, 2011).

4.4 Why Understanding Theories of Crime Matter

I will admit that simultaneously representing all of these constructs in a model with a variety of observational data is daunting. All theories presented have multi-faceted constructs, many of which could arguably be represented as reflective latent variables, direct causal constructs, or formative latent variables. Also many of the general theories presented have overlapping theoretical constructs that would be difficult to differentiate between even in optimal experimental conditions. To *test* the veracity of any particular theory it would likely be more appropriate to conduct an experimental intervention or identify unique natural experiments that allow one to identify an effect unique to one theory. This even ignores the potential for interaction between different causes of crime (Cohen et al., 2008; Roncek, 1981; Smith et al., 2000; Taniguchi and Salvatore, 2012).

So why bother with any cross-sectional observational research designs, or similarly what can we hope to learn from such research designs? I argue here that for the future accumulation of knowledge it is necessary to simultaneously understand the limitations of observational and aggregate data as well as the overlap of theoretical constructs that cause crime. Understanding both is essential for developing future research agendas to identify any specific causal link between the crime and its causes. Also understanding both is necessary to interpret the vast prior literature in criminology that has been accumulated, and is still being published with regular occurrence. Aggregate observational studies can be the start of identifying the causal link between crime and its causes (Savitz, 2012), but they will hardly be able to reach any definitive conclusions in regards to causality. In what Platt (1964) describes as strong inference, the role of experiments is to *eliminate* potential hypothesis as causes for any particular phenomenon. In this sense, aggregate level observational studies are very limited in what they can eliminate as potential causes for the observed patterns.

In instances in which experimental or quasi-experimental designs are not feasible, which is frequently the case when using spatial units of analysis, one should not ignore the potential to learn information from such observational settings. For instance, I suspect no one would make the argument that the knowledge gained from the PHDCN data collection is irrevocably tainted because of its limits on inference for such observational data. Sampson (2010) gives many other examples of past accomplishments that would have been impossible in an experimental design, and how further advances in econometric modelling provides the potential opportunity to identify causal antecedents of crime.

All models are essentially abstractions of reality (as the George Box quote goes 'all models are wrong, but some are useful'). Hence the goal is furthering knowledge of the specific causal processes, down to more and more specific causal processes (Blalock, 1972).

Chapter 5

The Theory and Measurement of Spatial Effects

Part of the motivation to evaluate the correlates of crime at micro places is to distinguish between local effects and spatial effects that are confounded when one aggregates up. This chapter will detail where spatial effects might occur and how one can empirically estimate those spatial effects.

5.1 Where Might We See Spatial Effects?

Here I outline four processes through which spatial effects might theoretically be present in criminology, and relate these to the place based theories of crime I elaborated on in chapter 4. Those four processes relate to the characteristics of people at places, the characteristics of the places themselves, the distribution of the population that may cause spatial effects, and spatial effects of crime, in and of itself.

Spatial effects here can be defined as the consequences nearby in space of the characteristics of a particular place (Townsley, 2009). The effect we are mainly interested in this dissertation is the effect on crime. For an(other) example of a spatial effect, one can imagine crime around (but not within) a public housing project might be elevated above normal, where normal is the level of crime a place would experience in the counter-factual world in which all else is equal but the public housing project does not exist. This may be due to the public housing project contributing more offenders to the local pool, or it may be public housing contributes more victimizable targets. Either way, the public housing project has a spatial effect on the amount of crime at nearby neighboring locations (Davies, 2006; Haberman et al., 2013).

5.1.1 Demographics

The most obvious way for the characteristics of people at places (e.g. demographic characteristics) to contribute to crime in areas proximate to where those people live is through the fact that certain places have more people who are crime prone. Those criminals may travel into neighboring local environments to commit their crimes (Bernasco and Block, 2010).

The previous discussion of the characteristics of people at places though focused on how certain places have differential levels of informal social control. This seems less obvious how it could result in spatial effects. The typical conceptualization of informal social control typically measures the process as a neighborhood one, i.e. a process that is actually defined by the interaction of multiple individuals within a neighborhood. More recently some scholars suggest utilizing variable buffer bespoke neighborhoods instead of discrete areal units (Hipp and Boessen, 2013; Li and Radke, 2012). This would result in a continuous gradient as opposed to discrete neighborhoods.

In fact, the suggestion to use such bespoke neighborhoods makes it impossible to differentiate between local and spatial effects, the same way, it just precludes one from specifying the relationship via some aggregate unit. It may be considered preferable in as much as the neighborhoods do not realistically represent the actual spatial relationships (Anselin and Arribas-Bel, 2013), but if one is theoretically interested in the distinction between local and spatial effects, this smoothing of the estimates confounds the ability to distinguish between the two (Ratcliffe and Taniguchi, 2008). For instance, positive spatial effects of being located near predominately white communities would be unidentifiable with the bespoke neighborhood model (Peterson and Krivo, 2009). Segregated black neighborhoods nearby white ones would just be measured as mixed racial neighborhoods with bespoke neighborhoods or larger aggregated neighborhoods.

5.1.2 Built environment

Much of the interest in how the built environment potentially impacts neighboring areas is focused on how CPTED initiatives introduce spill-over of crime into nearby areas. For an example, a publicly visible video camera may prevent a particular robber from targeting a victim within a parking lot, but if the offender is well enough motivated they will merely search (or forage) in the nearby area for a more vulnerable target (Jacobs, 2010; Johnson et al., 2009). A specific example can be found in Pirkis et al. (2013), a meta-analysis of CPTED initiatives to make it more difficult to commit suicides by jumping off bridges through the use of physical barriers. While the meta-analysis found an overall decrease in suicide at bridges, there was evidence of spill-over of suicides at other nearby locations. The reduction in suicides at the local site was greater than the spill-over suicides though, so overall there was a net reduction.

But CPTED articles frequently show *diffusion of benefits* into local areas (Bowers et al., 2011), as opposed to simply moving crime around the corner. The fact that overall crime is lower post-intervention is explainable in a rational choice perspective, if individuals were on the margin to commit crime (or suicide) previously, the added work of finding a new location should decrease the number of individuals who will perform the act. Diffusion of benefits though is more difficult to theoretically explain, as the CPTED initiatives don't directly impact neighboring crime, and indeed theoretically would have a negative impact. Continuing on with the suicide example, it would be similar to making one bridge difficult to commit suicide from by installing wire fencing, but then find fewer people commit suicide from jumping from a nearby high-rise building. Although Pirkis et al. (2013) did not find this, it is an analogous situation to the current criminological findings.

A potential explanation is that areas undergoing interventions have a sort of Hawthorne effect on the perceptions of would be offenders, and this perception diffuses into surrounding areas (Roethlisberger, 2001). A real diffusion in police patrolling of experimental decided hots spots is that police need to physically travel to wherever they are patrolling. Thus experimental units tend to have a greater exposure to patrol just from the travelling to and from the point of interest (Larson, 1975). Consistent with this theory is the fact that police interventions tend to have a decay in time on their deterrent effect (Cohen et al., 2003; Ratcliffe et al., 2009; Taylor et al., 2011; Wyant et al., 2011), although that time to decay varies substantially between those examples[1]. Weisburd et al. (2012) suggest that crime and place are so tightly coupled together that crime spill over would not be expected, but this fails to explain the diffusion of benefits.

Another way in which characteristics of the built environment generate effects that can diffuse into neighboring areas is the fact that certain visual cues can be observed from a specific distance, e.g. I see garbage on the other end of the street, so I assume at this end of the street no one will care if I litter. Or that certain places can be stigmatized as neighborhoods, and so any physical disorder at one particular place is indicative as to the nature of the entire neighborhood (Sampson and Raudenbush, 2004). The former would then result in diffusion within a specific visual area. The latter would result in spatial effects only as much as two places nearby in space tend to be encapsulated in the same neighborhood. If neighborhoods were clearly defined between people (like in places where physical barriers clearly delineate neighborhoods), this would result in a contextual/neighborhood data generating process.

A final process through which the built environment may cause spatial diffusion of crime is through public transportation, which reduces barriers to travel. As the antithesis to some aspect of the built environment being impermeable, public transportation may make travelling to distal areas of the city easier, and thus facilitate the journey to crime (Clare et al., 2009; Jackson and Owens, 2011).

[1] Another alternative explanation, posited by Taniguchi et al. (2009) relating drug sales to agglomerative economies, suggests that clustering of criminal events creates a more visible market and thus increases potential customers.

5.1.3 Population

The distribution of population can result in spatial effects of criminological processes because ambient population density (that is the walking around population, not just the residential population) tends to be a continuous gradient and rises and falls during the ebb-and-flow of other naturally occurring events. For instance, Ridgeway (2007) reports that in particular places of Manhattan the daytime population swells by a factor of 20 over the residential population (and thus uses the daytime population as a benchmark measure of stop rates). Average rates of the walking around population are unlikely to be indicative of special events as well (e.g. imagine the walking around population in Manhattan during the Thanksgiving Day parade or on New Year's Eve).

This can result in a spatial spill-over effects, because such a swell of individuals is unlikely to be contained into a prescribed area, but diffuses into nearby areas. For an example, Kurland et al. (2014) use the exogenous variation in the number of individuals during and after soccer matches (and use the LandScan ambient population estimates) to tell if the rate of crime is similar during soccer matches nearby a stadium. Rossmo (1995) coins the term *potentiation* to describe when bars close and the local streets are flooded with individuals, causing various intoxicated individuals (who are at both a greater propensity to commit violence and be victimized) to come together. Because humans can move, any location (or event) that attracts many individuals may diffuse victims and offenders into nearby areas, causing crime to increase nearby.

5.1.4 Crime itself

Crime *in and of itself* may result in spatial effects. Some examples given in prior literature are retaliatory violence (Short et al., 2010; Loftin, 1986; Radil et al., 2010), which would result in positive spatial autocorrelation of violence, or competition between gangs which would result in negative auto-correlation of the gang territories themselves, but is found to result in escalated violence nearby the edges of the gang territories (Brantingham et al., 2012).

Note that although crime may show positive spatial auto-correlation, this isn't *per se* evidence that crime in one location causes elevated amounts of crime at other nearby locations. Lesage and Pace (2009) give various examples of data generating processes which would result in positive spatial auto-correlation, and one might want to interpret positive spatial autocorrelation in terms of indirect effects of the independent variables as well. Positive spatial auto-correlation is typically taken as a problem that needs to be corrected to make proper inferences, as opposed to being interpreted in substantive theoretical terms (Townsley, 2009).

5.2 How Do We Measure Spatial Effects?

A usual regression model, predicting y from independent (or exogenous) variables x can be formulated as;

$$y = \beta_0 + \beta_1(x) + e \tag{5.1}$$

In (5.1) one then interprets the conditional expectation of y as the sum of the intercept, plus $\beta_1 \cdot x$ (for any value of x, assuming a mean zero error term). A spatial auto-regressive (SAR) model formulates a way to incorporate spatial effects of the dependent variable in the model via a spatial weights matrix, and can be written as;

$$y = \beta_0 + \beta_1(x) + \rho(W \cdot y) + e \tag{5.2}$$

Where W is a spatial weights matrix, that is sized n by n, where n is the number of observations. This matrix is then post-multiplied by y, producing a column vector. Note, an unbiased estimate of ρ in (5.2) needs to be estimated either via maximum likelihood, which is derived in Anselin (1988) and Lesage and Pace (2009), or via two-stage least squares (Land and Deane, 1992). As $\rho(W \cdot y)$ and y are endogenous, OLS estimates will be biased, and with other independent variables in the model will tend to over-estimate ρ assuming the spatial effect is positive.

Typically the W matrix has zeroes on the diagonal, and is scaled in other ways that either improve interpretation of the spatial effect or make estimation of the spatial effect simpler. For an example, if one uses a contiguity based matrix, that is W has ones for every observation that is a neighbor, and zeroes otherwise, and one subsequently *row standardizes* that matrix (i.e. the rows sum to one), the resultant product of $W \cdot y$ can be interpreted as the average y of the neighbors.

In (5.2) both the effect of $\beta_1(x)$ and $\rho(W \cdot y)$ are more difficult to interpret than in OLS. $\rho(W \cdot y)$ can be interpreted similarly to an auto-regressive component in a time series model, and can be considered in terms of of a causal relationship (e.g. crime in this location causes elevated amounts of crime in a nearby location) so the observed cross-sectional process is the result of unobserved temporal diffusion. Or $\rho(W \cdot y)$ can be interpreted as merely the outcome of some other unobserved process, and we merely wish to estimate its effect to provide other unbiased estimates. Note here that, unlike OLS, the change in y with an exogenous increase in x of size 1 has *a larger effect* on the conditional expectation of y than β_1 when ρ is positive. This is because one also needs to consider the fact that when the local value of y increases, so do its neighbors through the auto-regressive term. That is there are indirect spatial effects.[2]

A generalization of (5.2) leads to a spatial Durbin Model (SDM);

$$y = \beta_0 + \beta_1(x) + \rho(W \cdot y) + \theta(W \cdot X) + e \tag{5.3}$$

Here there are spatial effects in both the independent variables and dependent variable, along with local effects of the independent variables. A model where ρ equals zero would result in a spatial specification of a conditional auto-regressive model (CAR), which can either be interpreted in terms of heteroscedasticity in the error term (which is modelled explicitly via the other independent variables) or merely as spatial effects of the independent variables. All of these models can be extended to accommodate multiple spatial weights matrices as well (see Deane et al. (2008) for an applied example of multiple spatial weights matrices).

Considering full texts of how one goes about estimating spatial effects exist, I will not repeat them here. What deserves some elaboration on is how one represents those spatial

[2]An simplified example of this is that if y is a function of x and of a neighboring y values, Wy. If you increase x at one location the location i, it not only makes y_i increase, but also makes Wy increase for different nearby observations. So x_i increasing makes both y_i go up (the direct local effect) and other y's to increase as well (the indirect spatial effect).

effects via the spatial weights matrix. Earlier in the chapter where I defined spatial effects as *the characteristics of a particular place have an effect on areas nearby in space*, the spatial weights matrix how you define what nearby is. For example, in a binary contiguity matrix, if one's unit of analysis is an area (such as a census tract) one may define a neighbor based on adjacency of neighbor tracts (i.e. tracts that either share a vertex or an edge of their polygons are considered neighbors). Instead of relying on such fixed criteria, one could theoretically manipulate adjacency based on other external information. For an example, one may not want to consider two census tracts on opposite sides of a highway to be considered neighbors, as they are much less likely to influence one another due to the physical barrier (Fagan et al., 2012).

When working with either point data or data on a road network, there become even more choices. One can formulate a spatial weights matrix via the voronoi tessellation of the points, or consider other continuous based models such as inverse distance weighted (one can use these for areal lattice data as well, although they appear to be much less common in criminology).

Although Anselin (1988) suggests that one should form the spatial weights based on the theoretically posited relationship, theory in criminology (or much of the social sciences) is not strong enough to *a priori* distinguish between variants of different spatial weights matrices. This is opposed to a data based exploration to define W, such as that suggested in Negreiros et al. (2010) or what is similarly done in Box-Jenkin's ARIMA modelling for time series. Fortunately though we may consider this ambiguity (and resultant error from it) rather small. Various spatial weights matrices tend to be highly correlated, and so one might suspect that even with misspecification one would reach similar conclusions (Lesage and Pace, 2010).

5.3 Conclusion

As part of the motivation for examining smaller units of analysis is to decompose the difference between local effects and spatial effects, this chapter details in what situations spatial effects occur in criminology. It also goes into detail about how one represents spatial effects via spatial regression models and spatial weights matrices, and how spatial auto-regressive (SAR) models can be interpreted both in terms of local effects and spatial indirect effects.

Chapter 6

Data Management of Small Geographies

Part of the difficulty of conducting data analysis at small geographic units of analysis are the variety of data management tasks one needs to conduct; from geocoding crime data or using data geocoded by various address locators, to coercing explanatory variables measured at several different supports to a similar unit of analysis, and even to make a consistent small unit of analysis across a city with an irregular street grid.

In this chapter of the dissertation I present the data I will be working with for the analysis, along with several data management tasks to go from the raw data disseminated at separate geographic supports to a set of measures at the common street unit. I end the chapter with a description of how I generate a spatial weights matrix for the street units.

Care is taken to be explicit about how such data manipulations were conducted and to provide graphical examples to demonstrate the manipulations. Along the way I provide discussion of why I believe such arbitrary decisions are reasonable, why using census geographies with recorded crime data is difficult to justify, and potential bizarre examples of small units of analysis when working with an irregular street grid. Journal articles rarely afford such explicit descriptions, but such details are necessary to make the research reproducible as well as understandable, especially by those less familiar with geographic data analysis.

6.1 Describing crime data on DC.gov

In this dissertation I use publicly available data on crime and the built environment published at `http://data.octo.dc.gov/`, as well as census geographies. The crime event data include the date of the report, the type of crime, and the geographic location. The specific time of the incident is not available, and geographic coordinates are only disseminated at the specificity of the midpoint of a street block or an intersection.

The types of crimes disseminated are near verbatim Part 1 UCR crime events. Those crimes are; homicide, sex offense, robbery, assault with a deadly weapon (ADW), larceny, burglary, stolen auto, theft from auto (TFA), and arson. This differs slightly from the UCR Part 1 scheme in that TFA would be included as a larceny, and aggravated assaults do not *per se* have to be committed with a weapon. Crime is uploaded to the dc.gov internet site on a regular (near real time) basis, and exists back as far back as 2007. This dissertation will use crime events reported in 2011 in a cross-sectional research design.

The final unit of analysis that I use in this dissertation is what I will call the *street unit*. I assign a reported crime location to the nearest street unit in a 1 to 1 mapping: every reported crime was matched to the nearest street unit[1]. Figure 4 shows an example of this procedure,

[1]Nearest is in euclidean distances. The projection used for all analysis is NAD 1983 Maryland State Plane.

Table 3: Reported crimes from DC.gov. This table only includes events that had a geographic coordinate and where inside of the D.C. boundary.

Crime	**2011**
Arson	38
Homicide	108
Sex Offense	171
Assault with Weapon	2,481
Stolen Auto	3,746
Burglary	3,893
Robbery	4,166
Theft from Auto	7,766
Larceny	10,071
Total	32,440

with crimes shown in triangles and street units shown in circles. Crimes that share the same color as a street unit are matched. The majority of matches in this sample are similar; in that they tend to be very close to the street unit with some random displacement. The following section further details how street units are made. Crimes that either had missing geographic coordinates or fell outside of the D.C. boundary were discarded. This results in a total of 32,440 crimes used in this analysis for 2011, and Table 3 displays the distribution of crime types in the sample for 2011.

6.2 Describing street units

Here I define a street unit to be either the mid-section of a street, with two cross streets defining the begin and end of the street, or the intersection of two streets. Figure 5 shows a simple example for a location in Washington, D.C. For this project midpoints of street units were created from the midpoint of the street segments (from a street centerline file). Intersections are created wherever two separate street segments touch. I subsequently pruned

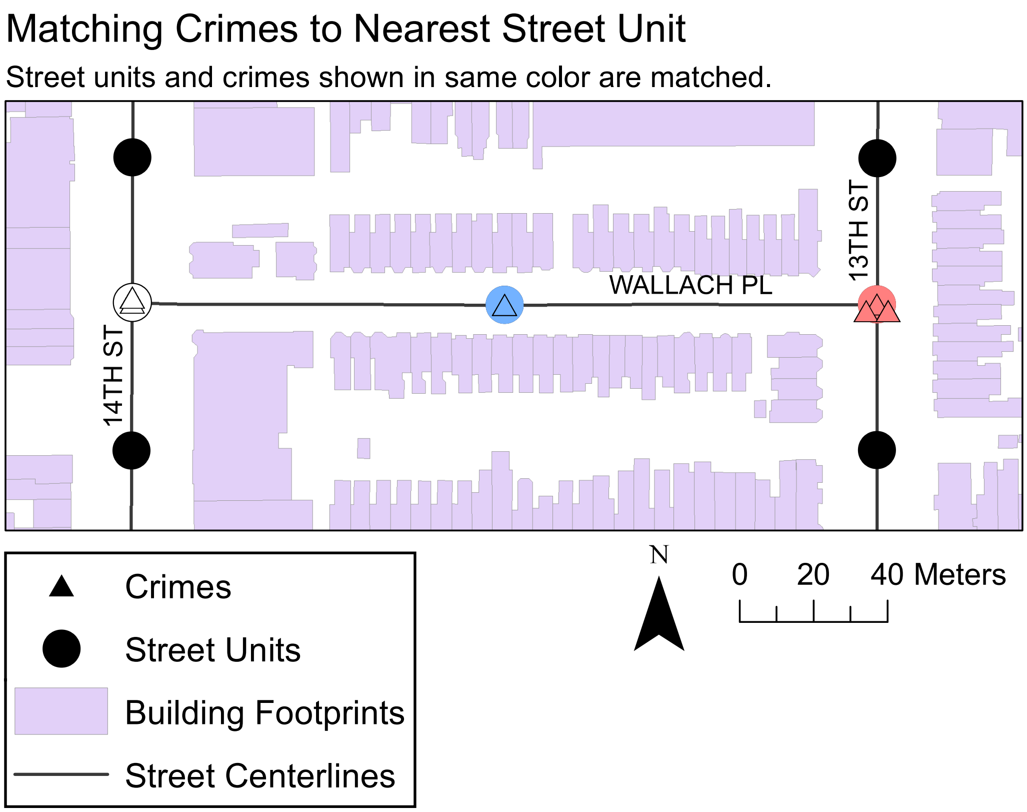


Figure 4: This figure shows how original crime incidents were matched to the nearest street unit. While some incidents are not reported exactly coincident at either a street midpoint or intersection, they tend to be quite close to a street unit.

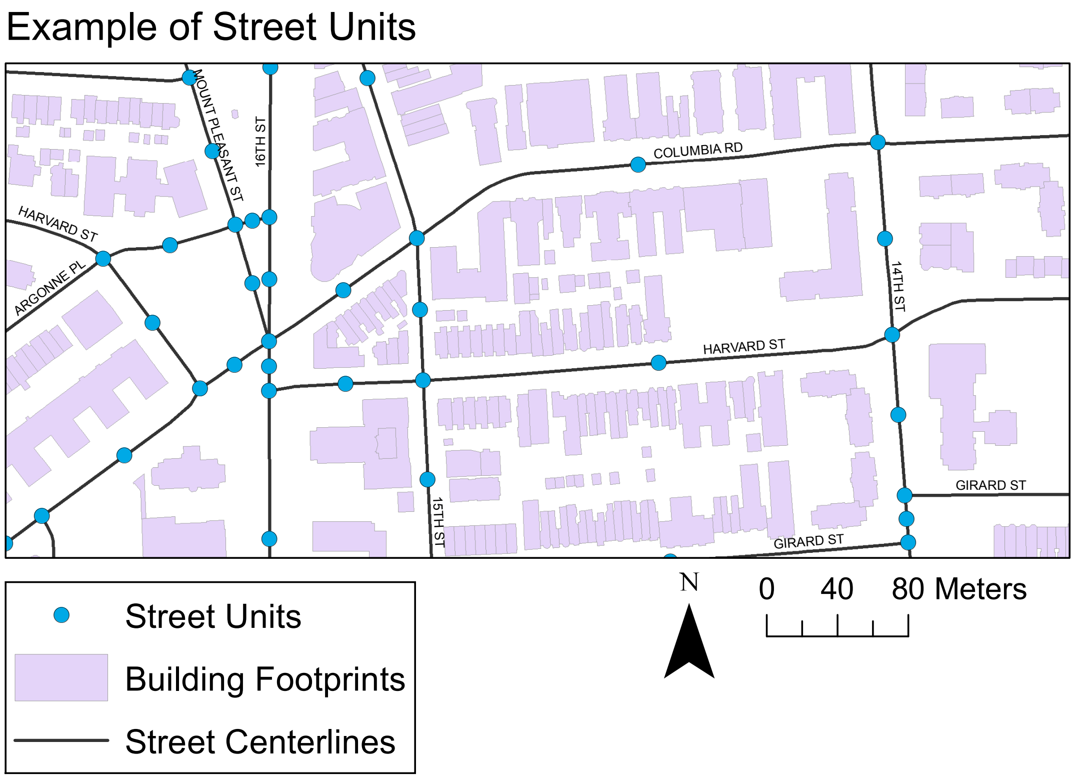


Figure 5: The above figure displays a sample of street units for Washington D.C. Some street midpoints tend to be very close to other intersections in areas where streets converge.

intersections and street midpoints that were near overlapping (which often occur at T intersections) by defining near as within 5 meters[2] and manually inspecting the locations. Such conflicting locations I always defined as an intersection. In total this results in 21,506 street units within D.C., with 8,172 being intersections and the remaining being midpoints of street segments.

Figure 5 shows both the simplicity of the approach of defining street units as I have here as well as some absurd situations. In this example one can see where Mount Pleasant St., 16th St., and Argonne Rd. *nearly* intersect, but manage to produce three street units all

[2]Wherever I choose a buffer to identify points as nearby it is admittedly completely arbitrary. I frequently take 5 meters to be judicious enough to correct for false alignment or spatial resolution issues while not flagging too many false positives.

within very close proximity (another example can be seen where Girard St. intersects 14th St. in the lower right of the map). This is problematic in that when recording crimes it is unlikely the midpoint of Argonee in between the intersections of Mount Pleasant and 16th is likely to ever be recorded as a crime location, or even be a location that will be returned depending on the geocoding service.

Unfortunately such locations occur in abundance in any natural street configuration that is not a perfect grid (I randomly chose this location as an example, they are not hard to find). If the analyst defines the *possible* geocoding locations, a simple solution would be to use street units that are within the potential population of locations that could be actually geocoded. With online geocoding services though this is not possible, as the potential population of locations is both hidden and likely changing over time.

Another potential solution would be to identify such locations where many street units are in close proximity, and manually delete those street units which are too close. This is quite arbitrary though, and given the number of such locations in the dataset a very time consuming task (unless otherwise automated, such as taking the centroid of a set of clustered points).

A final solution, and the one that I employ in these analyses, is to incorporate information into the model predicting crime that would identify a street segment as unlikely to be a location where a crime is geocoded (or in general occur). Simple ones that can be deduced from this map are the length of the street segment, whether any buildings are adjacent (or a similar proxy would be land use for neighboring parcels), and the closeness of neighboring street units[3]. This avoids needing to arbitrarily prune street units, and is a step that should

[3]Closeness of street units should be measured as an average of several of the neighbors. If you only choose

Table 4: Number of street units and number of crimes assigned to intersections.

Unit Type	Street Units	Crimes Reported
Midpoint	13,334	25,005
Intersection	8,172	7,435
Total	21,506	32,440

be taken when modelling crime at small street units anyway.

6.3 Intersections need to be included

Prior work by David Weisburd and colleagues have only used street midpoints and have discarded crimes that were listed as occurring at intersections (Weisburd et al., 2004, 2009, 2012, 2014). In this analysis I use both street midpoints and intersections as far too many crime incidents would be discarded if I only used street midpoints. Table 4 shows the number of street units and crimes that were assigned to intersections and midpoints, and Figure 6 shows the percent of crimes that were recorded at an intersection broken down by crime types. The labels in Figure 6 show the total numbers of crimes that are assigned to intersections for each crime type and the total amount of crime.

It is common for crimes that occur outside to be recorded as occurring at the nearest intersection (Braga et al., 2010; Levine and Kim, 1998). So for certain types of crimes that only occur at residential addresses (such as burglaries) it would seem reasonable to not include intersections if you had access to the original residential addresses (and could geocode them yourself). It is becoming popular for police departments to disseminate crime data already geocoded on online maps, and in most circumstances the data are already obfuscated

the distance of the nearest neighbour, the model will not be able to identify which of the nodes is the more likely winner of crimes recorded nearby.

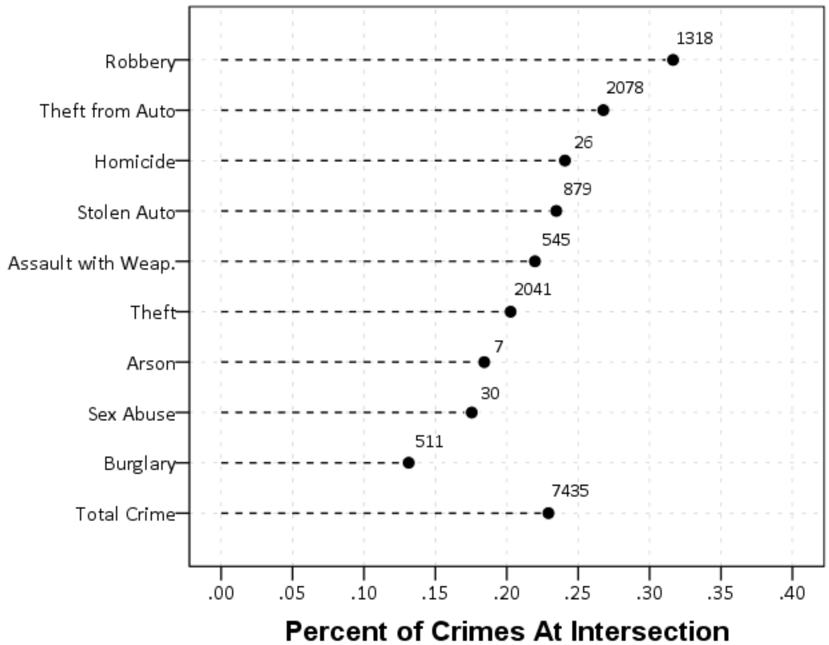


Figure 6: Percent of crimes that are associated with an intersection.

to either the intersection or the street midpoint. Figure 6 shows that 511 burglaries were associated with an intersection, which is a limitation of the data not being able to geocode the incidents directly - as it is obvious no one burglarized a public intersection.

But Figure 6 shows that crimes that are more likely to occur outdoors, such as robberies and motor vehicle related crimes, have a higher proportion of incidents associated with intersections. So in this particular circumstance, with the already geocoded D.C. data it is unreasonable to discard such a large number of events to only evaluate street midpoints. In general even in the optimal circumstance in which you have the original address data, for crimes that have any substantial portion of events that occur outdoors one should include intersections in that analysis to avoid discarding a large proportion of data. Even in the case in which a small number of events occur at intersections, one should be wary that discarding crimes at intersections introduces systematic bias in the analysis. For any type of spatial analysis of crime it is plausible crimes that occur outdoors are systematically different from crimes that occur indoors.

6.4 Problems with census units

When geocoding to the exact residential address, street centerline files[4] often have associated information to offset the crime to the appropriate side of the street.

When the crime location is not offset from the street centerline (or geocoded to an actual parcel) it is problematic to associate crimes with census geographies as the outlines of the areas are coincident with the streets. As previously discussed it is increasingly popular for departments to only release geographic data at the level of the street midpoint or intersection. Figure 7 shows an example of the crime locations disseminated by D.C., and how they almost always fall on the border between two census geographies (and when they do not it is just as likely that error in the spatial accuracy of the georeferenced location is what provides the identification).

Figure 7 displays census blocks, which are the smallest area at which the census disseminates population information. Aggregating up to larger census units, such as block groups or census tracts, only partly mitigates the problem. To show the extent of the problem I converted the census geographies to polylines, and then marked a street unit as near a census border if it was within 5 meters or less of the census geography border. This nearby definition is necessary because polylines have no area, and slight differences in the resolution of

[4]Street centerline files are those that contain a polyline with an associate set of begin address numbers and ending address numbers corresponding to the begin and end of the street. Often these files have coordinates that discriminate whether even or odd numbers are on a particular side of the polyline section. Then the location of a particular address is interpolated along the polyline according to the address ranges, and offset an arbitrary amount on the corresponding side of the street.

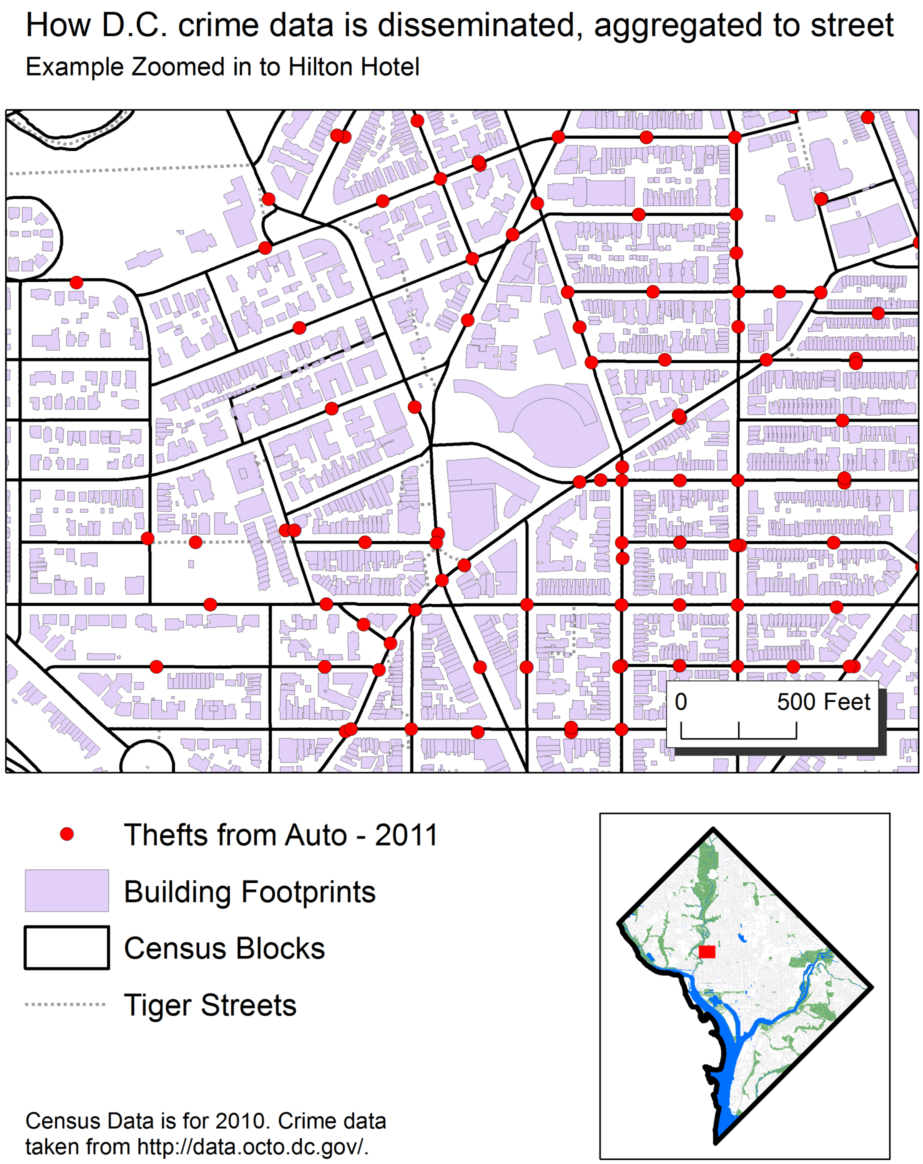


Figure 7: The Washington D.C. Metropolitan police disseminate crime data already aggregated to either the centroid of a street segment or the street intersection. This is useful in that small area crime patterns can be examined, but is very problematic if one wanted to use census aggregations as units of analysis. Crime almost always falls on the border of two census blocks.

Table 5: Percentage of street units near borders of Census Geographies.

Unit Type	**Block**	**Block Group**	**Tracts**
Intersections	98	46	30
Midpoints	94	30	18
Total	96	36	23

the recorded spatial data between different sources may make points appear within a census unit, although it is only due to alignment discrepancies between the geographic sources.

Table 5 shows the number of street units for Washington D.C. that are nearby census borders, broken down by street midpoints or intersections and by levels of the census geographies. This shows that when using the smallest census geographies of blocks, nearly every intersection or street midpoint falls on the border of a census block. The percentages decrease by a large proportion when using block groups or tracts, but are still non-trivial. If you have access to the original data that are not already obfuscated to the street midpoint, one could potentially use offset geocoding to identify which census geography a crime occurred in. This does not apply though to incidents that were originally recorded at an intersection. Even when using census tracts, 30 percent of intersections are near the tract border in Washington D.C. In this sample crime data, of the 32,440 serious crimes, one would be unable to identify the census tract of 1,640 (five percent) of the crimes that are geocoded to an intersection.

6.5 Allocating other measures to street units

While matching the crimes for DC.gov to the nearest street unit may cause little problem, associating other aspects of the built environment to those same street units will be problematic. We may want to allocate measures of areas, such as whether a street unit is nearby a park or the amount of residential floor space nearby. Or we may want to allocate the measure of a line to that location, such as the length of that particular street segment. To make these allocations I will calculate the Thiessen polygons for the set of street units and then take the intersection of areas and lines to allocate measures to those street units.

Given a set of points, the Thiessen polygon for a location is the space that is closest to the particular control point. That is, for every street unit, if I form a Thiessen polygon with respect to the other street units, any place within that polygon is closer to that focal point than any other point on the plane (Boots, 1986). Figure 8 provides an example of defining Thiessen polygons for a sample of street units.

For measures of areas and lines, I will take the intersection of the Thiessen polygon layer and the area (for polygons, e.g. parks) or length of line (e.g. streets) that covers it. For areas I will assign a measure of the intersection area with the focal point, and for lines I will assign a measure of the length. Area measures if needed later on can be normalized to be percent of area, e.g. 50% of the Thiessen polygon is covered by a park. For areas of the Thiessen polygon I will only count actual land area, I will not count area that is covered by water. This causes a few street units on bridges to assume zero area, which when calculating the logarithm of the area I will reassign these as have a value of $\log(1) = 0$.

One of the reasons I focus on 1 to 1 allocation of exogenous factors is that it allows easier interpretation of subsequent models. With allocation strategies that do not allocate all of the observations exclusively, one can not interpret model parameters as simply the expected increase in crime given a unit increase in some exogenous factor. This is because any other allocation strategy might increase the measures more or less than one unit, and the model

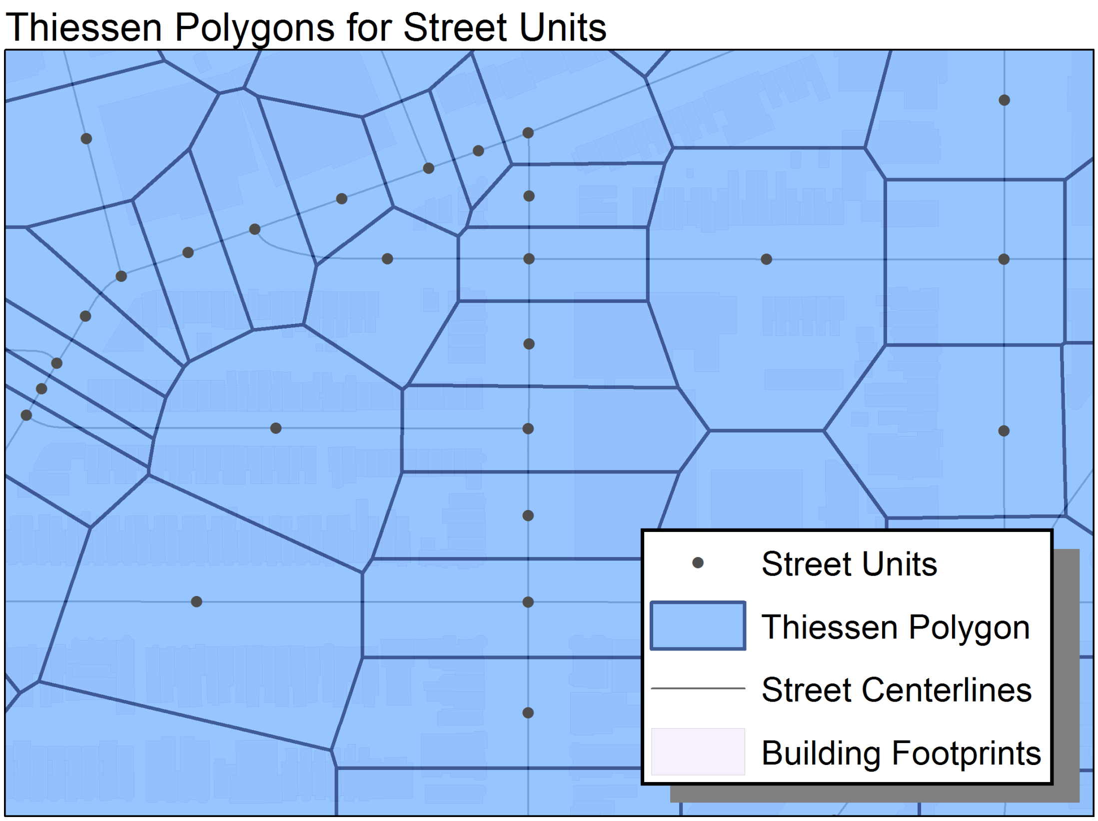


Figure 8: Example of constructing Thiessen polygons for street units. The polygon around a particular street unit marks the area where it is closer to that street unit versus any other street unit.

derivatives will be spatially variant. Tobler (1979) refers to such measures as pycnophylactic; if you take the sum of individual units to which you allocated the measures, it will equal the total of the measures with which you started.

For example, if I estimated the expected decrease in crime if a street is within 300 feet of a community garden, I can not say if you add one community garden in the city it will decrease the expected amount of crime by the parameter estimate. The actual expected value decrease would be dependent on where you put the community garden and how many street units are within the 300 foot exposure area. The location of street units are basically fixed aspects of the built environment, and so it does not make sense to interpret models as if you can exogenously change a street location to be within 300 feet of a community garden. For this reason I choose the strategy of discrete allocation based on the intersection of areas and lines with a street units Thiessen polygon. In the community garden example if I consider it to be a point location, I could say if you add a community garden to a particular street block it will decrease crime by a specific expected amount. The expectation will be spatially variant if one considers spatial effects (e.g. a community garden can affect crime on neighboring streets) or interaction effects, but as these are likely to be smaller than the direct local effects, the direct effects will still be of greater import and more easily interpretable.

For point locations, I will use the same procedure as crimes and allocate each individual unit to the nearest street unit in a 1 to 1 mapping. This is equivalent to assigning points within a street unit's Thiessen polygon, as the polygon is defined as all areas that are closest to that particular street unit. Another potential approach is to estimate smoothed distributions (e.g. the kernel density estimate of the points) (Hipp and Boessen, 2013; McCord and Ratcliffe, 2009). This could be viewed as useful as it allows a crime generating location to increase crime slightly further away than the exact address location (e.g. a bar can increase crime both outside its front doors as well as down the street). This bespoke approach however would not allow one to identify the differences between local effects and spatial effects.

To give a particular example, imagine we have a location with two neighboring streets, and on street A there are 5 bars and on street B there are 0 bars. If we make a smooth estimate of the density of bars, the end measure of bar density would be smaller than 5 on A, but higher than 0 on B. This is a convenient single measure, as we may expect some of the crime to spill over from street A's bars to street B. It is inconvenient though if one wants to answer the question, *Do bars increase crime on neighboring streets?* To be explicit, with one measure of bar density, one may estimate the model:

$$\text{Crime} = \beta_0 + \beta_1(\text{Bar Density}) + e \tag{6.1}$$

Here crime is a function of the estimate of bar density for any particular location or unit. To state a simple example of *why* this model can not answer the question do bars increase crime on neighboring streets, one can imagine the scenario presented above in which *only* streets like A have any crime and streets like B have zero crime. In that example β_1 would still be positive, even though bars *never* cause crime on neighboring streets! As opposed to making one estimate of the effect of bars at any location, I am suggesting to estimate two effects:

$$\text{Crime} = \beta_0 + \beta_1(\text{Bars on Same Street}) + \beta_2(\text{Bars on Neighboring Street}) + e \qquad (6.2)$$

In equation 6.2 β_1 would be considered the local effect. If you allow one more bar to be licensed on a particular street, β_1 would be the average increase in the number of crimes on that same street. β_2 would be the spatial effect, and could be interpreted as if a bar is licensed on street A, it would on average increase the number of crimes on street B by β_2. This is equivalent to the Spatial-Durbin model presented in chapter 5 that only includes spatial lags of the explanatory variables and does not include spatial lags of the endogenous variable. To construct these estimates it is necessary to make definitions of what neighbors are.

6.6 Data Sources

The same dc.gov website that provides the crime incident data also provides an incredible number of potential independent variables that have been stated in the past to affect crime. Table 6 provides a listing of independent variables that one can find at dc.gov that I have used in subsequent analysis. The variables are based on the prior literature review, and limited on an ad-hoc basis. I encourage all reviewing to go and check out for themselves the hundreds of variables present (and certainly more could be argued theoretically relevant or at least worthy of being examined). Practicality necessitated trimming the list to these particular variables.

Table 6: Sources of Data and Characteristics they should be measuring.

Name	**Support**	**People**	**Places**	**Opportunities**
Crime Data	Points			
311 Service Requests	Points		X	
Registered Vacant Property	Points		X	
Green Site	Points		X	
Trees	Points		X	
Parks	Polygon		X	X
Sidewalks	Polygon		X	
Street Lighting	Points		X	
Outdoor recreational facilities	Points			X
Litter Cans	Points		X	
Toxic Release Inventory Sites	Points		X	
Roads	Polygon		X	
Public Housing	Polygon	X	X	X
Alcohol Licenses*	Points			X
Places of Worship	Points			X
Halfway Houses	Points	X		X
Hospitals	Polygon			X
Police Stations	Points			X
Place of Worship	Points			X
Bus Stops	Points			X
Subway Entrances	Points			X
Halfway House	Points	X		
HIV Clinic	Points	X		
Public Housing	Points	X		X
Library	Points			X
Schools	Polygon			X
University Area	Polygon			X
Recreation Area	Polygon			X
Wireless Hot Spot	Points	X		X
Shopping Centres	Points	X		X
Sidwalk Café	Points	X		X

*Alcohol licenses are subsequently split between types liquor stores, grocery or convenience, and bars or restaurants.

The table also includes a column stating the support of the variable, the spatial unit at which the variable is measured or disseminated, and the potential theoretical association. That is the column **People** refers to the demographic characteristics of people who *reside* at that place, **Places** refers to the characteristics of the built environment that may influence crime, and **Opportunities** refer to the number of people interacting in a particular place. Note that not all measures theoretically discriminate between the different theoretical categories. I applied the checks in a purely ad-hoc way as well, and certainly argument could be made for including (or not including) check marks in any particular column.

Point locations in subsequent analysis were simply counted up if they fell within the

Thiessen polygon of a street unit. Areas, with the exception of sidewalks and road area, were recorded as dummy variables. So if the Thiessen polygon of a street unit intersected the area of public housing complex it was recorded as 1 for that location, and as 0 if it did not intersect one of these locations. The *area* of sidewalks and roads were included in the model, as more sidewalks would be a good indicator of foot traffic, and more road area would be an indicator of areas in the periphery of the city with little foot traffic.

6.7 Choosing independent variables

There is a near limitless amount of data one can incorporate into such a model. Chapters 8 and 9 focus specifically on the relationships between bars, 311 calls for service, and crime. These factors could ultimately be confounded with other variables though as well, so additional control variables are included in Chapter 10 in which I build a general model of crime.

The strategy used here was to incorporate a series of covariates based on the prior literature at small places based on a series of relationships to general theories of crime. Particular attention was paid to exogenous variables that would increase the general walking around population at a particular location (Wilcox and Eck, 2011), but several variables measuring other general theories of crime were incorporated.

The main focus in later chapters is on specific analysis of the relationship between bars, 311 calls for service, and crime. In those later chapters more detail will be given to those specific measures, but here I briefly recap the other control variables used in subsequent models and their theoretical motivation.

- Alcohol licenses include all licenses in D.C. reported in October 2010. These include both locations that are only intended for the alcohol to be drunk off-site (e.g. liquor and convenience stores) or on-site (e.g. bars, restaurants, clubs).

- 311 calls for service related to garbage on the street (Detritus), and infrastructure complaints (e.g. pothole, broken sidewalk) are included as measures of broken windows. Registered vacant houses and toxic release sites are included as measures of physical deterioration as well (Taylor, 2001).

- Green sites, which include community gardens and rain collector barrels, were included as a measure of collective efficacy (Sampson et al., 1997). Both signify an investment in the community by its constituents, and community gardens facilitate informal interactions that help to build social capital (Branas et al., 2011; Eizenberg, 2013; Kuo and Sullivan, 2001).

- Sidewalk Cafes and wireless hot spots as measures of gentrification (Papachristos et al., 2011).

- Parks, trees (that are alive) and street lights are included as measures related to visibility (Donovan and Prestemon, 2012; Farrington and Welsh, 2004; Groff and McCord, 2012).

- Halfway houses and HIV Clinics that may be crime attractors (Groff and Lockwood, 2014).

- The area of sidewalks, roads, whether the street unit is an intersection, and the area of the Thiessen polygon for a street unit are measures of the built environment. These may relate to either the accessibility (Johnson and Bowers, 2010) or the general nature of that place. E.g. a location with no sidewalks is unlikely to have a sizable walking around population, and so may have very few exposed persons to be the victim of a crime.

- Areas that are crime generators; such as public transportation areas (e.g. bus stops and subway entrances) (Levine et al., 1986), schools (Roncek and Lobosco, 1983), recreational areas (mainly basketball and tennis courts), public housing (Roncek et al., 1981), university areas, hospitals, places of worship (Beyerlein and Hipp, 2005; Desmond et al., 2010), shopping malls, and libraries (Eck et al., 2007). Police stations are also included, although they may be expected to be a guardian preventing crimes.

A notable absence from this table are measures related to the residential demographics. In Chapter 9 I detail how using fixed effects can control for omitted neighborhood level variables, and use several different definitions of neighborhoods to show that the local effects of bars and 311 calls for service are not confounded with omitted variables. While this is different than much of the macro criminological work, the model specifications used here are not wildly different from others using small units of analysis. Typically there is a trade-off when using small units, as fewer census variables are available and would require allocation to the smaller units, such as dasymetric mapping (Poulsen and Kennedy, 2004). Local institutions become more predictive of increased levels of crime at small units, so models often incorporate many of these local crime generator locations, see Bernasco and Block (2010) for one recent example. Instead of interpolating census measures to small units, which potentially introduces measurement error, I use the fixed effects analysis as a robustness check to make sure the inferences I make are not confounded by omitted neighborhood variables.

6.8 Describing spatial weights

Based on the Thiessen polygons previously discussed, one can create spatial weights matrices based on contiguity of those polygons. That is two points are neighbors if any part of their Thiessen polygon touch one another (queen contiguity). Figure 9 shows an example of defining a set of neighbors based on this criterion.

The weights matrix is arbitrary, so one may consider alternative neighbor matrices. For one particular example, Elizabeth Groff has recently suggested to incorporate spatial weights based on connectivity of the street network (Groff, 2011, 2013). Other potential weights matrices may be based on a certain number of nearest neighbors, inverse distance weighting, or neighbors within a buffer (or combinations of any of these). Fortunately the correlation of *any* two continuous measures pre-multiplied by different binary spatial weights matrix is *only* a function of the shared number of neighbors in the two matrices (Lesage and Pace, 2010). So in the end most spatial weights matrices will produce similar results, as most reasonable choices in the matrices will produce similar sets of neighbors. Here the binary contiguity weights matrix is chosen partly out of convenience; criminological theory is not precise enough to theoretically choose between the different options I previously mentioned. Given the large number of observations (21,506) choosing a contiguity based matrix (as opposed to an inverse distance matrix) provides easier computations of neighbor statistics as well as makes the weights matrix sparse.

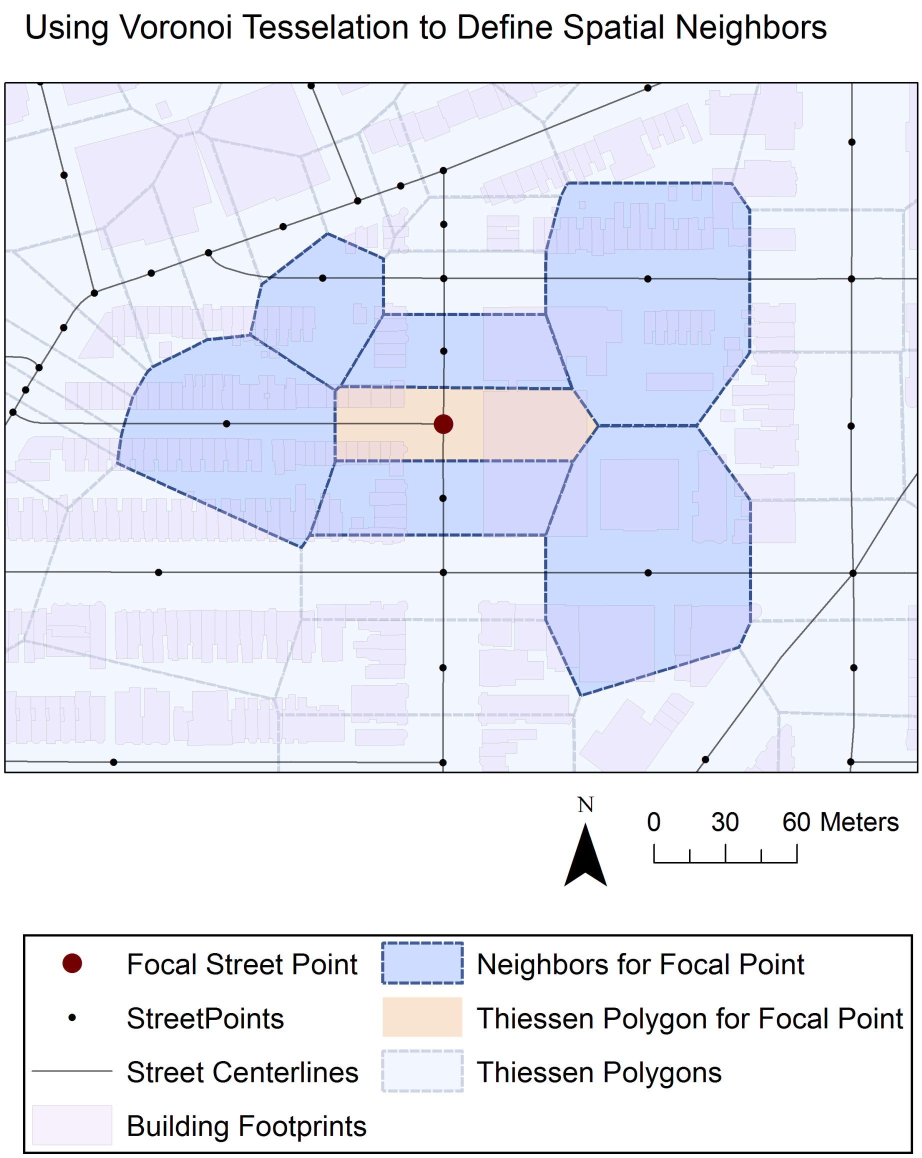


Figure 9: This figure shows how spatial neighbors are defined using queen contiguity of the Thiessen polygons (also known as the voronoi tessellation) of the points. This chooses the points that, given their voronoi tessellation, touch each others borders.

6.9 Conclusion

This chapter of the dissertation describes the construction of units of analysis, and why including intersections as a spatial unit is necessary. It also discusses the independent variables that will be used that are publicly available from dc.gov, as well as how a contiguity based spatial weights matrix is constructed from point data.

Subsequent analysis will present a set of simulated results illustrating how some of the aggregation bias that was discussed in prior chapters can occur. Then later chapters go on to present some observational data analysis of the effects of bars and 311 calls for service on crime, and then attempt to fit a more general model of crime.

Chapter 7

Simulation of a Spatial Process

In this chapter I will generate a very simple simulated dataset that will show the problems with interpreting aggregate correlations when the data are known to come from a micro level spatial process. The spatial data used will be the same D.C. data as has been described in Chapter 6, and use the same definitions of neighborhoods.

A set of data was simulated with the following specifications:

$$y = -8 + 2(x) + 0.5(Wx) + e \tag{7.1}$$

Where -8 is a constant intercept, x is normally distributed with a mean of 5 and a variance of 1, and Wx is a binary contiguity spatial weights matrix post multiplied by x (note W is not row normalized). W is defined by the Thiessen polygon of street units for Washington D.C. The error term e is normally distributed with a mean of 0 and a variance of 9.

The point of this particular specification is to show that a set of data with a spatial term,

Table 7: Correlations when aggregating to neighborhoods

	β	S.E.	R^2
Grid Cells	3.3	0.4	0.10
Block Group	4.4	0.2	0.45
Census Tract	4.8	0.3	0.66

here $0.5(Wx)$, when *presuming* a different type of model will cause spurious conclusions. Here x is generated without reference to space, so x and Wx are uncorrelated.

The first misspecified models I will consider are of the form:

$$\bar{Y} = \beta_0 + \beta_1(\bar{X}) + e \tag{7.2}$$

Where the $\bar{Y}$ and $\bar{X}$ terms are mean values aggregated up to neighborhoods. So these results will show what happens when you aggregate up the micro level equations. Table 7 displays these results for three different types of aggregations; aggregating to a regular grid of 500 meters square laid over top of D.C. (Grid Cells), block groups and census tracts. To identify which block group or tract a particular street unit falls within (as I previously discussed they are often times on the border) I just treat them where they lay. That is, small differences in the resolution of the geographic data files will make it seem like the street unit falls on a particular side of a border, but in reality this identification is only due to geographic error. Of the few street units that do fall directly on a border, I assign them to the census geography that has the nearest centroid.

If one interpreted the coefficient estimates as the local effect (e.g. making the ecological fallacy) all three spatial aggregations over estimate the local effect of x on y, and the bias

Table 8: Correlations when aggregating to neighborhoods based on Observed Variables

Variable	Aggregation	β	S.E.	R^2
Bar	Grid	4.72	0.26	0.33
	Block Group	4.63	0.13	0.75
	Tract	5.26	0.18	0.83
Detritus 311	Grid	4.99	0.03	0.97
	Block Group	4.94	0.03	0.98
	Tract	5.22	0.03	0.99
Infrastructure 311	Grid	4.54	0.07	0.88
	Block Group	4.71	0.05	0.95
	Tract	5.03	0.06	0.98

gets worse for larger aggregations. When one aggregates up, it will combine the covariances of spatially near observations, so what we see in the aggregate equations is the confounding of the local and spatial effects.

To show that this results holds when using actual independent variables, I simulated the same outcome but with using bars and 311 calls for service (both infrastructure and detritus related calls). Table 8 shows these results, which are largely comparable to the simulated variable. The higher R^2 value for 311 calls for service results from the fact that they have a much larger variance to begin with than the simulated variables, so the error term is relatively much smaller. The same trends hold though, in that the effects confound local and spatial, with larger effects at larger aggregations, and the R^2 values increase for larger aggregations.

One particular point I would like to make that is misleading in prior criminological literature is that larger correlations (or regression coefficients) for larger aggregate units **does not make the larger unit more theoretically relevant** to the data generating process (Hipp, 2007; Pratt and Cullen, 2005). Here *we know* the data generating process is a micro one because I artificially created the data to be that way. Larger aggregations

Table 9: Contextual Effects Models

	True Model		Grid		Block Group		Tract	
	β	S.E.	β	S.E.	β	S.E.	β	S.E.
x	2.0	0.02	1.9	0.04	2.0	0.04	2.0	0.04
Wx	0.5	<0.01						
$\bar{X}$			2.1	0.22	1.7	0.26	2.2	0.40
Inter.	-7.9	0.13	-2.1	0.22	-1.3	1.30	-4.0	2.00
R^2		0.73		0.12		0.12		0.12

provide larger effects (and larger correlations, as can be surmised from the R^2 values in the tables) but the larger aggregations *are further* from the true model I made. This also makes comparing R^2 values from studies with different aggregations nonsensical as R^2 is not invariant to the spatial aggregation (Weisburd and Piquero, 2008).[1] One can simply aggregate up to make it appear as if the model is doing a better job of explaining variation in y, when again in reality you are travelling further away from a realistic model of how the data were generated. This is not new, and are the same findings and warnings Robinson (1950) gave over 60 years ago.

The second set of models estimates a contextual effects model of the form:

$$y = \beta_0 + \beta_1(x) + \beta_2(\bar{X}) + e \tag{7.3}$$

Where $\bar{X}$ is the mean value of x aggregated again to several different definitions of neighborhoods. Table 9 provides these results.

In this example simulation run, the observed coefficients for x and Wx are almost exactly

[1] Weisburd and Piquero (2008) do discuss how different units of analysis have different average R^2 values (see their Table 2). Higher R^2 again does not indicate that the models at higher levels of aggregation are better specified or privy one to make correct inferences.

(within rounding) to the actual simulated data model. The intercept is also within one standard error. Because x was generated randomly, one can see that the local x effect is very close to the value of 2 in the subsequent equations, without regard to the aggregate unit used for $\bar{X}$. This is because x is independent and identically distributed, but if x were generated with spatial auto-correlation this would not be the case, and the local x effect would be biased as well.

Here even with x not having any spatial auto-correlation, the contextual effects estimated for $\bar{X}$ are quite different than the spatial effects that the data are generated under the true model. This is to show that even if you put $\bar{X}$ on the right hand side of a regression equation, this does not validate a neighborhood or contextual effects model. The effect of $\bar{X}$ can be generated by a micro level process at a lower level. If you assume a neighborhood exists you are likely to find a contextual level effect at whatever aggregate unit you choose. This does not rule out a micro level spatial process generating the data.

Again because x and Wx are independent here, you can think of Wx or $\bar{X}$ as representing any process that you are interested in. If you only have measures pre-aggregated up the neighborhood level, you will *never* be able to verify if a micro-level model is generating the data. Even with the micro level data, the contextual effect of $\bar{X}$ is still subject to the ecological fallacy if the micro equation is misspecified (Hauser, 1970). Similarly if the local (and spatial) effects of x are not modelled correctly in the micro level equation the effect of $\bar{X}$ is potentially subject to the individualistic (or atomistic) fallacy (Dogan and Rokkan, 1969). That is you may infer a neighborhood effect, but the data generating process is really only at the local level or is a misspecified spatial effect.

Subsequent chapters will build on these simulated examples to show what one can learn from small units of analysis, and identify if a neighborhood level process adequately describes crime data in D.C. on street units compared to a micro level spatial process.

Chapter 8

Local and Spatial Effects of Bars on Crime

The motivation for this particular chapter, separately estimating the local and spatial effects of bars on crime, is twofold. The first motivation is to supply a realistic example where criminologists may be interested in separate estimates of local and spatial effects. The second is the pertinence the particular example has to both criminological understanding of the effects of bars on crime as well as policy on how bars are licensed.

The relationship between bars and crime is a popular subject in both criminology and epidemiology. Generically, the specific relationship between elevated occurrences of crime in places with more bars is linked to the following observations:

- Inebriated individuals are more vulnerable to being targets of victimization, as well as losing inhibitions that may prevent them from committing particular crimes.
- Bars are locations in which many people congregate in a small space.

- Some bars attract motivated offenders.

Various more in depth theoretical perspectives exist, and the majority of literature is based on examining bars at aggregate neighborhood levels. This research typically specifies some additive structure between the number of bars in the neighborhood and the crime rate (Gruenewald et al., 2006; Lipton et al., 2013; Livingston, 2008; Pridemore and Grubesic, 2012, 2013; White et al., 2012; Zhu et al., 2004). The distinction between local and spatial effects of bars (Murray and Roncek, 2008; Roncek and Maier, 1991), and in particular whether bars within closer proximity to one another have larger than non-additive effects, could have various policy implications for zoning of drinking establishments (Jennings et al., 2014).

The relationship between bars and crime is well established, but one question of the current findings is whether the relationship is spurious because bars are not randomly placed in the built environment. It is plausible that bars self select into criminogenic neighborhoods (Block and Block, 1995; Briscoe and Donnell, 2003; Lugo, 2008; Treno et al., 2008) and so correlations between bars and crime may be a spurious association. I take this to be the biggest difficulty in establishing the relationship between bars and crime, and thus spend particular effort to provide evidence whether that is the case where possible.

Because the relationship between bars and crime is not in doubt, I take the added benefit of such an analysis to be the actual estimate of the local and spatial effects of bars on crime - and its subsequent pertinence to policy. The preponderance of criminological work that examines whether bars have spatial effects only attempts to estimate a uniformly radial effect around bars (Groff, 2011; Murray and Roncek, 2008; Ratcliffe, 2012), and here I am suggesting to *separately* estimate a local and spatial effect of bars. This allows one to estimate an interaction effect between the two, which is what Rossmo (1995) refers to when he describes *potentiation* of multiple bars within close proximity. It also allows one to describe the size of the spatial diffusion, which past literature is not able to comment on.

Again this has particular pertinence for urban planners that supply liquor licenses for drinking establishments. If bars being close to one another increase crime by a multiplicative factor, it may suggest particular actions by bars to mitigate crime caused by the interaction of patrons should be taken (Felson et al., 1997). Or it may suggest licensing of drinking establishments should be limited in areas in which another bar(s) already exists (Ahern et al., 2013).

So based on the prior literature and motivating question of the analysis, the following hypotheses are presented and examined in this chapter:

Hypothesis: Bars will have separate local and spatial effects on crime. A secondary hypotheses is that local and spatially proximate bars have interaction effects on crime.

The following section goes into greater detail about how bars are measured. Following that section empirical analysis of the relationship between bars and crimes is presented. To end the chapter some critical assessment of the model presented is given, suggesting that the effect estimates presented should only be critically accepted. There are several potential biases that appear to be evident in the data that impede interpreting the effect estimates as being reasonable estimates of the causal effect of bars on crime.

8.1 Measures of Bars

Here I generically use the term bars to refer to any location that has an alcohol license. The locations of alcohol licenses in Washington D.C. for this analysis were as of 9/29/2010, so are reasonable to estimate the correlation with crime in 2011. Alcohol licenses provided by DC.gov differentiate between off-premise retailers, such as liquor stores (license type A), grocery stores (license type B), or on-premise retailers (license type C). Type C license include places such as taverns, restaurants, nightclubs, hotels or multi-purpose vendors (e.g. a sports stadium).[1] Licenses for 58 caterers and wholesalers were eliminated from this analysis, and this resulted in a total of 1,549 separate licenses across the entire city. Of those 211 were liquor stores, 267 were grocery stores, and 1,071 were on-premise retailers. The majority of on-premise establishments were classified as restaurants (691) and the next highest were taverns (183). All other types (hotel, nightclub and multi-purpose) have fewer than 100 licenses each citywide. Restaurants are classified based on the proportion of sales that are for foodstuffs. The distinction between taverns and nightclubs is based on the amount of floorspace devoted to a dance floor and nightclubs are required to charge a cover during certain times and events.

Figure 10 displays the spatial locations of those alcohol licenses in Washington D.C. Although some work has suggested that liquor licenses, especially off-premise licenses, tend to concentrate in minority neighborhoods (Franklin et al., 2010), this does not appear to be a reflective description of the predominant pattern of the locations of alcohol licenses. The vast

[1]See `http://abra.dc.gov/page/abc-license-types-and-classes` for a description of the D.C. liquor license types. Last accessed on 4/13/2014. On premise license types also distinguish between whether they allow for sale of liquor or only allow sale of beer or wine. The vast majority of on premise licenses permit sale of liquor.

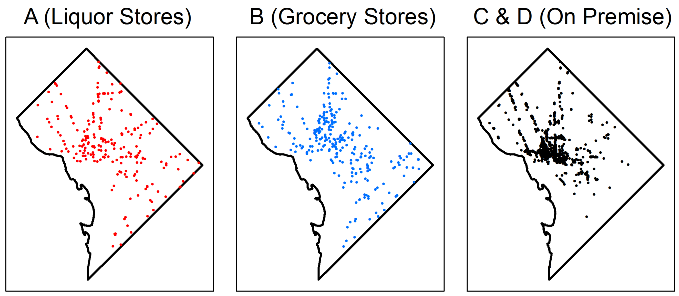


Figure 10: Location of liquor licenses by type in Washington D.C.

majority of licenses are concentrated in the central portion of D.C., which is more reasonably described as a mainly non-residential business district. Figure 11 displays the proportion of non-white residents and the proportion of female headed households, two typical demographic indicators predicting crime in neighborhood level criminological research. Blocks with fewer households are drawn at a higher level of transparency to reflect the underlying uncertainty in the estimates, creating a value by alpha map (Roth et al., 2010). Each of the distributions are classified for the choropleth maps by using quantiles to ease comparisons between the two maps (Brewer and Pickle, 2002). The swaths of white in the map either represent large parks or rivers where there is no residential population. Places with lower percentages of white residents and higher percentages of female headed households tend to be in locations south of the Anacostia river, and to a lesser extent the north-eastern part of the city. These are far away from the densest clusters of alcohol licenses that are mainly clustered in the area to the north of the national mall that has many fewer residents.

Block and Block (1995) suggests that liquor stores can serve as similar places of congregation as do actual bars. From the onset I decided to base the analysis here on all alcohol licenses without differentiating between them as the marginal relationship between the three

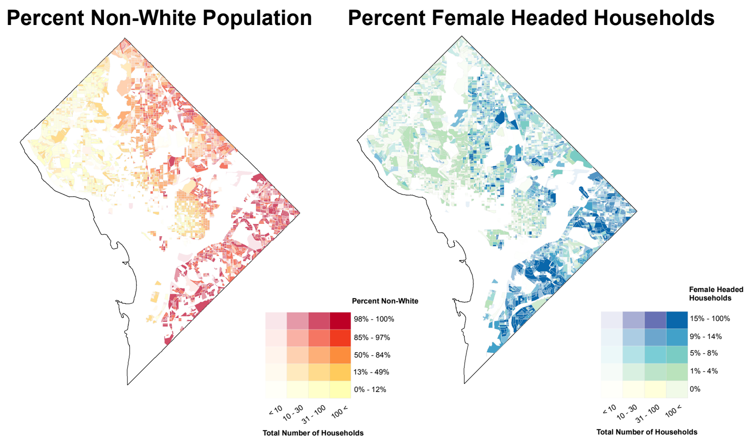


Figure 11: Demographic characteristics for D.C. census blocks. The map on the left displays the percentage of non-white population per block, and the right map displays the percentage of female headed households per block. Information taken from the 2010 census. Colors are moderated by the total number of households within the block by making blocks with a smaller number of households more transparent (Roth et al., 2010). Each color ramp is specified by the quantiles of the respective distribution, so the color gradients are quintile ranges of the data.

types (grocery, liquor store and on-premise) and crime does not appear to be great.[2] Figure 12 shows the mean number of crimes for street units with no alcohol license compared to streets with *only* one license for each type, grocery stores, liquor stores or on-premise locations. The error bars signify plus or minus two standard errors for the mean estimate, and the label signifies the total number of street units that fall under that category. The error bar for street units with no license is so small it is entirely covered by the point representing the mean number of crimes.

There is certainly an increase between street units with no license (around a mean of 1) to street units with at least 1 license (around a mean of 3 to 4). But there is not a

[2]Findings are mixed as to the difference between on-premise and off-premise licenses (Pridemore and Grubesic, 2013). One other study for Washington D.C. found very similar effects between on premise and off-premise licenses (Franklin et al., 2010), White et al. (2012) find similar effects for a sample in Norfolk, VA, and Jennings et al. (2014) find similar effects for a sample in Baltimore, MD. Others though have found differential effects for different license types (Lipton et al., 2013) and interaction effects between license types and other demographic characteristics (Livingston, 2008).

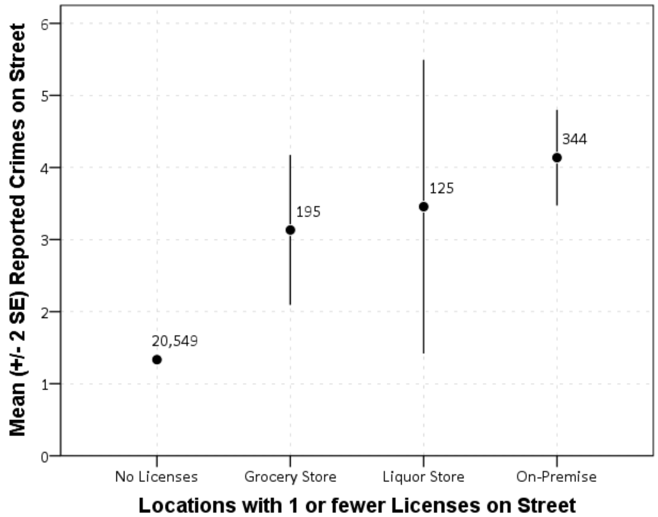


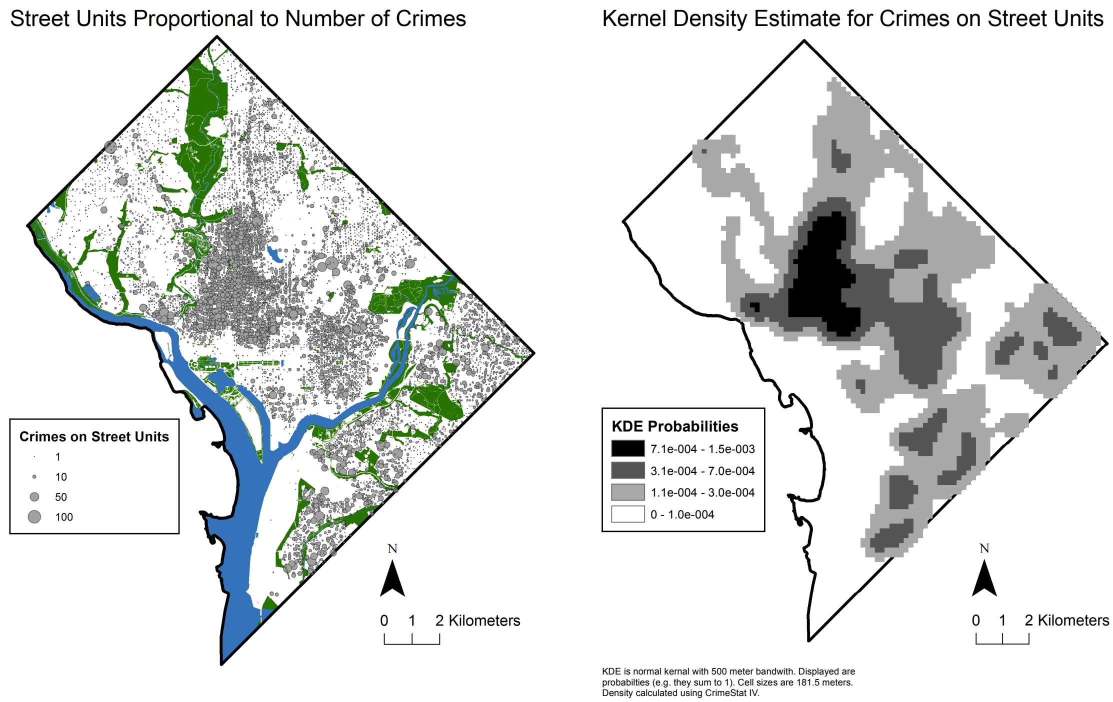
Figure 12: The mean number of crimes for street units with no licenses or only one liquor license for a grocery store, liquor store, or on-premise retailer (e.g. restaurant, bar). The text labels show the number of street units in that particular category, and the error bars display the standard error of the mean number of crimes plus and minus 2. Note the total number of observations does not sum to 21,506, as some streets can have more than one license and are excluded from this plot.

large difference that discriminates the means for grocery stores, liquor stores or on-premise retailers. One can further disaggregate the license types later on if one feels the need or want to. The aggregate effect is a function of the disaggregated effects (same as with aggregate spatial data) and so finding an effect for all license types is not invalidated even if there are differing effects between the different licenses, as they all likely increase crime.

The subsequent sections presents some bivariate analysis of the relationship between bars and crime.

8.2 Empirical Analysis

As expected, crime tends to concentrate in particular areas of Washington D.C. Figure 13 displays a proportional dot map on the left, where larger dots represent more crimes on that street unit. Given the extreme overplotting, I estimate a kernel density of the crimes from



(a) Crimes by Street Unit

(b) Kernel Density of Crimes by Street Unit

Figure 13: All 32,440 crimes for the 21,506 street units. The map on the left displays a proportional dot map of the crimes, with larger dots for more crimes. The map on the right displays a kernel density estimate of the crimes. The kernel is normal with a fixed bandwidth of 500 meters. Displayed are probabilities that sum to one (so an actual density).

the weighted street unit data to more clearly show the peaks and valleys of the density of crime in Washington D.C.

An alternative, non-spatial display of such clustering is to plot the empirical cumulative distribution function of the number of crimes on individual street units. Figure 14 displays the cumulative proportion of crimes, sorting the data values so the streets with the highest number of crimes are counted first and street units with zero crimes are counted last. In Figure 14 we can clearly see the Pareto principle in action, with twenty percent of the street units having over 80 percent of the total number of crimes in this sample. You can also see from this plot that all crimes are recorded on just under 40 percent of the street units. Over

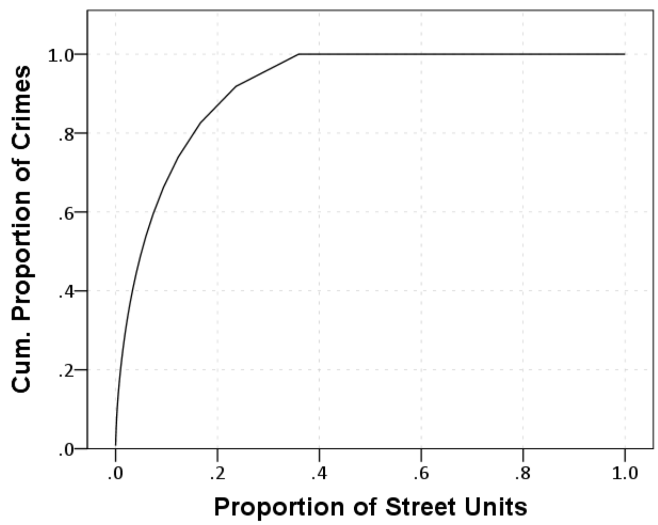


Figure 14: Empirical cumulative distribution function of crimes on street units. The X axis displays the proportion of street units, and the Y axis displays the cumulative proportion of crimes from the sorted distribution. Consistent with other crime literature and the Pareto principle, over 80% of the crimes occur on 20% of the street units. Over 60% of the streets unit have zero reported crimes, as can be seen by the flat line at the top of the graph.

60 percent of the street units do not have a single crime attributed to them.

So what of the bivariate relationship between bars and crime? We have already established that streets with at least one alcohol license in Figure 12 have a larger number of crimes than do streets with no license, but here we are primarily interested in whether *more* bars in close proximity result in more crime. That is, if city planners license a bar in a location proximate to other bars, is the effect on crime larger than if the bar was licensed in an area where there were no other bars proximate? Figures 15 and 16 attempt to answer that question.

Figure 15 displays on the Y axis the mean of crimes on street units with a particular number of alcohol licenses on the local street (left panel) or the neighboring street (right panel). You can see from this chart that the number of crimes increase with added bars both on the local street unit and in the neighboring street units. To get a better sense of the relative sizes of these two effects, the two mean effects are superimposed on the same graph in Figure 16. Here the red line corresponds to the local effect and the black line corresponds to the effect of the sum of bars in the neighboring street units. Contrary to the potentiation hypothesis, the effects appear to be linear and additive for both local bars and neighboring bars. There is only a limited window with which to address such differences though; Figure 15 shows the number of street unit observations that encompass the higher numbers of bars. While there are a few street units with more than 5 licenses on the local street (22 in total), the vast majority of street units have fewer than 2.

There is slightly more variation in the number of licenses on neighboring streets, but again the effect appears to be additive. The slope for the local effect appears to be around 2 extra crimes per added bar, and the slope for the neighboring number of bars appears to be less than 1 added crime per added bar. This decay of the spatial effect makes theoretical sense as you would expect a bar farther away to have less of an effect on crime than a bar closer in proximity.

The following section fits a multiple Poisson regression equation predicting the number of crimes that are on a street unit.

8.2.1 Model fitting

As from examining the CDF of the crimes on street units, the variance of the distribution is clearly zero inflated. The mean number of crimes on a street unit is 1.51, and given this value a Poisson distribution would only be expected to have have a density of 22% for zero observations. Because of this zero inflation I consider a negative binomial regression model

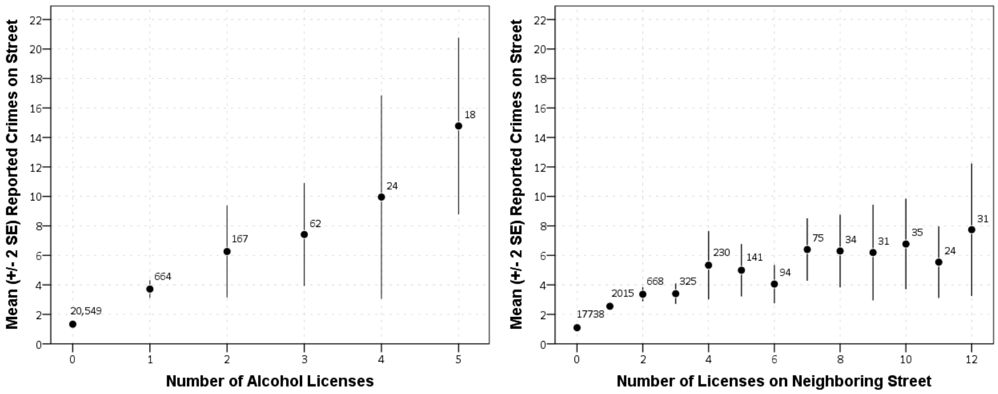


Figure 15: In the left panel this plot shows the mean number of crimes (as a point) and ±2 standard errors (as a vertical bar) on street units with 0 to 5 licenses on the *local street*, and in the right panel 0 to 12 licenses on the *neighboring* street in black. One can see that the number of bars on the local street has a stronger effect on the number of crimes than do the number of bars on the neighboring street. The labels show the number of street units that fall within that category of number of licenses. Note these aggregated estimates are not mutually exclusive of streets, i.e. one street can be both within 1 license on the local street and 1 license on the neighboring street. There are a total number of 21,506 street units, so you can add up the labels and subtract that from 21,506 to determine the number of street units not displayed on these graphs.

(Berk and MacDonald, 2008; Osgood, 2000), where the logarithm of the expected value of crime is conditional on a set of covariates.[3] This model is represented in Equation 8.1.

$$\log\left(\mathbb{E}\left[\text{Crime}\right]\right) = \beta_0 + \beta_1 \text{Bars}_L + \beta_2 \text{Bars}_N + \beta_3 \left(\text{Bars}_L \cdot \text{Bars}_N\right) + \beta_4 \text{Intersection} + f(\log(\text{Area})) \quad (8.1)$$

Where the items in the model correspond to:

- β_0 is the intercept

[3]Graphs later on will show that zero inflated models are not necessary, the predicted number of crimes given the negative binomial model matches very closely to the observed density. Even later on for models of a subset of burglaries in which around 88% of the street units had zero observations the negative binomial model fit quite well, suggesting no empirical need for zero inflated or hurdle models. Subsequently these models are not further considered, as I see no theoretical motivation for the more complicated zero inflated or hurdle models.

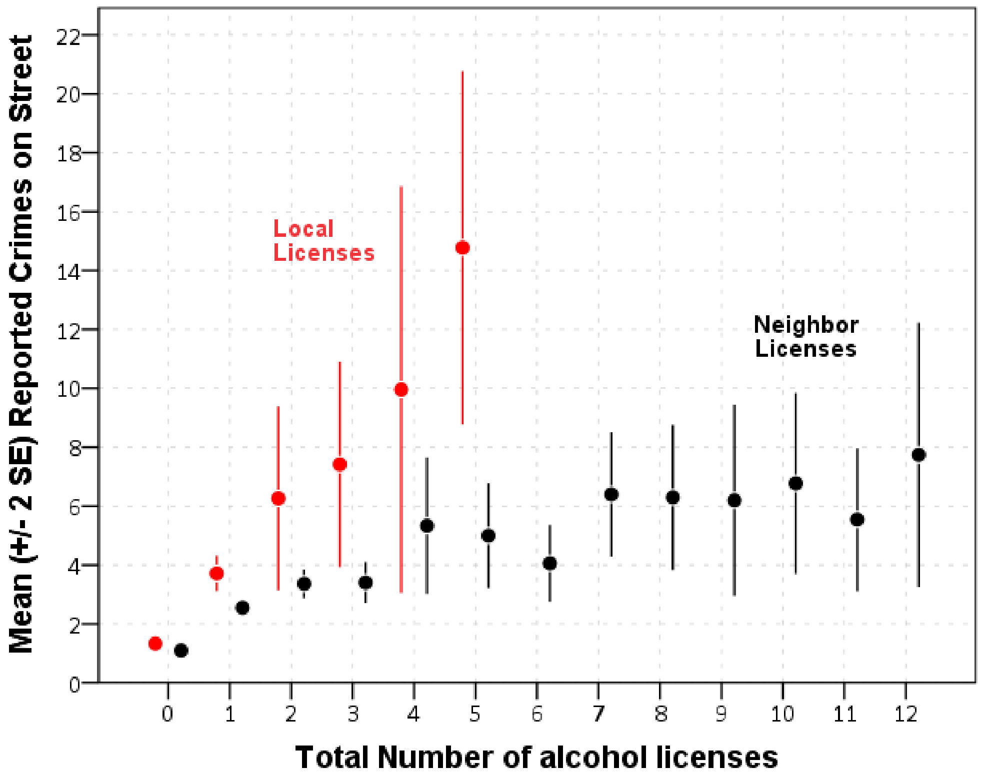


Figure 16: This plot shows the mean number of crimes (as a point) and ± 2 standard errors (as a vertical bar) on street units with 0 to 5 licenses on the local street in red, and 0 to 12 licenses on the neighboring street in black. One can see that the number of bars on the local street has a stronger effect on the number of crimes than do the number of bars on the neighboring street.

- Bars_L is the number of local bars on the street unit
- Bars_N is the sum of bars on neighboring street units
- $(\text{Bars}_L \cdot \text{Bars}_N)$ is the interaction between local bars and neighboring bars
- Intersection is a dummy variable equal to 1 when the street unit is an intersection
- $f(\log(\text{Area}))$ is a set of restricted cubic spline basis for the log of the area of the Thiessen polygon.

The model as specified is the same as the Poisson regression model, but the negative binomial model includes a separate dispersion term, a, that inflates the variance of the estimate and spreads out the predicted density (Long, 1997). If one sums the covariates on the right hand side of Equation 8.1 and then exponentiates them, this is then the conditional mean estimate. This is the first line in Equation 8.2. The variance is then a function of the dispersion parameter a and the conditional expectation μ as shown by the second line. This is typically called the NB2 model (Long and Freese, 2006). Here you can see that the case of the dispersion parameter being equal to 0 is the usual Poisson case (i.e. the variance of the Poisson is the same as its mean). When a is larger than 0 the variance is larger than the mean. Again this does not change the mean equation or the parameter estimates (Gelman and Hill, 2007), but it does increase the standard errors for the estimated coefficients in the model.

$$\mathbb{E}[Y|X] = e^{(\beta_k X_k)} = \mu \tag{8.2}$$

$$\mathbb{V}\mathrm{ar}[Y|X] = \mu(1 + a \cdot \mu) \tag{8.3}$$

In this model since the link function is the logarithm it forces the effect of bars on the expected number of crimes to be multiplicative, so any positive effect for the number of bars on crime by necessity implies the potentiation effect previously discussed. Besides the number of local bars, neighboring bars, and their interaction effect the model includes terms for whether a street unit is an intersection and a set of restricted cubic spline (RCS) basis for the natural logarithm of the area of Thiessen polygon (in square meters). Intersections are included for reasons already discussed in Chapter 6, in that crimes are more likely to be recorded at specific addresses. A non-linear term for the area of the Thiessen polygon is included because areas with very small Thiessen polygons are likely to be in dense clusters of street units, making them less likely to have a crime associated with them (e.g. competition between nearby street units). For street units with large areas though, this is indicative of no nearby street units, and this is more likely to be associated with fewer crimes, as they are areas near the border of the city (or on major thoroughfares with no intersecting streets). These do not tend to be areas of high human traffic conducive to being victimized.

When estimating RCS in a regression model one also includes the original value of log(Area) in addition to the $K - 2$ basis terms. The model is then restricted to be linear in the locations beyond the outermost knots, but can be approximated by cubic terms between knot locations. Also the first and second derivatives of the model are defined ev-

Table 10: Knot locations for log(Area)

Knot	log(Area)	**Quantile**
t_1	6.76	0.25
t_2	7.87	.1833
t_3	8.23	.3417
t_4	8.50	.5
t_5	8.77	.6583
t_6	9.16	.8167
t_7	10.34	.975

erywhere, so the function is smooth and has no discontinuities at any location. The 7 knots for the splines are placed at the suggested locations per Harrell (2001) for specific quantiles for log(Area), and are contained in Table 10.[4] Using restricted cubic splines forces the predictions in the tails to be linear and only adds $K - 2$ terms into the model. Following the notation in Durrleman and Simon (1989), the knot locations are signified by $t_1 \ldots t_7$ and the new variables are specified as $K_1 \ldots K_5$. The spline basis that are estimated in the model are normalized (i.e. divided) by the square of the range between the first and last knot locations (e.g. $(10.34 - 6.76)^2 = 12.82$ for this model).

Table 11 displays the coefficient estimates for the model specified in Equation 8.1. Displayed are the linear predictors, the standard error of the linear predictors, and the 95% confidence intervals of the exponentiated coefficients (i.e. incident rate ratios) where reasonable. One can see that the effect of the local number of bars is 0.5, and the effect of the neighboring number of bars is 0.2. The interaction term is negative and is statistically significant at the .05 level (as is evident from the confidence interval not containing zero). Graphs later on showing effect estimates for reasonable ranges of the number of bars will show that this negative interaction term is so small it would only create a decrease in the ex-

[4] Typically the number of knots is more important than their exact location. Placing the knots at specific quantiles provides the most data to estimate the cubic terms in between the knot locations.

pected number of crimes when going far beyond typical values for the local and neighboring numbers of bars.

As expected a street unit being an intersection is associated with a decrease in the expected number of crimes on a street unit. Also as expected the dispersion parameter a is greater than 0. The non-linear effect of area is impossible and nonsensical to interpret in terms of the individual coefficient(s) for the spline terms, and so Figure 17 displays the predicted number of crimes (black line) and the 95% confidence interval setting the covariate values of Intersection equal to zero, one local bar and one neighbor bar, and varying the size of log(Area) between 0 and 12.83 (the range of the data). Between values of 0 and 6.76 one can see how the fit is forced to be linear and the confidence interval is very wide. Where the bulk of the data is, between values 6.76 and 10.34, one can see the hypothesis of a non-linear effect, increasing at first for small areas but then decreasing for larger areas is confirmed. Then when the estimates get much beyond 10 the predictions start to trend upward, but the standard errors begin to explode, making any sort of conclusion about the tails beyond a log area of 10 (approximately 22,000 square meters) unwarranted.

While the parameter estimates displayed in Table 11 on their face confirm our prior suspicions, that bars increase crime and bars farther away have decreasing effects, the linear effects displayed by the Poisson regression model do not directly address the most direct policy question of interest: If one bar is added to a particular street, how much does it increase crime? Unlike OLS regression, the derivatives of interest are conditional on other covariate values, so one number can not be used to evaluate the marginal effect of increasing one bar on crime. What we can do though is estimate that marginal effect given a reasonable set of covariate values - and this provides a more direct quantity of interest than can be

Table 11: Negative Binomial Model Predicting Crime at Street Units: Parameter estimates & 95% Confidence intervals for the exponentiated estimates

Variable	β	S.E.	CI_L	CI_H
Intercept	-2.01	1.03	0.02	1.00
Bars_L	0.50	0.06	1.48	1.85
Bars_N	0.22	0.01	1.22	1.28
$(\text{Bars}_L \cdot \text{Bars}_N)$	-0.03	< 0.01	0.96	0.98
Intersection	-0.44	0.04	0.59	0.70
log(Area)	0.18	0.14		
K_1	3.24	0.99		
K_2	-42.99	13.95		
K_3	121.48	38.40		
K_4	-136.37	39.71		
K_5	55.76	14.72		
a	3.69	0.06		

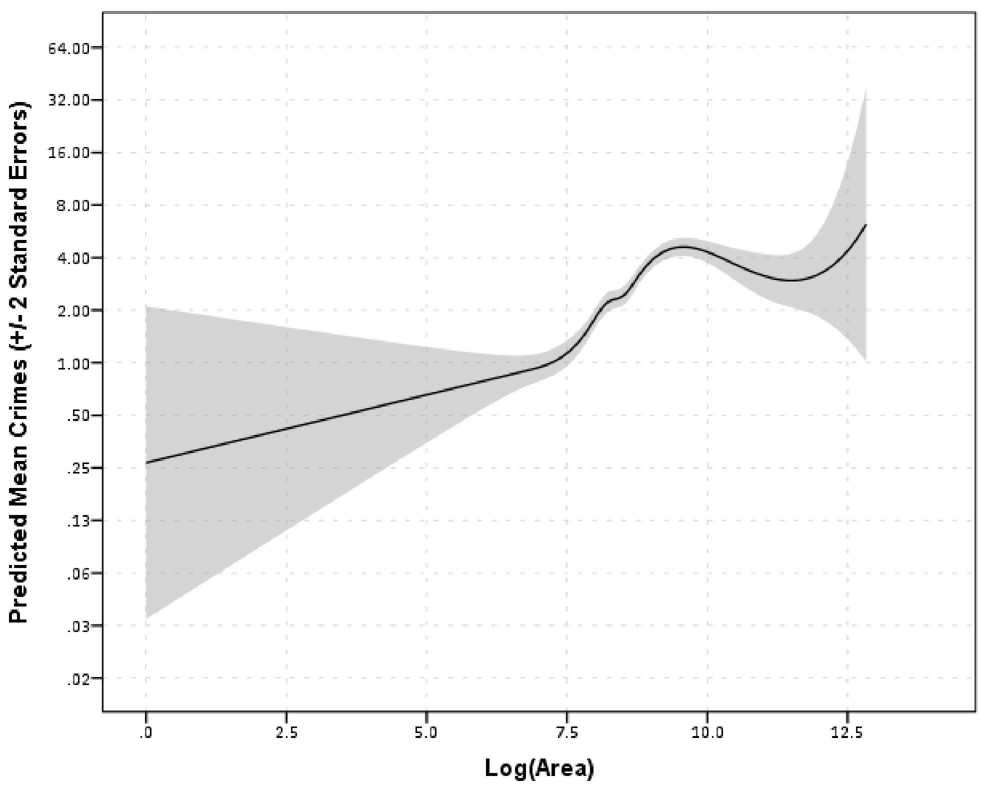


Figure 17: Non-linear effect estimate of log(Area) on the expected number of crimes per street unit. Other covariate values were fixed at a predicted values of Intersection = 0 and log(Area) = 8.5.

gleaned from either the linear predictors or the incident rate ratios listed in a table. Figure 18 displays the expected number of crimes for differing number of local bars on the X axis, and different lines correspond to increasing number of neighboring bars. The other covariate values are set to zero for Intersection and 8.5 (the median) for log(Area) (which corresponds to a RCS basis of .41, .02, .002, 0 and 0 for K_1 through K_5). Thus in terms of an equation, Figure 18 displays:[5]

$$\log\left(\mathbb{E}\left[\text{Crime}\right]\right) = \beta_1 \text{Bars}_L + \beta_2 \text{Bars}_N + \beta_3 \left(\text{Bars}_L \cdot \text{Bars}_N\right) + 0.23156 \tag{8.4}$$

$$\log\left(\mathbb{E}\left[\text{Crime}\right]\right) = \mu \tag{8.5}$$

$$\mathbb{E}\left[\text{Crime}\right] = e^{\mu} \tag{8.6}$$

Where the values of Bars_L and Bars_N are varied between 1 and 5. To estimate the change in the number of crimes on a street by increasing the local number of bars by a value of one, holding all other covariates constant, can be calculated by tracing up each line by one value on the X axis. For example, with one bar on the local street and one bar on the neighboring street, the expected number of crimes given this model is predicted to be 2.5. If you add one bar to the local street unit, the expected number of crimes is 3.9, resulting in a marginal increase of 1.4 expected crimes (remember the crimes being predicted are Part 1 offences for 2011). One can subsequently see the potentiation implied by this model in the upward sloping exponential curves. Adding one bar to a street unit that say already has 4 local bars

[5] The value of 0.23156 is calculated as follows: $-2.01 + 0.18(8.5) + 3.24(.41) + -42.99(.02) + 121.48(.002)$ where -2.01 is the intercept and the following terms are the effects for the log of the area and the associated RCS basis.

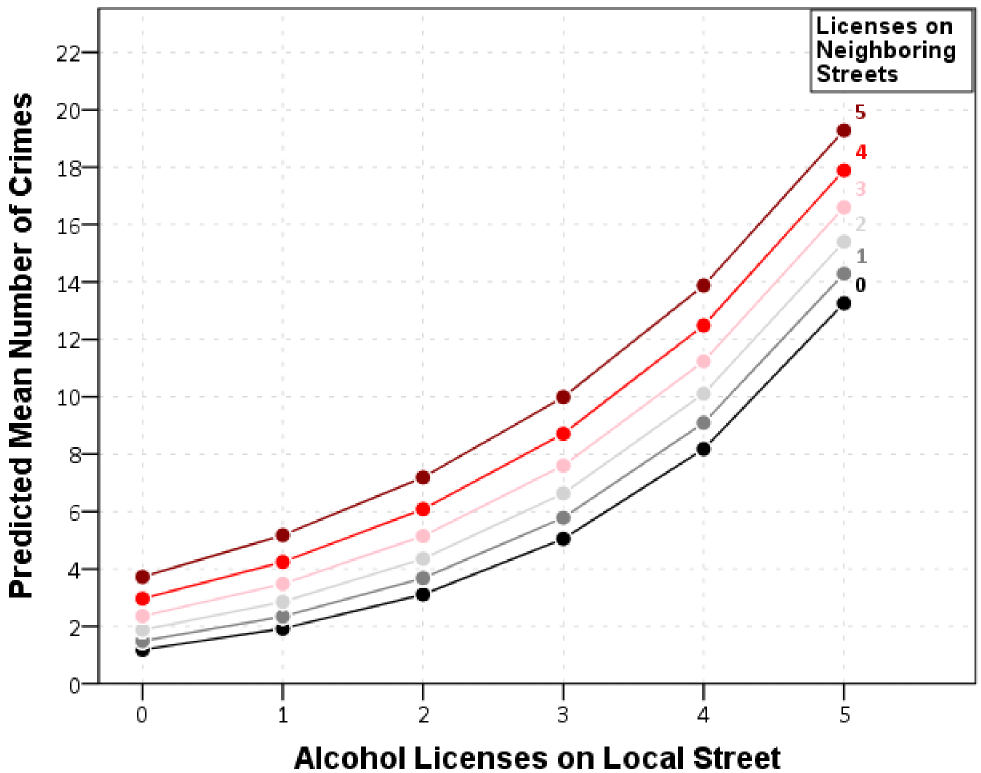


Figure 18: This graph displays the predicted values for crimes on a street unit, setting Intersection to 0 and log(Area) to 8.5. Tracing up a line will show the increase in the total number of crimes to the local street if you added a bar. Different lines correspond to differing numbers of bars on the neighbor streets.

and 1 neighboring bar increases the expected number of crimes from 8.2 to 13.1 crimes.

Given the discrepancies between various studies (units of analysis, different subsets of crime, measure of alcohol outlets) it is difficult to compare the effect size of bars on crime found here to other studies (Bushway et al., 2006). For example Franklin et al. (2010) in a similar negative binomial regression model predicting violent crime find the effect (in the form of an incident rate ratio) of alcohol outlets on crime at the census tract level in Washington, D.C. is 1.04 for the year of 2006. This is seemingly much smaller than the effects found here, both for the local and spatial effects.

But, the average number of violent crimes per census tract in Franklin et al. (2010)'s study was 26. So an increase in one bar in a census tract if the conditional expectation was originally 26 violent crimes is $26 \cdot 1.04 \approx 27$, an increase of only one violent crime. So in terms of reasonable marginal changes in the number of crimes given an increase of one bar the two findings do not appear to be terribly discrepant with one another. Comparing the linear models estimated in White et al. (2012) (predicting violent street crimes at the block group level for different years in Norfolk, VA), model estimates for alcohol licenses presented vary mainly between 1.5 and 3 (ignoring the endogenous spatial effect, which would make the marginal effect on crime *larger*). So on its face it is again a similar range of effects sizes, though a very different model is implied for the effect of bars on crime by estimating a linear model. In White et al. (2012) there is no additional effect for being in a location with more bars, so no matter the location adding one bar always increases the expected number of street crimes somewhere between 1.5 and 3 given the reported models. Jennings et al. (2014), in a study of violent crime at the census tract level for 5 years in Baltimore, find incident rate ratios around 1.02 (so appear similar in size to Franklin et al. (2010)), but the mean number of crimes per census tract is 259, so an increase of around 5 violent crimes is implied for a census tract starting with the mean number of crimes. Dividing this by 5 (so on a yearly basis) then gets us back to the yearly estimate of a marginal increase in 1 violent crime per year with the addition of 1 additional alcohol license. So upon perusing the literature the effects displayed in the model here appear to be slightly larger, but not drastically so, although quantitative comparisons between studies is made difficult by differences in aggregation and types of models estimated.

This ends up being only part of the marginal effect of interest. Having the neighboring number of bars in the model provides additional complexity, as one can not *only* increase the local number of bars by 1, as increasing the local number by 1 implies you increase the neighboring number for *multiple* neighboring observations. So, if you added a bar to *Street A*, it would increase Bars_L by one on Street A, but would also increase Bars_N by one on **multiple** neighboring streets (say Streets B, C and D).

A similar predicted effect chart is plotted in Figure 20, but this chart is different in two ways. First, it plots the marginal effect *increase* of adding a bar to a neighboring street, and then multiplies that effect by 3 to 10 (represented by different lines). I choose to multiply by 3 to 10 as these are the typical number of neighbors in the dataset, with a mean of 6. Three neighbors is the minimum in the dataset, and there are only a few locations with more than 10 neighbors (less than 300). The brown to blue-green color scheme for the plot is taken from ColorBrewer (Harrower and Brewer, 2003) and the more saturated colors are used for the more typical neighboring values from 5 to 8, de-emphasizing the lines for the number of neighbors that occur much less frequently.

To explain this plot I will use an additional example. For our Street A example and its three neighbors, B, C and D, lets take the original expected value for crime on B, C and D to be $\log(\mathbb{E}[\text{Crime}]) = 1$. Then we add one to the number of neighboring bars to each street, which would increase the effect by $\log(\mathbb{E}[\text{Crime}]) = 1 + 0.22 = 1.22$ for each individual street. (0.22 is the linear predictor of the logged expectation for the number of neighbors in the original regression equation.) So the increase in the expected number of crimes *for one neighboring street* is $e^{1.22} - e^{1} = 6.1$. Since this increase is diffused over three streets, the expected marginal increase in crime of placing a new bar on Street A to all of the neighboring streets is $3 \cdot 6.1 = 18.3$. If the original expected number of crimes on street A is the same as the neighboring streets, $\log(\mathbb{E}[\text{Crime}]) = 1$, the expected marginal increase for crimes on *only* Street A would be $e^{1.5} - e^{1} = 7.2$. So with only 3 neighbors (which is the minimum number of neighbors in the dataset), the diffusion effects are **much larger than the local effects when taking into account the typical amount of diffusion**.

To illustrate this, Figure 19 displays a hypothetical street layout. In this scenario, a bar is added to the street unit represented by the blue dot, and the model then implies crime increases at the blue dot *and all* of the neighboring orange dot street units. The areas of the circles represent the size of the crime increase at that location given a bar added to the blue dot street unit, and the blue dot indicates that crime on the local street increases by 4 times more than a location at *one* of the neighboring streets. The one blue dot and the one orange dot are what the model coefficients tell us directly. But, even though the impact of crime on the local street (i.e. at the blue dot) is larger than one orange dot, which demonstrates distance decay in the effect, summing the area of all the orange dots the total spatial effect *of adding a bar* is 1.5 times the area of the blue local effect.

The Y axis in Figure 20 plots these added spatial effects to the neighboring streets for the same fixed set of coefficients in Figure 18 and varies the number of bars on the local street on the X axis and has separate panels for the number of bars on the neighboring streets. The X axis displays the original number of bars on the local street, so the smallest added effect of one bar can be seen as having zero bars on either the local or neighboring streets (the top most left panel at the start of the lines). Having between 3 and 10 neighbors produces an added number of crimes between 1 and 3 for this particular set of circumstances, with only an increase in 1 crime for neighboring streets if there are only 3 neighbors, and an increase in 3 crimes if there are 10 neighbors.

The separate lines are then these added effects multiplied by 3 through 10. Since I used a

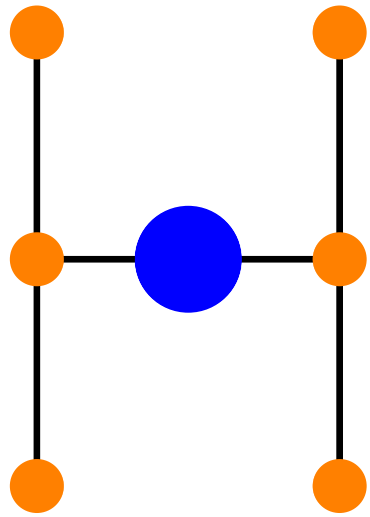


Figure 19: This diagram shows a hypothetical street layout, with the blue dot having a bar added to that street unit, and the orange dots the neighboring street units. The area of the dots are proportional to the amount of crime increase at that street unit, and the blue dot has four times the area of one orange dot. But when summing up the area of all 6 of the orange dots, the total spatial effect of all the orange dots is larger than the blue dot by 1.5.

spatial weights matrix that is not normalized to make the neighbor calculations, this diffusion effect is much plainer to calculate and graph. One could multiply the linear effect of 0.22 by the mean number of neighbors to make the effect sizes between the local and neighboring effects more comparable in the original table, but I prefer to plot the effect estimates directly over a reasonable set of covariate values.

A popular effect size metric is to calculate discrete measures of change, as I have given in these prior examples, for every individual observation in the sample. One then has a distribution of the expected increase in the number of crimes if you added a bar, one at a time, to every single street unit in the sample. I have compiled these statistics, breaking down the difference in the local effect, the neighboring effect, the sum of those two, and the difference (Neighbor minus Local). These are presented in Table 12.

Here we can see that the variance of the discrete change effects are very large, although it is partly a function of some great outliers in the data. The mean effects are less than

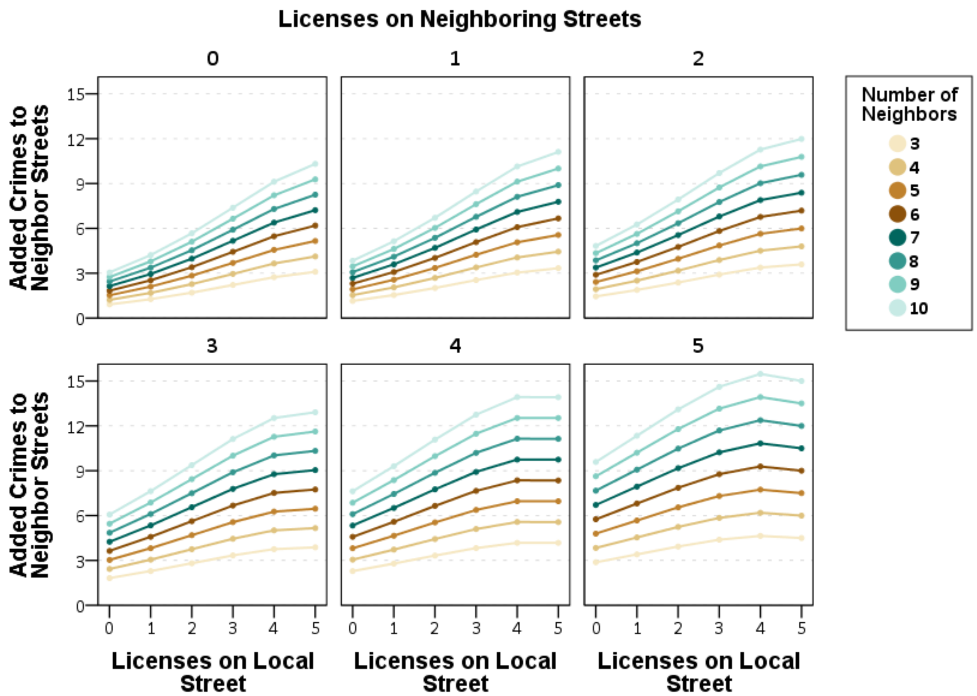


Figure 20: The graph displays the diffusion effects of bars on crime. The Y axis displays the added number of crimes to neighboring streets if one bar is added to the local street. Lines represented 3 through 10 signify that if a location has more neighbors it will diffuse the number of crimes to more locations.

Table 12: Average discrete change effects for adding one bar to every observation

Effect Type	**Mean**	**St.Dev.**	**Min.**	**Max.**	**25th**	**Median**	**75th**
Local Effect	0.80	12.41	-1669.48	610.34	0.48	0.76	1.11
Neighbor Effect	2.96	56.60	-1041.61	7846.26	0.88	1.71	2.95
Sum of Effects	3.76	44.77	-431.27	6176.78	1.35	2.47	4.12
Difference in Effects	2.17	68.63	-1651.95	9515.74	0.39	0.91	1.85

one crime for the local effect, almost three for the neighbor effect, and slightly short of four crimes for the total effect. The quantiles of the effects located in the right most columns provide a more reasonable spread of the effects, with a median discrete change effects of 0.76, 1.71 and 2.47 for local, neighbor and the total effects respectively. Figure 21 graphs the majority of these effects in a scatterplot (although note the range of the axes do cut off some of the outlying observations). On the Y axis is the neighbor effect, and on the X axis is the local effect. The red line signals the line of equality, and a point above the line signifies that the spatial effect is larger than the local effect. The vast majority of the point cloud can be seen to add less than 10 crimes to the neighboring streets and 5 crimes to the local street.

Although I did demonstrate distance decay in the effect, e.g. a bar further away would increase crime by less than a bar nearby, evaluating the *total* spatial effect as being nearly double the local effect is a surprising finding. It is difficult to relate this effect size to prior work, as those not using weights matrices that simply count up the number of bars on neighboring streets are near impossible to recreate guesses as to the size of the spatial effect based on summary statistics. One article that does have a similar model specification though is Murray and Roncek (2008). In that article Murray and Roncek (2008) estimate negative binomial regression models predicting assaults based on the number of bars at the block level, and include a measure of the number of bars on neighboring blocks in some models (e.g. the left most regression results in Table 3). In that model they find the local effect of bars (in the linear parameters) to be 0.64, and the spatial effect to be 0.18. These are very similar to the regression model results presented here, with the linear local and spatial effects of bars being 0.50 and 0.22 respectively. Because they also use queen contiguity to

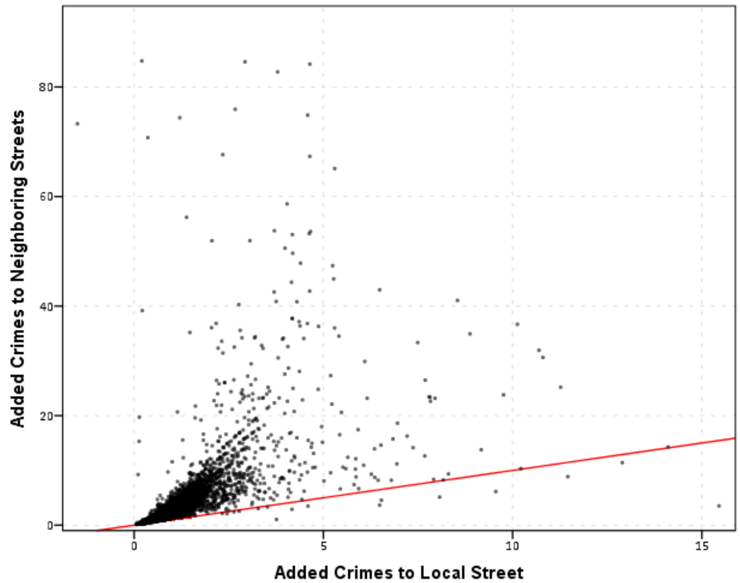


Figure 21: On the graph the effects of adding one bar to each individual street unit are calculated, and the spatial effect (Y axis) and the local effect (X axis) are decomposed. The red line signifies where the local effect wouold equal the spatial effect. In total, spatial effects tend to be much larger than the local effect.

define neighbors, it is very likely the typical number of neighbors is between 6 and 8 for their sample. If one multiples the spatial effect by 6, it would produce a total effect for the spatial neighbors twice that of the local effect, essentially synonymous with the results here.

Now that the model and marginal effects implied by it have been sufficiently explained, the following section will critically examine the model for flaws that would suggest the estimates are biased.

8.3 Checking the Model

The first check for Poisson regression models is typically to see if the predicted density from the model approximates the observed density of the integer values in the data. Figure 22 superimposes these two density estimates, with the black line and points depicting the

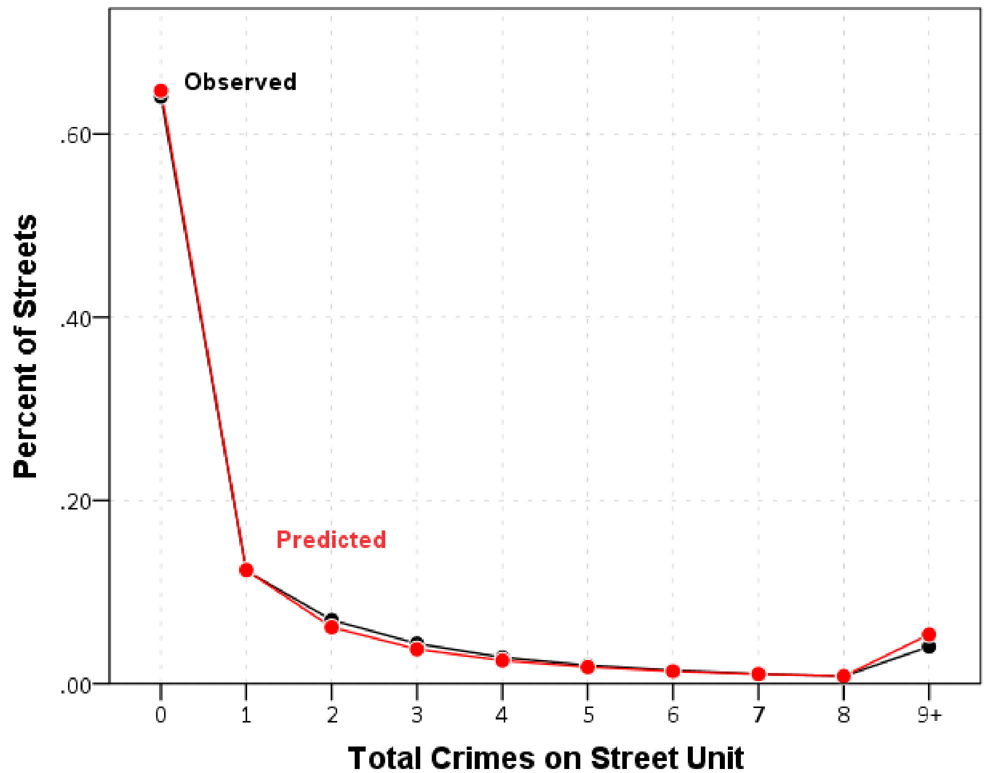


Figure 22: Predicted versus observed values for the negative binomial regression model predicting crimes at street units.

percentage of observed integer values and the red displaying the predicted values. They both agree very closely, and one can see that the zero inflation is adequately modelled with the negative binomial specification.

I further critically examine the model for abnormal data points (in terms of influential observations and nonsensical predictions), spatial autocorrelation, and provide some additional tests to identify if bars self-select into criminogenic neighborhoods.

8.3.1 Outliers

Upon examining the model residuals there were 11 locations that were greatly over or under predicted. The large over predictions tended to be locations that had an outlying number of

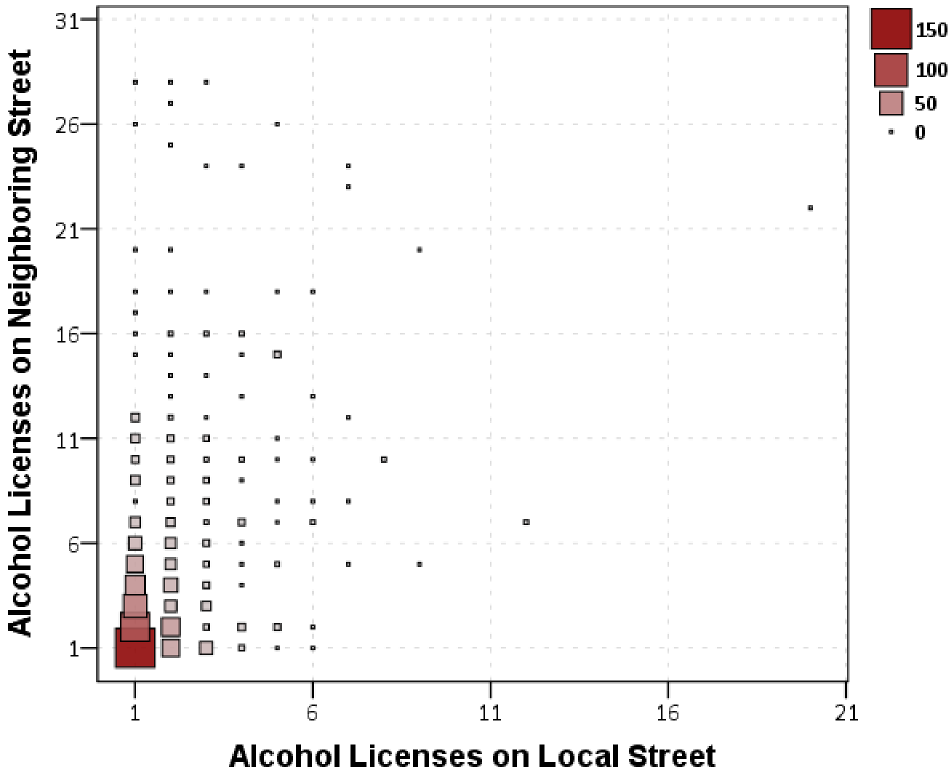


Figure 23: A binned scatterplot showing high leverage locations for interaction of local bars and neighboring bars. Streets with zero bars on the local or the neighboring street are not shown, as the interaction of those two would be equal to zero. The *area* of the square corresponds to the number of observations that fall within that bin.

bars on the local and/or neighboring streets - making the interaction effect of the two a likely culprit to produce a highly influential value. In Figure 23 I show a binned scatterplot to demonstrate the number of locations that potentially have high leverage values, in particular for the interaction effect of local bars and neighboring bars. Streets with either zero bars on the local street or neighboring street are not shown, as they will be equal to zero for the interaction effect. One can see that the vast majority of the distribution of bars falls within the lower left corner of the plot, but there are a few outlying locations that have either a high number of local bars, neighboring bars, or both.

These 11 locations were spatially clustered in central D.C., slightly north of the national mall. I have created an online map where one can view the locations of the aberrant predictions along with several of the values for key variables in the model at `https://dl.dropbox.com/s/jxg2mjss6pvjam3/OutlierMap.html`. See Figure 24 for a screen shot of this map. To check the robustness of the presented regression model I conducted two separate regressions; the first simply eliminating these 11 locations and estimating the same regression model, and the second adding dummy variables for these 11 locations and re-estimating the model. Neither of these had any substantive impacts on the parameter estimates for bars on the local or neighboring street.

Examining the two under predicted locations, one is at large mall, and the other is clustered within a location with high predicted values. Based on these observations I would make two general points about the model. The first is that some missing covariates (such as a place being a mall or measures of gentrification) are likely missing from the model. The second is that the vast majority of locations in D.C. only have a few bars, and extrapolating the estimated bar effects to areas with a large concentration of bars is inappropriate. This is unfortunate, as these bar hot spot areas are likely some of the most concern of city planners, but because they are outliers, they are difficult to model in this cross-sectional design.[6] In terms of omitted variables from the built environment, it may be expected that these omissions moderate the effect of bars on crime. For instance Pridemore and Grubesic (2012, 2013) find various interactions between measures of land use and alcohol outlets predicting crime at the block group level.

[6] My limited experience is that this is a common occurrence with Poisson regression models and the exponential link function. Outliers for covariate values tend to cause predictions to explode far beyond reasonable values.

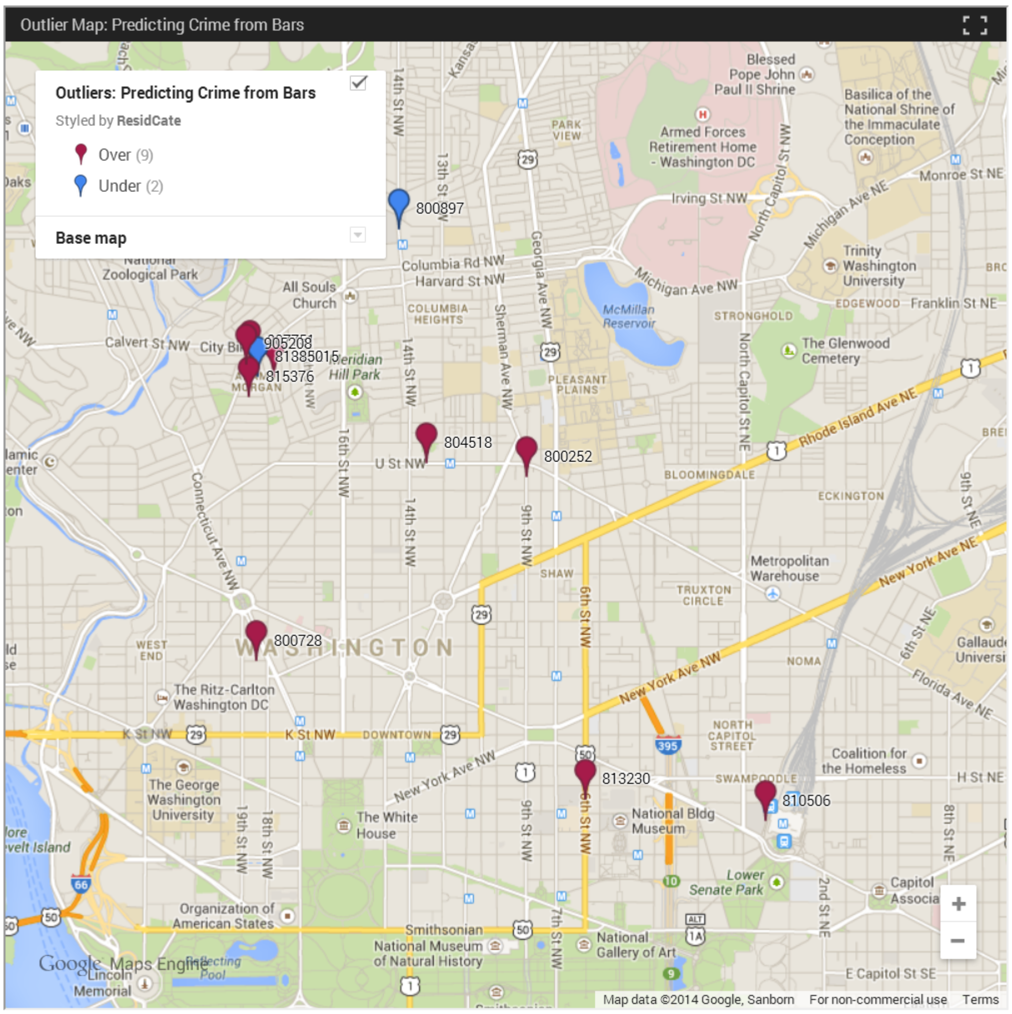


Figure 24: Screenshot of online map showing areas where the model predicts poorly. An interactive version can be accessed at `https://dl.dropbox.com/s/jxg2mjss6pvjam3/OutlierMap.html`.

8.3.2 Spatial autocorrelation

Spatial autocorrelation is a concern with the regression model presented because if present will cause standard errors to be biased downward, and in the case of a positive spatial auto-regressive process will cause parameter estimates to be biased away from zero (Lesage and Pace, 2009). There does appear to be residual spatial autocorrelation in the model residuals. The global Moran's I estimate of the deviance residuals is 0.13, and the map in Figure 25 displays those deviance residuals. It appears that in the central part of D.C., where crime is the highest, the deviance residuals tend to be positive (that is crime is under-predicted) and going towards the periphery of the city is predominately blue, and so crime is over-predicted.

Subsequent models either need to consider additional covariates that explain the spatial autocorrelation, add additional spatial terms to the model to make the residuals white noise (such as Eigenvector filters (Patuelli et al., 2011; Hodges and Reich, 2010)), add deterministic spatial terms such as distance from the city center or some flexible function of the spatial coordinates (Ridgeway, 2006), or estimate more explicit spatial models such as SAR, Spatial-Durbin, or CAR models. If one assumes that the auto-correlation in the residuals is simply due to the auto-correlation in the errors, then the effects presented would be unbiased for OLS models, but this is not the case for generalized linear models (Freedman, 2006; Giles, 2013) and the coefficient estimates are not consistent. In OLS models the standard errors would be too small, but given the size of the sample here that is not of much concern. However even in OLS models, if the true data generative process is an endogenous auto-regressive process and that lagged process is positive, it will bias the estimates away from 0 (Lesage and Pace, 2009).

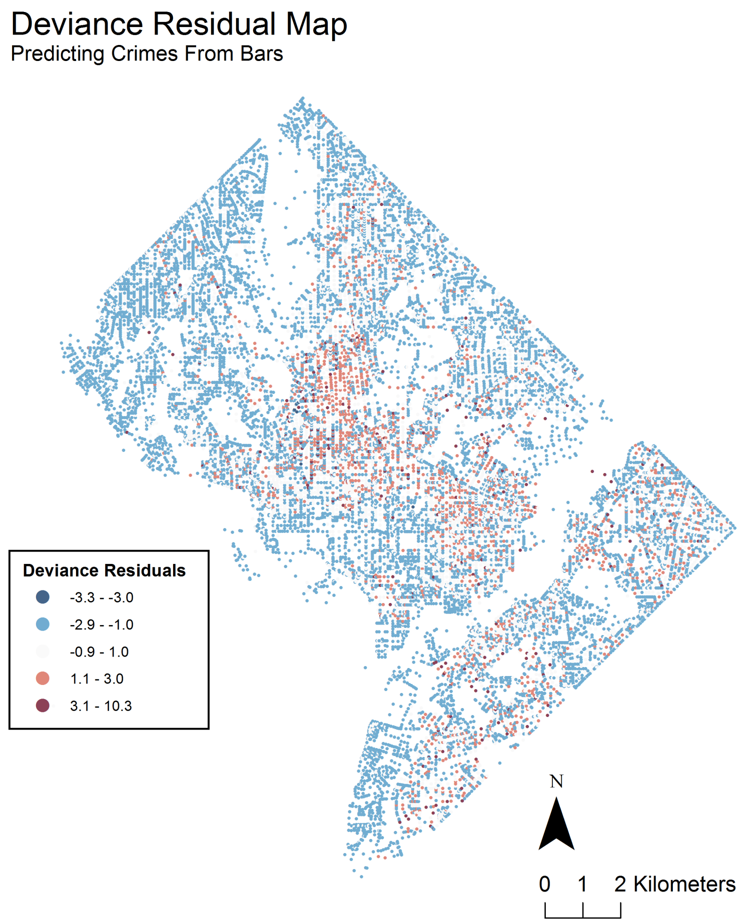


Figure 25: This map displays the deviance residuals for the negative binomial regression model predicting crimes. Blue areas represent low residuals (areas that are over-predicted) and red represent under predicted locations. The residuals visually appear to have positive spatial autocorrelation, and the global Moran's I value is 0.13.

8.3.3 Selection effects

The biggest challenge with working with observational data is that omitted variables may account for some or all of the observed effects that are estimated. In this case when examining bars and crime it is certainly plausible that bars self select into areas that already have elevated occurrences of crime. Possible mechanisms for this may be those locations have lower rents or housing prices, or the local residents lack the social capital to block such commercial establishments.

To estimate whether this is the case, one can notice that the theoretical relationship between bars and crime explicated at the beginning of the chapter is mostly only applicable to interpersonal crimes. For example people being inebriated should not increase the risk of a *home* being burglarized. This allows for a non-equivalent dependent variable test (Shadish et al., 2002) to determine if bars self-select into already crime prone neighborhoods.

The logic of the test is as follows; we are concerned that an omitted variable U might effect crime and be correlated with bars. For a certain set of data, say Y_{alt}, we theoretically *know* the effect of bars to be zero, but U affects this set of data just the same. If we then estimate the effect of bars on Y_{alt} and find it to be non-zero, this is evidence that there is a relationship between U and bars.

To attempt to estimate this indirect (but unobserved) effect, I propose a graphical model in Figure 26 (Pearl, 2000). Making a few assumptions, from this graphical model we can theorize the total, direct and indirect effects of bars on crime. The two particular assumptions of interest in this model are; 1) that there is no direct effect of bars on burglary, and 2) to estimate the indirect selection effect of bars on other crimes it will be assumed that $a_1 = a_2$.

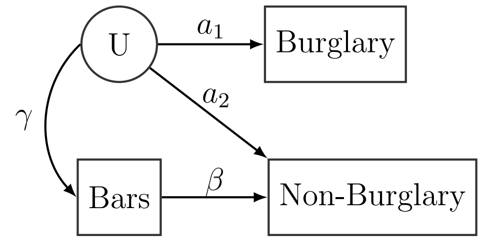


Figure 26: Path diagram depicting non-equivalent dependent variable design. U is unobserved, but the indirect effects can be estimated assuming Bars have no direct effect on burglaries and that $a_1 = a_2$.

In reality neither of these assumptions are likely to be true, but such violations are likely to *decrease* our estimate of the unobserved selection effect.

Because U is unobserved in Figure 26, with the given data we can only estimate the total effect of bars on burglaries and non-burglaries. By assuming that bars can not directly affect burglaries, we can assume the indirect effect, following along the inward path from Bars to U to Burglaries, is equal to the total effect we can estimate.

$$\text{Total Effect}_{\text{Bars}\rightarrow\text{Burglary}} = \text{Indirect Effect}_{\text{Bars}\leftarrow\text{U}\rightarrow\text{Burglary}} \tag{8.7}$$

Here I am using Total Effect as shorthand for $\mathbb{E}\left[\text{Burglary}|\text{Bars}\right] = f(\text{Total Effect})$. In linear models this total effect would be equal to $\gamma \cdot a_1$ (Wright, 1934) but in non-linear models the decomposition is more complicated (Buis, 2010; Imai et al., 2010). For simplicity I will write following these paths to be $g(\gamma, a_1)$, an anonymous function g of the two path coefficients. It may be possible for bars to directly effect burglaries. For example, bars can be locations where motivated offenders spend a disproportionate time committing other deviant activities, such as selling drugs (Rengert, 1996), and nearby areas are exposed a greater amount. But note if this is the case, our estimate of the indirect effect in equation 8.7 would be *larger* than it is in reality.

The total effect on non-burglaries is a sum of both a direct effect, β, and an indirect effect, $g(\gamma, a_2)$.

$$\text{Total Effect}_{\text{Bars}\rightarrow\text{Non-Burglary}} = \beta + g(\gamma, a_2) \tag{8.8}$$

I assume that $g(\gamma, a_1) = g(\gamma, a_2)$ to decompose this direct and indirect effect. This assumption makes the most sense if one were estimating standardized path models, e.g. neighborhood selection has the same affect on all crimes. Because the variance of burglaries is much smaller than non-burglaries (0.8^2 and 4.1^2 respectively), it would make more sense to assume that $a_1 < a_2$. These most likely violations would cause the estimate of the effect of the indirect effect to be smaller, so this is a conservative test of selection effects for bars on crime.

Table 13 displays these separate models, predicting burglaries, all other crimes besides burglaries, and the difference between the coefficients as the estimate of the direct effect of bars on crime. Each model uses the exact same specification of other variables as does the original model, and diagnostic checks for each resulted in similar conclusions (e.g. the negative binomial predicted the near 90% of zero observations for the burglary model very closely). Model estimates and standard errors are placed in columns, as comparing numbers within one column is more convenient than comparing across one row (Feinberg and Wainer, 2011). One can see that the local and spatial effects of bars are much smaller, although

Table 13: Effects of bars on crime for burglaries and non-burglaries.

Dependent Variable	**Local**	**(S.E.)**	**Neighbor**	**(S.E.)**
Non-Burglary	.55	(.06)	.24	(.02)
Burglary	.15	(.06)	.08	(.02)
Direct Effect	.40	(.07)	.16	(.02)

still statistically significant, for predicting burglaries. Effects for non-burglaries are slightly larger for local effects and spatial effects compared to the grand model. This provides some evidence that part of the effect displayed earlier is due to selection, but that selection effect appears to be much smaller than the total effect. Standard errors of the direct effects are calculated using 99 bootstrap replications.

The estimate of the unobserved indirect effect in Table 13 is $g(\gamma, a_1) = g(\gamma, a_2) = 0.15 + 0.08 = 0.23$ (including both the effect of local bars and neighboring bars). The observed total effect for non-burglaries is $0.55 + 0.24 = 0.79$. So the estimated direct effect of bars on non-burglaries is then $\beta = 0.79 - 0.23 = 0.56$. This is a change from an incident rate ratio of 2.2 to 1.8, still quite large although greatly moderated. This estimates that 29% of the observed total effect (in the linear parameters) is due to selection. For pragmatic reasons, to estimate the variance of these indirect effects one typically uses the bootstrap (Imai et al., 2010; Preacher and Hayes, 2004). In 99 bootstrap replications the estimate of the total effect ranged from 0.68 to 0.97, the estimate of the indirect effect ranged from 0.09 to 0.44, and the estimate of the total minus the indirect ranged from 0.34 to 0.76.

The estimates in table 13 include additional control variables, so the path diagram listed are conditional on the control variables, but the estimates of the indirect effects work the same (with the control variables in the equation they *block* the paths from Bars to crime, so there is no need to estimate the indirect effects along those paths). Still, even with a large proportion of the local and neighboring effects of bars on crime mediated it still has quite a large impact on crime that seems unlikely to be entirely explained by selection into particular criminogenic areas.

8.4 Conclusion

The contribution of this chapter are separate estimates of the local and spatial effects of liquor licenses on serious crime separately at the street unit level. It was found that the *total* size of the spatial spill-over effects were larger on the neighboring street units than they were on the local street unit. This theoretically suggests that bars have a diffusive effect into the environment beyond the location of the bar itself.

This also has implications for how bars are zoned, as they do not simply impact the local area but the surrounding areas a great deal as well. The median partial effect of adding a bar on a street is under 1 additional part 1 crime on the local street, and under 2 additional part 1 crimes on the neighboring streets in one year.

The model as specified though is problematic. Spatial auto-correlation still exists in the residuals, and several locations have wildly inaccurate predictions. It also appears that some selection of bars is occurring, as theoretically bars should have little to no effect on burglaries. These suggest that the presented effects of bars on crime are overestimates of their true effect.

The following chapter will present additional analysis of the relationship between 311 calls for service and crime, and attempt to see if some of the problems with the model are mitigated by using neighborhood fixed effects. The chapter after that will consider a more general model of crime, and incorporate a wider array of place based predictors in an attempt to make a more realistic model of crime at micro places.

Chapter 9

311 Calls for Service and Crime

This chapter will examine the relationship between 311 calls for service, which are public supplied reports requesting service for particular problems in the city, and crime. The majority of these calls are for problems related to public property (e.g. pothole or broken sidewalk, garbage on the street) and thus are public supplied measures of physical disorder. Similar to the previous chapter, the motivation for examining the relationship between 311 calls for service and crime is based on:

- 311 calls for service are a proxy for physical disorder. Thus it provides an empirical test to validate or refute broken windows theory (Wilson and Kelling, 1982).
- It provides an illustration where the distinction between neighborhood effects, local effects and spatial effects are of theoretical interest to criminologists.

The logic behind the relationship between 311 calls is as follows: in broken windows theory, *visible* signs of disorder are cues that deviant behavior will be acceptable within that area. This would suggest that visible signs of disorder should have the strongest effects in the immediate visual arena, and should have very small to non-existent spatial and neighborhood effects. The existence of either may be better explained by omitted variables, or broken windows theory as typically stated is incomplete[1].

So based on this, I propose the following hypothesis:

Hypothesis: Broken windows will *only* have local effects on crime.

One problem with identifying the effect of broken windows on crime is that particular neighborhoods may be more likely to have signs of physical disorder and higher rates of crime, but those higher rates of crime may be for other reasons besides those broken windows. The following section describes how I use a regular grid to approximate neighborhoods (which we can not observe), and how this identification strategy relates to the use of census geographies to approximate neighborhoods.

9.1 On Fixed Effects and Neighborhoods

One may be concerned with the models presented that do not account for theoretical neighborhood processes that might explain the levels of crime at street units (Sampson, 2012). In particular for a theoretical process between broken windows and crime, it has been suggested that individuals associate physical signs of disorder either based on prejudicial attitudes towards minority neighborhoods (Sampson and Raudenbush, 2004) or with crime itself (Gau and Pratt, 2008). It is also possible that the opposite effect occurs due to the self selecting nature of the 311 calls for service; neighborhoods with more affluent individuals may

[1]See chapter 4 for a description of broken windows theory

be more likely to report physical damages in the built environment. One example from a Boston service that used a smart phone application to report potholes found that minority and elderly populations were under-represented in the calls for service (Executive Office of the President, 2014).

Here I will illustrate how using fixed effects regression equations (Allison, 1990) can potentially control for unobserved neighborhood processes under the condition you *know* where the neighborhoods are. Because we do not know explicitly know where the neighborhood boundaries are, I then go on to show how using smaller neighborhoods provides more protection against omitted variables bias.

First consider a *micro-place* model of the form:

$$y_{ij} = \beta_1(x_{ij}) + \beta_2(Z_j) \tag{9.1}$$

In this equation y_{ij} represents the dependent variable at place i nested within neighborhood j, and respectively x_{ij} represents the explanatory variable at the same location. The item Z_j is a characteristic of the *neighborhood* j, and so is invariant for all locations within a neighborhood but varies between neighborhoods.

In this formulation one can aggregate these micro place equations up to the neighborhood level. So for illustration one can write this micro level equation for all i units in one specific j neighborhood:

$$y_{1j} = \beta_1(x_{1j}) + \beta_2(Z_j) \tag{9.2}$$

$$y_{2j} = \beta_1(x_{2j}) + \beta_2(Z_j) \tag{9.3}$$

$$y_{3j} = \beta_1(x_{3j}) + \beta_2(Z_j) \tag{9.4}$$

$$\vdots \tag{9.5}$$

$$y_{ij} = \beta_1(x_{ij}) + \beta_2(Z_j) \tag{9.6}$$

And if one sums each of the i units, one then has the equation (where n is the total number of units within a neighborhood):

$$\sum_{i=1}^{n} y_i = \beta_1(\sum_{i=1}^{n} x_i) + n \cdot \beta_2(Z_j) \tag{9.7}$$

If you then divide this equation by n on both sides one then has the function represented per the mean of the neighborhood (frequently referred to as the between effects estimator (Snijders and Bosker, 2012)). Using the relationship that $\frac{\sum_{i=1}^{n} a_i}{n} = \bar{a}$ we then have the mean equation for neighborhood j:

$$\bar{y}_j = \beta_1(\bar{x}_j) + \beta_2(Z_j) \tag{9.8}$$

To control for the between neighborhood model then, one then subtracts 9.8 from 9.6 to obtain the fixed effects estimator.

$$y_{ij} - \bar{y}_j = \beta_1(x_{ij} - \bar{x}_j) + \beta_2(Z_j - Z_j) \tag{9.9}$$

Because $\beta_2(Z_j - Z_j)$ equals zero, one can then drop this term from the equation. So now we can estimate β_1 without observing the neighborhood effect Z_j simply by subtracting out the neighborhood level means of the observed variables. This model is equivalent to fitting a model where there are a set of dummy variables for each neighborhood (Hardin, 1996), and so it can be extended to generalized linear models such as the Poisson regression equations used in this research. Unfortunately though not all neighborhoods will be identified under this procedure for limited dependent *or* explanatory variables (Mundlak, 1978). If the variable nested within the neighborhood does not vary, that neighborhood and all of its nested micro level units will need to be dropped from the equation.

Clearly the weakness of such a fixed effects approach is that *we do not know where the neighborhoods are.* Typically they are chosen more out of convenience (e.g. census units, a regular grid) than they are based on any theoretical considerations.[2] One needs to keep in mind that this conception of neighborhoods is based on the causal outcomes, e.g. the neighborhood is itself defined by the contextual effect of $\beta_2(Z_j)$.[3] Typically neighborhoods

[2]In some situations exogenous interventions can create a neighborhood effect, such as in state level policy implementations (Deane et al., 2008).

[3]This causal outcome view of neighborhoods is unlikely to be a satisfactory definition of neighborhoods for the sake of studying neighborhoods. Consider an extreme example in which a city contains two neighborhoods, A and B. Lets say A and B are on opposite sides of the city, but are otherwise identical, e.g. have equivalent values for Z_j. For our fixed effects regression model we could aggregate these two neighborhoods together and still identify the local effect, β_1, even though these two neighborhoods have nothing to do with one another. The equivalence of Z_j allows them to be considered the same neighborhood, even though they have no interaction with one another.

are conceptualized as having some type of internal homogeneity (Sampson, 2012), and this is often given as a reason why census units are reasonable to use as proxies for neighborhoods. Demographic homogeneity seems unlikely to rule out other possible neighborhood effects not captured by those static units. There may be alternative neighborhoods of influence, such as gang territories or school districts that affect crime that street units may be nested within.

Because the borders of neighborhoods are unknown, *smaller* neighborhood units provides a stricter test of controlling for unobserved neighborhood characteristics. Consider a simple example of a space delimited by two neighborhoods, A and B. In the first diagram on the left lists the actual neighborhoods, and the second diagram on the right shows our smaller neighborhoods nested within the larger neighborhoods.

A
B

a_1	a_2
b_1	b_2

Here y is a function of being in either neighborhood A or neighborhood B (where if a unit is nested within neighborhood A the dummy variable takes on a value of one and zero otherwise). But lets say we end up estimating the equation with our smaller neighborhoods, a_1, a_2, b_1, b_2.

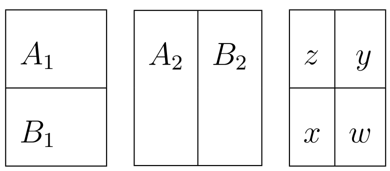
$$y = \beta_1(A) + \beta_2(B) \tag{9.10}$$

$$y = \beta_1(a_1 + a_2) + \beta_2(b_1 + b_2) \tag{9.11}$$

$$y = \beta_3(a_1) + \beta_4(a_2) + \beta_5(b_1) + \beta_6(b_2) \tag{9.12}$$

So equation 9.12 it is equivalent to the model with the larger true neighborhood boundaries. It happens that $\beta_3 = \beta_4 = \beta_1$ and similarly $\beta_5 = \beta_6 = \beta_2$. So the estimates based on the smaller neighborhoods will be less efficient (as we estimate four parameters instead of only the two that are needed) but they will not be biased.

Now consider the case of cross-classified neighborhoods, illustrated in the following diagram with two neighborhood boundaries that are not mutually exclusive areas. The diagram in the left and the middle represent our two neighborhoods, and the rightmost diagram represents our smaller neighborhood boundaries.



These neighborhoods are represented in a similar set of equations below where a unit is affected by membership in either neighborhood A_1 or neighborhood A_2 (and an intercept β_0).

$$y = \beta_0 + \beta_1(A_1) + \beta_2(A_2) \tag{9.13}$$

$$y = \beta_0 + \beta_1(z + y) + \beta_2(z + x) \tag{9.14}$$

$$y = \beta_0 + (\beta_1 + \beta_2)z + \beta_1(y) + \beta_2(x) \tag{9.15}$$

$$y = \beta_3(z) + \beta_4(y) + \beta_5(x) + \beta_6(w) \tag{9.16}$$

The relationships become more complicated, but the smaller set of neighborhood dummies will still provide an unbiased estimate of the neighborhood effects. It happens that the smaller neighborhoods become confounded and can not distinguish between the larger neighborhood effects, e.g. $\beta_3 = \beta_1 + \beta_2$. Here we can still see that the smaller neighborhoods are less efficient, the larger neighborhoods take three coefficients (including the intercept β_0), but the smaller neighborhoods need to estimate four parameters (in this notation $\beta_6 = \beta_0$). In the case that the different neighborhood units A_1 and A_2 have interaction effects, it would be equivalent to the smaller nested neighborhood model. Again though the true neighborhood effects are not recoverable, so we can merely control for them, not identify them in any realistic way in these fixed effects equations.

This fixed effects approach for partialing out the neighborhood is limited because we will never be sure that our neighborhood unit is small enough to capture the unobserved neighborhood effects. For instance, Taylor (1997) argues that street blocks are themselves a particular neighborhood context. One could always shrink the neighborhood down to smaller and smaller units, and likely account for nearly all of the variation in crime, making identification of the effects of other predictors on crime impossible.

In subsequent analysis I will use a regular grid that is based on cells of size 500 meters^2. In comparison to prior criminological literature delineating neighborhoods, the work of Shaw and McKay (1969) used square mile grid cells for most of Chicago, agglomerating less populous areas in the periphery of the city (one square mile is slightly more than 1,600 meters square). In other cities though Shaw and McKay (1969) tended to concatenate census tracts to create areas of around 1 to 2 square miles. For example, Cleveland (and the included surrounding towns) was listed as a total area of 94 square miles (pg. 272) and comprised of 252 census tracts (pg. 274), but was then aggregated to a total of 45 areas for subsequent analysis.

More contemporary research tends to use census tracts (Boggs, 1965; Browning et al., 2010; Griffiths and Chavez, 2004; Hipp, 2007; Krivo and Peterson, 1996; Tita et al., 2005) or block groups (Socia, 2011; Taniguchi and Salvatore, 2012; Weisburd et al., 2008; Yang, 2010) or similar to Shaw and McKay (1969) aggregations of these census units. The most notable of these are studies using data collected from the Project on Human Development in Chicago (Mears and Bhati, 2006; Morenoff et al., 2001; Sampson et al., 1997) in which census tracts in Chicago are aggregated into homogeneous neighborhoods. Wang (2005) does this as well for the city of Chicago, although does not use PHDCN data nor the same aggregated units, and Britt et al. (2005) use neighborhoods defined by the The City of Minneapolis Planning Department that are aggregations of census blocks and tracts that align well with the city's perceptions of neighborhoods. So from this (with the exception of the work of Ralph Taylor) the *smallest* unit used to represent *neighborhoods* is a census block group, and the majority of researchers use larger areas (sometimes much larger) to represent neighborhoods.

The use of 500 meters^2 grid cells is similar in average size to that of block groups in Washington, D.C. The mean size of block groups in D.C. is 538 meters^2, but of course is irregular, so the sizes range from 180 to 3,400 meters square. Census tracts in D.C. have a mean size of slightly under 900 meters^2 and a range of 400 to 3,400 meters square.[4]

[4] These area figures are calculated from 2010 census geographies without removing water bodies. The block group/tract that is around $3,400$ meters^2 is the national mall and includes some area to the west of the Potomac.

9.2 Description of Measures

Measures of 311 calls for service in 2010 are taken from DC.gov and are already geocoded. In 2010 there were a total of 429,676 calls for service, with over half of the calls for service being complaints about parking meters. Other calls for service are not directly related to physical disorder either (such as parking enforcement) and so are not included in this analysis of calls for service. Table 14 shows the most common labels for complaints within the 311 calls for service in 2010, and highlights the labels that are taken to be measures of disorder. Grey lines are measures of *detritus* related aspects of physical disorder, and blue lines are measures of *infrastructure* related aspects of physical disorder. Included in the table are the frequencies of such complaints.

Taylor (2001) considers the distinction between transient types of disorder (synonymous with my label of detritus here) or more permanent damage to physical infrastructure. Each may signify different types of problems in a particular community, and so I continue to use the same distinction here in this research. Transient types of physical disorder, such as garbage on the street or an unkempt lawn, are relatively easy to fix. Specifically, *one* motivated individual can clean up the area in a relatively short amount of time. For problems related to physical infrastructure though, such as fixing a sidewalk or repainting over graffiti, it takes more monetary investment.

These different types of disorder also may have different potential confounds in their relationship to crime. It has been found that areas proximite to commercial establishments tend to have more of detritus in plain view (Forsyth and Davidson, 2010; Taylor et al., 1995). Physical deterioration of infrastructure is plausibly related to general poverty, as

Table 14: Most common 311 calls for service, D.C. (2010). Blue are infrastructure related and grey are detritus related.

Call Type	**Frequency**	**Call Type**	**Frequency**
Parking Meter Request	222,682	How's My Driving - Complaint	944
Bulk Collection	52,109	DMV - Driver's License/ID Issues	907
Parking Enforcement	19,230	DMV - Vehicle Title Issues	892
Snow/Ice Removal	10,833	Alley Repair	883
Sanitation Enforcement	7,765	Vacant Lot	794
Trash Collection - Missed	6,729	Sign New	735
Recycling Container Delivery	6,268	Grass & Weeds Mowing	554
Alley Cleaning	5,870	Curb and Gutter Repair	546
Pothole	5,681	DMV - Adjudication Supervisor	516
Streetlight Repair	5,565	DMV - Driver's License/ID Reinstatement	512
Abandoned Vehicle - On Public Property	5,352	Sign Missing	512
Yard Waste- Missed	4,369	Utility Repair	457
Tree Removal	4,125	DMV - Refunds - Tickets	416
Tree Inspection	3,771	xxx_Abandoned Vehicle LOOK UP ONLY	387
Supercan - Delivery	3,763	XXX_Curb & Gutter Repair - Major	387
Rat Abatement	3,123	XXX_Tree Trimming	384
DMV - Vehicle Registration Issues	2,919	XXX_Street Repairs	383
Abandoned Vehicle - On Private Property	2,900	DMV - Insurance - Lapse in Coverage	343
Illegal Dumping	2,808	Litter Can - Collection	341
TRU Report	2,706	DMV - Processing Center Manager	339
Trash Container - Delivery	2,682	Illegal Posters	317
Recycling Collection - Missed	2,604	XXX_DCRA - Trash and Debris	314
Supercan - Repair	2,544	DMV - Ticket Payment Dispute	278
Dead Animal Pickup	2,526	Leaf Collection	253
Street Cleaning	2,389	Sign Removal	243
Tree Pruning	2,380	DMV - Tag Surrenders/Registration Fee R	236
Residential Parking Permit Violation	2,355	Litter Can - Installation/Removal/Repai	228
Sidewalk Repair	2,330	Eviction	224
DMV - Send Forms, Applications, Manuals	1,691	Parks and Recreation	220
Alleylight Repair	1,670	Pedestrian Safety Program	217
Graffiti Removal	1,494	Street Sweeping	187
Out of State Parking Violation (ROSA)	1,456	DMV - Call Center Supervisor	186
Sign Replace	1,369	Christmas Tree Removal - Seasonal	165
Street Repair	1,362	Resident Parking Permit	161
DMV - Copy of Ticket	1,300	Bulk Collection - Unscheduled	155
Tree Planting	1,262	Street Paving Schedule	152
DPW Correspondence Tracking	1,238	Insects	146
Traffic Signal Maintenance	1,168	Marking Maintenance	129
DMV - Hearings	1,161	*Other Labels with fewer than 100 cases*	1,084
		Total	**429,676**

Table 15: Summary statistics for 311 calls for service

Call Type	**Mean**	**St.Dev.**	**Min.**	**Max.**	**25th**	**Median**	**75th**
Detritus	4.3	7.1	0	83	0	1	6
Infrastructure	1.1	2.0	0	44	0	0	1

people in impoverished places lack the capital to fix or update the structures. Both of these types of measures have been used in the past to proxy types of disorder, for instance Cerd et al. (2009) uses a measure of damage to infrastructure in schools as a measure of physical disorder, while Sampson and Raudenbush (1999) measure disorder by coding garbage on the street from video recordings.

Figure 27 displays the spatial variation of detritus and physical infrastructure 311 calls for service in D.C. for 2010. The maps on the left display sized circles for several classes of the number of calls for service at the street unit, and the maps on the right display a smoothed kernel density estimate (of the original coordinates). The sized circle map is not effective for visualizing the distribution, but does display that calls for service are distributed across the entire city. The kernel density estimates are calculated using a normal kernel and a 300 meter bandwidth (displayed are probabilities, so the estimates over the entire raster sum to 1) and the cell sizes are 180 meters square.

The maps generically show the same type of spatial patterning as that of crime in the previous chapter. Only squinting for an extending period of time will produce noticeable differences between the two. Table 15 shows the summary statistics for 311 calls for service aggregated to the street unit level. Detritus related calls are much more frequent, although one can see based on the relationship between the mean and the median that each distribution is positively skewed.

To visualize the regular grid that I will be using as a proxy for neighborhoods, Figure 28

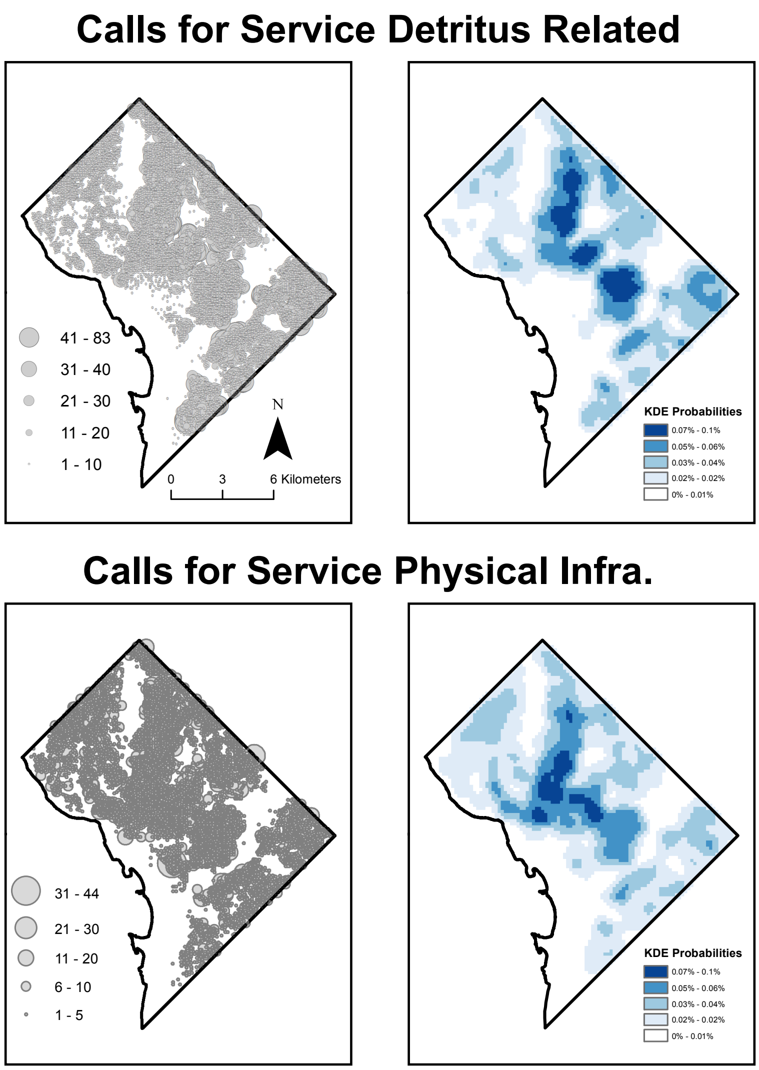


Figure 27: Spatial distribution of calls for service. Maps on the left show proportional symbols, and maps on the right show the kernel density estimate of detritus related or infrastructure related calls for service.

displays the aggregate measures of the *mean* number of crimes or calls per service for the nested street units. The map in the top left corner displays sized circles to show the number of street units within a particular grid cell. One can see from the sized circles in the top left that the majority of grid cells in the city have a large number of observations, 30 or more. Only a few cells (mostly those that share much of their area with waterways) do not have very many nested street units.

The choropleth mapping scheme uses quintiles to define the five color bins. For infrastructure calls for service this provides a similar visualization to that of the maps in the prior chapter, but crimes and detritus calls for service it does not differentiate between the central portion of D.C. and the neighborhoods in the southern portion of D.C. to the east of the Potomac and south of the Anacostia rivers. This is because the distributions for each are highly skewed, and so there is quite a large difference even within the largest quintile displayed on these maps.

The relationship between 311 calls for service and crime is presented in Figure 29. This figure shows the mean number of crimes per street units at various values for calls for service on street units. For detritus related calls for service (the left panel) the average of crime only appears to slightly rise between 0 and 9 calls for service (at a rate of perhaps 0.1 crimes per an increase in 1 call), but the relationship is quite noisy. The aggregated values of 10 or more show a distinct jump, suggesting the trend continues to rise outside of the plot range. Calls for service for infrastructure related calls for service shows a much more noticeable increase in the range of the plot, increasing by around 0.5 crimes per 1 increase the number of calls for service. Again these effects appear to be linear, suggesting the exponential link function in Poisson regression equations may not be appropriate.

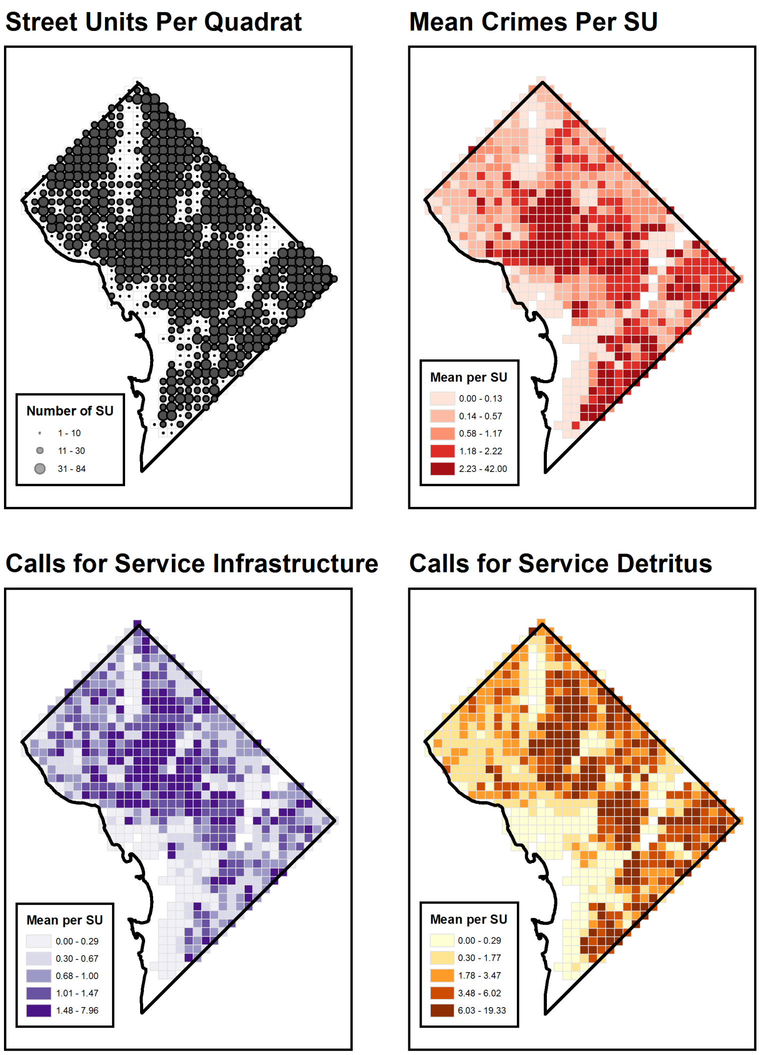


Figure 28: Summary measures for crime and 311 calls for service at the grid used for the fixed effects estimation. Grid cells are 500 meters square.

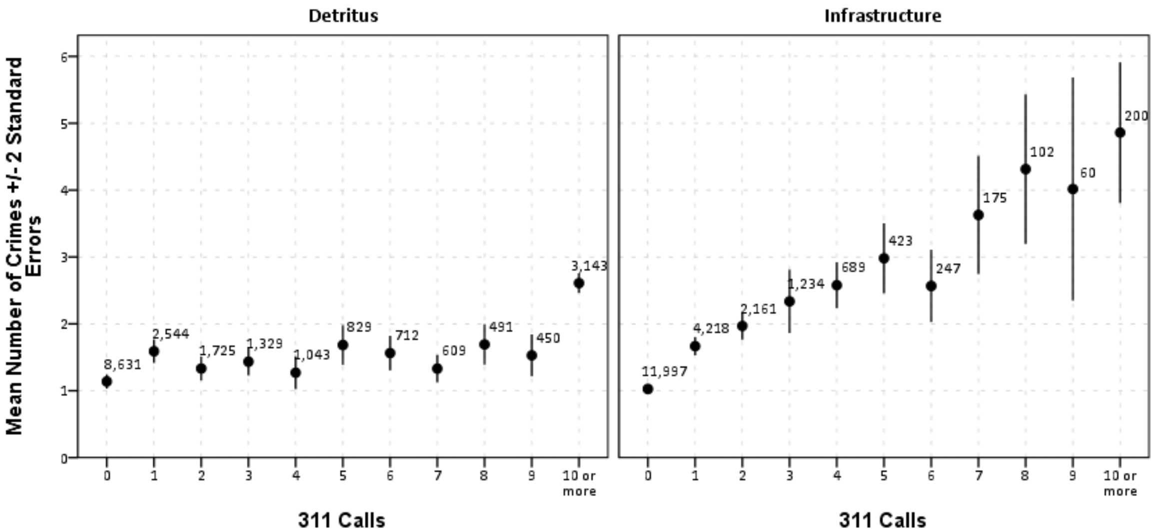


Figure 29: Marginal effects of 311 calls for service on crime at the street unit level.

To examine the extent to which *neighborhood* effects for physical disorder on crime exist, a simple test is to evaluate the correlation between the aggregate mean per the neighborhood and the outcome at the micro level (Firebaugh, 1978). Table 16 displays the Pearson correlations for the micro measurements of crime and calls for service, along with the correlations based on the aggregate measures of calls for service at the neighborhood quadrant level. Because of the large sample size (21,506 observations) only a correlation of absolute size of 0.014 is needed to reach a t-value of 2 (i.e. statistically significant at a level of .05) and only a correlation of 0.034 to reach a t-value of 5, so all coefficients presented will be statistically significant even if they are quite small. Here we can see that the correlations for the mean number of calls for service are around the same magnitude as the local effects (examining Kendall Tau correlations leads to the same conclusions).

I take this to mean that there are omitted neighborhood variables driving the relationship between 311 calls for service and crime at the neighborhood level. So subsequent models need to incorporate the neighborhood fixed effects to control for this potential confounding of the

Table 16: Pearson correlations for crime and calls for service at the micro and neighborhood level (n = 21,506)

	Crime	Detritus	Infra.	Mean Detr.	Mean Infra.
Crime	1.00				
Detritus	.13	1.00			
Infra.	.15	.34	1.00		
Mean Detr.	.10	.47	.19	1.00	
Mean Infra.	.20	.24	.37	.51	1.00

local effects of physical disorder (King, 1996). This also suggests that you *can not* identify the local effect for the relationship between physical disorder and crime at the aggregate neighborhood level, you need micro level data *within* the neighborhood to properly identify the relationship. The subsequent section will fit such multiple regression models to examine if the local effect of physical disorder is entirely confounded with the unobserved neighborhood effects.

9.3 Model Fitting

Building on the same Poisson regression models estimated in the prior chapter, here I estimate additional parameters for detritus and infrastructure related 311 calls for service. These models are simply extensions of the same Poisson models presented in the prior chapter, predicting the log of the expected number of crimes on a street unit as a function of detritus and infrastructure 311 calls for service *in addition to* the prior variables included (bars on the same street, bars on neighboring street, non-linear effect of area, and whether the street is an intersection). As several neighborhood units are not identified in the subsequent models because they are invariant with respect to the explanatory variables, they are dropped from all other models presented in this chapter. This results in a reduction of 1,135

Table 17: Linear coefficient estimates on the log expected value of crime at the street unit

	Det. Local			Det. Spat.			Infra. Local			Infra. Spat.		
Model	β	S.E		β	S.E		β	S.E		β	S.E	
(1) Only Local	0.022	0.002	**				0.068	0.008	**			
(2) Local and Neighborhood	0.011	0.002	**				0.031	0.006	**			
(3) Local and Spatial	0.008	0.002	**	0.004	0.001	**	0.048	0.008	**	0.019	0.003	**
(4) Local, Spat. and Neigh.	0.009	0.002	**	0.001	0.001		0.030	0.006	**	0.002	0.003	

* indicates significant at the .05 level

** indicates significant at the .01 level

street units, from 21,506 to 20,371.

Table 17 presents the linear effects of detritus and infrastructure calls for service on crime for four different model specifications. Model 1 estimates only the local effects of 311 calls for service without neighborhood fixed effects. Model 2 estimates only the local effects of 311 calls for service with the neighborhood fixed effects. Model 3 estimates the effects of the calls for service on the local street and the sum of calls for service on the neighboring streets, and Model 4 is the same but also includes the neighborhood fixed effects.

While one could possibly conduct a series of tests to select one of these 4 models (models 1,2 and 3 are all nested within model 4), I prefer simply to interpret the local and spatial effects of 311 calls for service in the presence and absence of the neighborhood fixed effects. The inclusion of the fixed effects are not to define the best model, but to identify the effects of the calls for service on crime absent of potential omitted neighborhood effects. One could always make them smaller, or even perhaps use a data based definition of neighborhoods (Weeks et al., 2010). These will likely give better fits, but do not necessarily help identify an unbiased effect of 311 calls for service on crime.

The estimates present evidence that the effects of 311 calls for service are small, but the local effects are only partially moderated when the neighborhood fixed effects are included in the model. In all models the spatial effects are about half the size of the local effects (although remember the typical diffusion would make them of around equal size for assessing actual marginal changes in crime), and the spatial effects when including the neighborhood fixed effects are positive but very close to zero and have approximately the same size standard errors (and so fail to reach statistical significance).

So given the model uncertainty, I interpret the results to suggest;

- 311 calls for service have a small effect on crime, invariant to different model specifications including neighborhood fixed effects.

- The spatial effects of 311 calls for service are confounded with neighborhood effects.

Table 18 presents the full results for Model 2 (excluding the 555 neighborhood dummy variables), including only the local effects and neighborhood effects for calls for service on crime. This is included to illustrate the effects on other portions of the model, in particular the effects of bars on crime, with the inclusion of the neighborhood fixed effects (and the reduction of 1,000 cases). Note again this model uses 20,371 observations, so is not directly comparable to the model in the previous chapter.

How small are the effects of 311 calls for service? Figure 30 presents marginal effect plots for detritus and infrastructure calls for service where the values of $0, 1 \ldots k$ are plotted on the X axis and $e^{\mu + k}$ is plotted on the Y axis. Values of μ are varied as well, between integer values of 0 and 3 ($e^3 \approx 20.1$), so this provides support of the potential effects of 311 calls for service over the most typical values of crime for street units.

Table 18: Negative Binomial Model Predicting Crime at Street Units: Parameter estimates & 95% Confidence intervals for the exponentiated estimates

Variable	β	S.E.	CI_L	CI_H
Intercept	-1.70	1.11	0.02	1.60
Bars_L	0.36	0.04	1.33	1.55
Bars_N	0.08	0.01	1.06	1.11
$(\text{Bars}_L \cdot \text{Bars}_N)$	-0.02	< 0.01	0.98	0.99
Detritus	0.01	< 0.01	1.01	1.01
Infrastructure	0.03	0.01	1.02	1.04
Intersection	-0.50	0.03	0.57	0.65
log(Area)	0.02	0.14		
K_1	2.77	0.88		
K_2	-24.85	11.85		
K_3	63.19	32.50		
K_4	-73.18	34.50		
K_5	31.81	14.37		
a	2.22	0.04		

$$\log\left(\mathbb{E}\left[\text{Crime}\right]\right) = \mu + 0.011 \cdot (\text{Detritus}_k) = \mu_{+k} \tag{9.17}$$

$$\log\left(\mathbb{E}\left[\text{Crime}\right]\right) = \mu + 0.031 \cdot (\text{Infra.}_k) = \mu_{+k} \tag{9.18}$$

Figure 30 shows that the conditional predicted effects of detritus related 311 calls for service on crime are so small that only in places with a large amount of crime and over a large range of calls can one identify an appreciable change in the predicted number of crimes. For infrastructure related 311 calls, if the expected number of crimes were log(2) and you increased the number of infrastructure 311 calls for service from 0 to 15, the expected increase in crime would be less than 5 crimes within the year. The typical increase on this particular grey curve is around 1 extra crime per 3 calls for service. So while detritus related calls are statistically significant in the model, the conditional effects in this model appear to be too

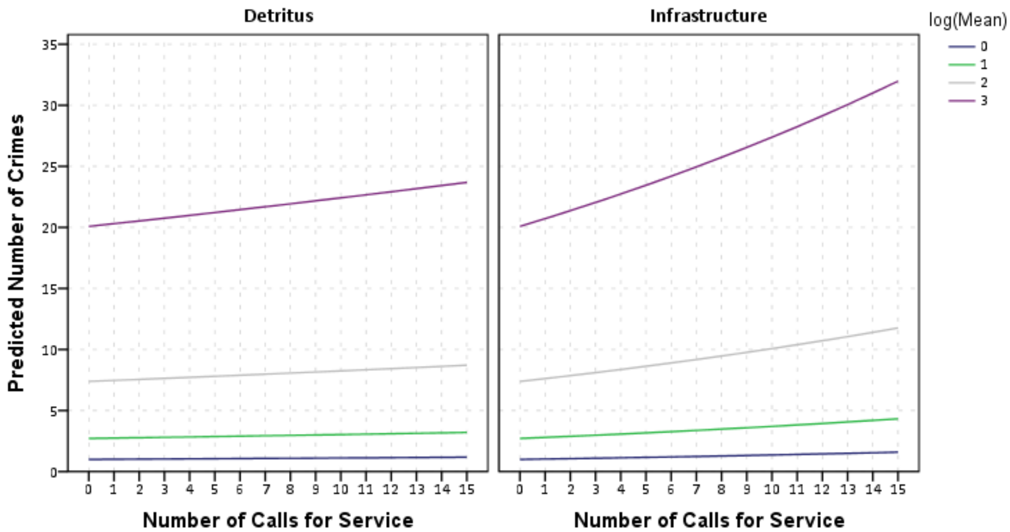


Figure 30: Predicted Marginal effects of 311 calls for service on crime at the street unit level.

small to be of practical utility. Tracing down the lines are of course a more appropriate way to think about crime policy implementation. That is, one would think about reducing the number of calls for service to decrease crime.

311 calls for service are of course simply proxies though for measures of physical disorder[5], and so the observational model should be taken with a heavy dose of scepticism. It is possible that other reasonable models estimate these effects to be larger (e.g. absent of the neighborhood fixed effects) or even estimate these effects simultaneously. For instance, if one thinks that *all calls for service* are just proxies for an underlying latent trait of disorder in the environment, estimating the conditional effects separately for detritus and infrastructure calls in the model could be misleading. If that is the case, one can't decrease the number of detritus related calls for service without also decreasing the infrastructure calls for service. On its face the correlation between the two measures is small enough that this assertion is not supported though. It is also the case in observational models like this, in which

[5]Otherwise the cynic may say they should simply stop taking messages, as the number of calls for service would decrease and the city would be a safer place!

the explanatory variables are not constructed by the investigator to be orthogonal that one can have indirect effects that are pertinent to any application. For example, Forsyth and Davidson (2010) find that liquor stores tend to have more detritus nearby, so adding a liquor store could theoretically have both a direct effect on crime and an indirect effect by increasing the amount of detritus.

While what is large or small effect may be a normative question, these marginal effects suggest that any policy implementation to decrease the number of physical disorder calls for service could only be expected to have a small effect on reported crimes. Also this shows it is likely that in any experimental study will be difficult to achieve a sufficiently powered study based on the low baseline rates for street units (Hinkle et al., 2014).

For illustration purposes I fit the same model using different definitions of neighborhoods. Although I previously mentioned that street units are only ambiguously identified as belonging to a particular census geography, here I treat them as they lay in the geographic file. That is, when identifying which polygon a street unit falls within, minor differences in the scales of the geographic data files will allow GIS software to determine which census geography a point belongs to, but this is only a false determination because of inconsistent geographic data. Here I ignore this inconsistency and pretend like the units fall within particular census geographies. For the few units that the GIS software did determine fell exactly on the line, I simply assign them to the census geography with the nearest centroid. So these results are useful as potential comparisons to what might have happened if the results were aggregated to different census geographies, but should not be interpreted as being more appropriate. Table 19 presents these results.

Statistical inferences on the effects do not change between models using different defini-

Table 19: Linear coefficient estimates on the log expected value of crime at the street unit for different aggregate neighborhoods

	Detritus		**Infra.**		**Bars Loc.**		**Bars Neigh.**	
Model	β	S.E	β	S.E	β	S.E	β	S.E
No Neighborhood	.022	.002	.068	.008	.532	.058	.208	.014
Grid Cells	.011	.002	.031	.006	.361	.039	.084	.011
Block Groups	.008	.002	.043	.007	.407	.048	.094	.011
Tracts	.007	.002	.047	.008	.451	.058	.115	.013

tions of neighborhoods, but the size of the effects diminishes going from no neighborhood to a model with neighborhood fixed effects for all the different neighborhood definitions. One might expect that with smaller neighborhood units the effects would be consistently smaller, but relative to their variances they appear quite similar. A strict test of coefficient differences between the models, e.g. $\text{H}_\text{o} : \beta_{\text{Grid Cells}} - \beta_{\text{Block Groups}} = 0$ (and the alternative hypothesis is that the difference does not equal 0), might presume that the two estimates are uncorrelated, and so the variance of the difference would be the sum of the variances (Paternoster et al., 1998). Using that as the denominator for a Wald test of the differences between model estimates using different neighborhood definitions for most contrasts in Table 19 would *fail* to reach statistical significance. For example, testing the effect of Detritus between grid cells and block groups, the z-score for the test would equal $\frac{.011-.008}{\sqrt{.002^2+.002^2}} = 1.06$. This makes the false presumption that the coefficients have no shared variance, which would *decrease* the standard error of the difference and increase the test statistic. So presuming no covariance between the coefficients biases one to fail to reject the null more often.

Including the neighborhood dummies, both for 311 calls for service and bars, diminishes the spatial effects to a much greater extent than the local effects. This is to be expected given the prior chapters detailing the process of aggregation bias. Neighborhood effects and spatial effects are generally confounded, and it is unlikely in noisy data that one will be able to identify data generating processes that strictly confirm one model or another. Also the lack of change between the local effects when aggregating up to the larger nested neighborhoods (e.g. from block groups to tracts) simply suggests that the covariances between points within neighborhoods are practically zero when going up to larger neighborhoods. The following section goes into further details about potential residual diagnostic tests that one can use to identify whether spatial autocorrelation still exists (even in the neighborhood fixed effects model), and a potential graphical test to see if a neighborhood model is appropriate given the data.

9.4 Residual Diagnostics

In the prior chapter it was suggested that one possible remedy to the problem of residual spatial autocorrelation was to include other explanatory variables in the model. One might suspect that including the fixed neighborhood effects in the subsequent reported models might correct for this, but this specification seems unlikely on its face to account for the complexities of spatial autocorrelation (Anselin and Cho, 2002; Anselin and Arribas-Bel, 2013). Here I will show how one can use a correlogram of the spatial residuals, and in particular a correlogram distinguishing between pairs of observations within neighborhoods versus between neighborhoods to identify the relevance of neighborhoods to explaining observed spatial autocorrelation.

It is typical in exploratory spatial data analysis to plot local indicators of spatial association (Anselin, 1995; Messner et al., 1999) to examine if the outcome of interest is spatially autocorrelated by showing a Moran's I scatterplot. If the outcome is z whose expectation equals 0, Anselin (1995) defines the *global Moran's I value to be*:

$$I = (n/S_0) \cdot \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2} \tag{9.19}$$

Where $w_i j$ are the elements of spatial weights matrix W, n are the number of observations, and S_0 is the sum of all the elements in W. In the case that z is mean centered and is standardized to have a variance of 1 (e.g. Z-scored) $\sum_i z_i^2$ will drop from the denominator of the final term as it is equal to 1. In the case that one row-standardized the spatial weights matrix W, this will in turn make $n = S_0$, so the first term drops from the equation, and then one is only left with $\sum_i \sum_j w_{ij} z_i z_j$. For every i observation of z, Anselin defines these to be the *local* Moran's values, as when summed up they equal the global Moran's I coefficient. One can then plot the values of z_i versus the local Moran values, (lets call them z_w) in a scatterplot, and in the case of no spatial autocorrelation the scatterplot should appear to be random. All z_w are in practice are the spatial averages of the neighbors, where what is a neighbor and what weight it receives in the averaging is dictated by your spatial weights matrix. If one estimates a regression equation of $z_w = \beta_1(z_i)$ forcing the intercept to be zero, the value of the slope coefficient β_1 can be used to test for global spatial autocorrelation, but more simply one can evaluate particular locations as being influential locations on the global spatial autocorrelation estimates.

One should notice in equation 9.19 that the numerator term is synonymous with an estimate of a *weighted* covariance (in the case that the variable has a value of zero for its mean). Thus the global Moran's I value can be thought of as a spatially weighted measure of the autocorrelation function that is more common in time series analysis. So here I will propose a synonymous use of the correlogram to that of the autocorrelation function plot (i.e. ACF plot) in time series analysis as a tool to diagnose if spatial autocorrelation exists at varying scales (McDowall et al., 1980; Negreiros et al., 2010). One can also use different specifications of W to test particular hypotheses about the structure of neighborhood effects.

Typically one considers the values of W to be fixed by the investigator before examining the data. This has the problem in practice though that social scientists rarely have enough theory to dictate the potential form W takes (Loftin and Ward, 1983). Here in subsequent diagrams I will suggest examining varying W at different distances, so *all* locations within a particular distance band are equally weighted neighbors and those outside of the distance band have zero weights.[6]

Here instead of global Moran's I I use a similar function to estimate the correlogram for values within specific distance bands (Ostrom, 1990). The estimate for the autocorrelation for distance lags of h can be written as $r_h = c_h/c_0$ where c_h is the covariance between two points (of a mean detrended series):

$$c_h = \frac{1}{n_{[i,i+h)}} \sum_{i=1} (z_i)(z_{i+h}) \tag{9.20}$$

And c_0 is the variance of the original series. Here I use $n_{[i,i+h)}$ to signify all of the pairs

[6]Note this procedure is different than those that examine the covariance between z and $W(z)$ for subsequently higher powers of W, e.g. W^1, W^2, W^3 etc. High powers of W is not the same thing as simply observations further away in space, and it amounts to diffusing the observations over multiple locations. This is appropriate for examining diffusion effects, but not appropriate for assessing residual autocorrelation in a model in which you question the actual construction of the weights matrix itself.

of observations within a distance band $[i, i+h)$ of each other. One can then estimate the autocorrelation for varying bands of distance and then make the typical ACF plot used in time series analysis.

With a row-standardized matrix, the expected value of equation 9.19 under the null of no spatial autocorrelation, takes the value of $\frac{-1}{n-1}$. So with 21,506 cases I presume the expectation of the autocorrelation is practically zero. Because of the skewed nature of the outcome (and its residuals) here I use the permutation approach suggested by Anselin (1995), and permutate the spatial labels for the original values of z_i and then subsequently re-estimate the autocorrelation value with the permutated data. This generates a reference distribution with which I evaluate the variance of the estimate under the null hypothesis. In any correlogram plot I then plot these traces for visual reference to the null distribution, so one can evaluate the variance of the statistic as well as whether the observed values deviate from those expected under the null and the size of that deviation.

Because the total number of neighbors for a set of 21,506 observations is too large in practice to manipulate (the total number of pairwise observations is $n(n-1)/2 = 231,243,265$) the strategy to identify the sample for each distance lag is as follows. For each observation i, I identified the nearest 1,000 neighbors along with their distance. This strategy assures that every individual observation has at *a minimum* 1,000 neighboring pairs, and given the large number of neighbors practically ensures that neighbors at smaller locations (where we expect to uncover the most of the spatial autocorrelation) are adequately represented. The smallest distance of the 1,000th neighbor in this D.C. dataset is 1,133 meters, so I *know* I have sampled *every possible* neighbor pair within 1,133 meters in this particular dataset. I then stacked these sets of nearest neighbors into two columns with the distance as a third column, and then eliminated redundant neighbor pairs. This allows a much simpler calculation of equation 9.20 without having to manipulate a square matrix of 21,506 rows and columns. I only need to merge in the values of z_i and z_j (for each pair of observations), and then perform a simple aggregation of the product $z_i \cdot z_j$ within a particular distance band to estimate the mean of the product. Creating the original set of neighboring observations is the most expensive operation (in terms of time). Once this is created, it is only slightly painful to estimate the autocorrelation for one (or multiple) values of z.

Once eliminating duplicate pairwise observations this produced over 12 million observation pairs. The number of neighbor pairs that the correlograms will be based on are 11,674,853 for distance bins of 50 meters between 0 and 2,000 meters, and larger distances are discarded. Figure 31 plots for the autocorrelation values on the Y axis for the original crime data (that is standardized to have a mean of zero and a variance of 1). On the X axis are sequential lags based on the distance between two points. The bright red line is the sample estimate, and the smaller grey lines are the same auto-correlation value estimated based on 19 permutations of the original crime data.

One can clearly see in Figure 31 that the autocorrelation over the range of distances (extending to 2,000 meters) is positive and much larger than any of the 19 permutations in the light grey lines, which hover around 0 for all distances in the plot. The autocorrelation peaks at the 100 to 150 meter bin, and then gradually declines until around 1,300 meters, where the correlation flattens at a value slightly below .02 and then gradually rises again.

Because of the difference in calculating the auto-correlation in the bands here, it is likely not appropriate to compare the size of Moran's I to previous criminological research as an effect size measure (this is more intended to be a residual diagnostic measure to visualize

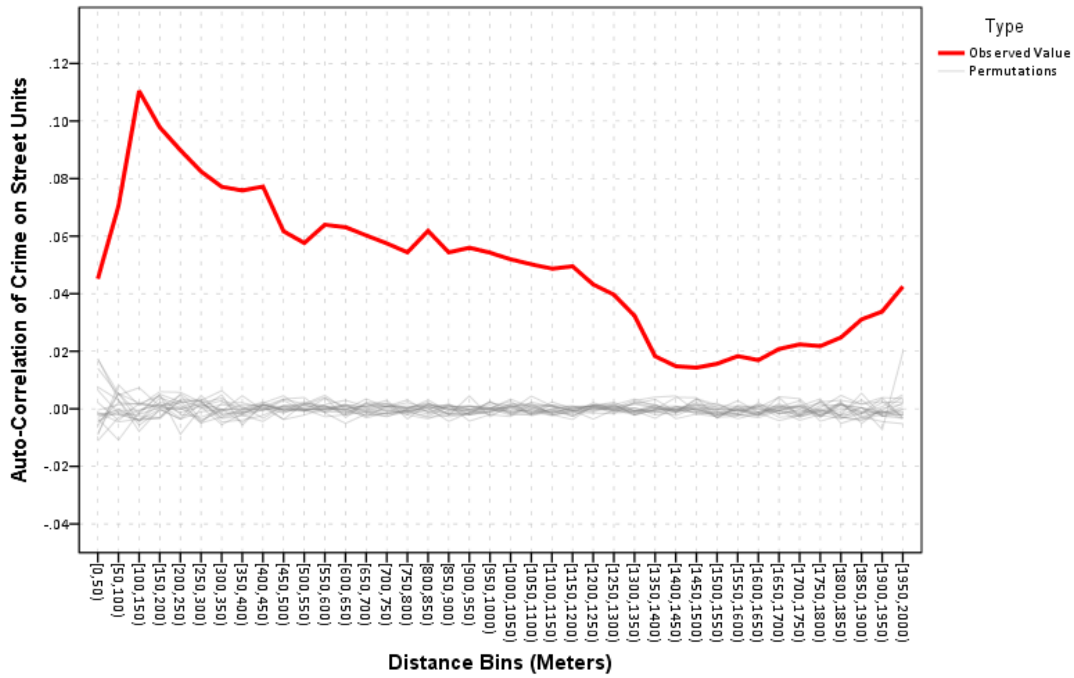


Figure 31: Correlogram based on spatial distances of 50 meters up to 2,000 meters. Auto-correlation is for the original crime data series.

whether spatial auto-correlation still exists in the model). But, the smaller auto-correlations appear to be consistent with prior research. In one example, for yearly county level homicide rates in the St. Louis metropolitan area, Messner et al. (1999) report Moran's I values between 0.02 and 0.17 over a 17 year period. Unlike the case of the time series data, small spatial auto-correlations of around .1 do not appear to be uncommon.[7]

The prior chapter mentioned that including additional covariates may account for the spatial auto-correlation in the data. So in Figure 32 plots the auto-correlations of the residuals from the model presented in Table 18, the model only including local effects of 311 calls for service and the neighborhood dummy variables.

Interestingly, we can now see a negative auto-correlation for the very short distance bins

[7]For time series models of count data Brandt et al. (2000) suggest the ACF plot is still a reasonable tool to diagnose a time series, and I so I do not suspect the low count nature of the crime data to have a strong impact on the spatial auto-correlations displayed here.

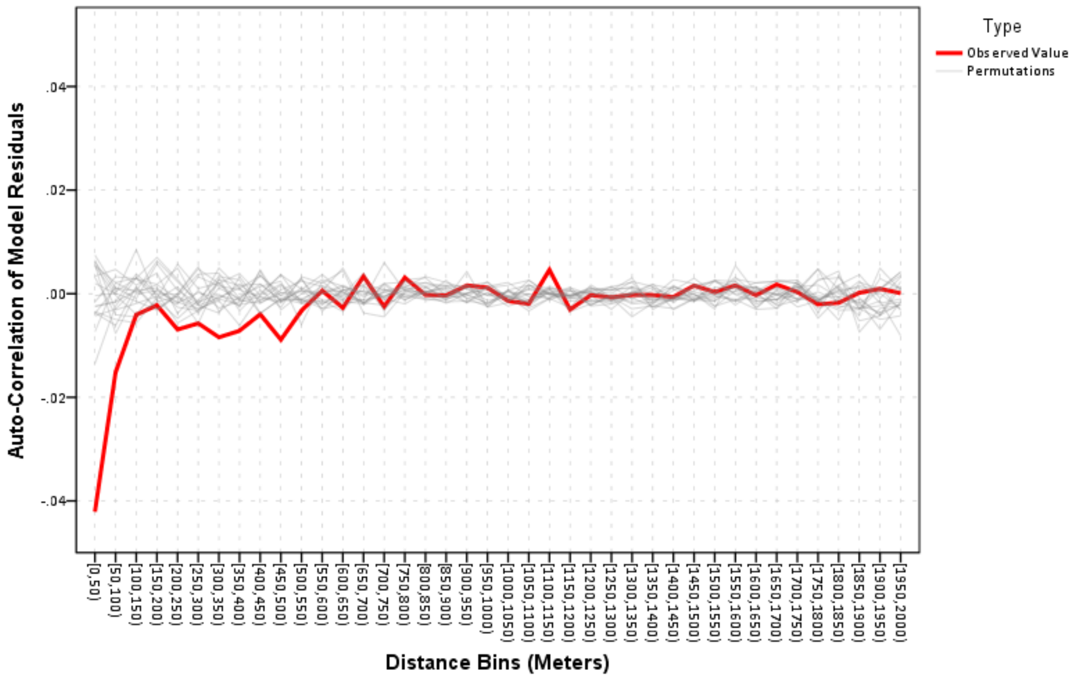


Figure 32: Correlogram based on spatial distances of 50 meters up to 2,000 meters. Auto-correlation is for the residuals from the neighborhood model on crime.

between 0 and 100 meters that are well outside of the 19 permutations. Negative spatial auto-correlation is extremely rare (see Tolnay et al. (1996) for the only counter-example I am aware of in social science data), and so I suspect this is more likely a result of the model mis-specification than it is a real phenomenon in the data. For instance, if one differences a time series which does not have a unit root, it will introduce a negative auto-correlation into the differenced series. Here I speculate the neighborhood dummy variables are playing a similar role. One potential example where this negative auto-correlation could be introduced by a series of neighborhood dummy variables is as follows: imagine the number of crimes on a street are only a linear function of the distance from the center of a city, and one estimates a series of aggregate neighborhood dummies over a regular grid (just as I have done in this analysis). Within the neighborhood boundaries the data will still be smooth and show positive spatial auto-correlation, *but* on the borders of the neighborhoods there will be

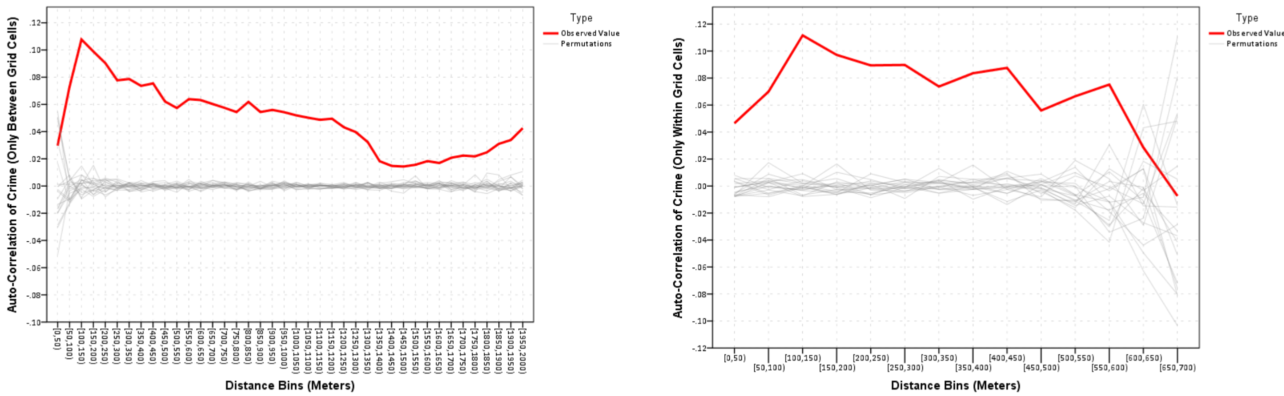
Figure 33: Example of how neighborhood dummies can cause negative auto-correlation if the underlying data are a smooth function of space.

extreme differences contributing negative values to the auto-correlation. Figure 33 shows an example of this for a fake set of data, where the plot on the left is the original crime data, and the plot on the right is the crime data differenced at a regular grid. The colors are normalized within each plot to range from 0 to 1, and one can see the neighborhood differenced data on the right introduces many irregularities that will spuriously cause negative auto-correlation (at least for some spatial weights matrices) compared the original smooth distribution on the left. If one only examines the auto-correlation between neighborhoods the negative auto-correlation should become more clear.

While it may be necessary to include the neighborhood fixed effects to identify particular effects, they do not appear to be a reasonable representation of reality in this particular sample, nor are they a cure for spatial auto-correlation (Anselin and Arribas-Bel, 2013). So this brings up the question, *what would a neighborhood effects model actually look like in the data?* A neighborhood effects model ends up being quite restrictive in how different pairs of data are correlated, in that only pairs of data *within* neighborhoods should share a correlation, where neighbors that do not share a neighborhood should be uncorrelated. Also within neighborhood pairs should be equally correlated, without regard to the actual distance between the two points. That is, two points that are 100 meters away should share the same correlation as those that are 500 meters away as long as they share the same neighborhood, but two points that are 100 meters away but do not share the same neighborhood should have no correlation with each other. Based on this notion, we can then decompose the correlograms previously shown to only estimate correlations for pairs within neighborhoods versus a correlogram for pairs between neighborhoods.

In a theoretical neighborhood model, a correlogram of the correlations between neighborhoods should be theoretically zero, while those within a neighborhood should be positive but flat. Figures 34a and 34b display the same correlograms decomposing the auto-correlations to between neighborhood pairs (on the left) and within neighborhood pairs (on the right). Here we can see that the between neighborhood model has a higher variance for the smaller distance bins, as very few points are a small distance away and have different neighborhood boundaries, but generally follows the same pattern as the overall correlogram. The within correlogram has a smaller variance for the shorter distances, but only extends to 700 meters. Table 20 lists the overall number of pairs that are used for the calculations in the correlograms at each distance bin. The maximum distance away two pairs of points could be in a square grid of 500^2 meters is under 708 meters. Before 1,300 meters these include all pairs within the population. After 1,300 meters it is only a sample of the potential pairs in the dataset.

The number of pair-wise comparisons for the cross correlations are quite large, so fairly



(a) (b)

Figure 34: The correlogram on the left displays the auto-correlations for the original crime data only between neighborhood grid cells. The auto-correlation plot on the right only displays auto-correlations for the same crime data within grid cells. The similarity of the plots suggests that the neighborhood boundaries used here have no substantive effect on the spatial auto-correlation of crime.

precise estimates can be calculated for all of these bins. Only in the very small bins are within neighborhood correlations the predominate type, and for distance bins of 200 meters or larger the sample size starts to swing towards being predominately between neighborhood grid cells. Distance bins for the within neighborhood correlograms have quite a large sample until over 600 meters. A distance of 600 meters is sufficient to evaluate the difference between spatial and neighborhood effects that are examined here. Spatial weights matrices for smaller windows are used, and often only use contiguity of the areal points to define the spatial effects (Bowers, 2013; Murray and Roncek, 2008), which can be variable but would likely be much less than 600 meters for micro place units. Some provide evidence of typically smaller windows of spatial effects. For instance, Ratcliffe (2012) finds that bars' influence on crime extends less than 100 meters. Groff (2013) suggests that bars influence crime up to 610 meters (2,000 feet), but use an inverse distance weighting method that would make bars further away contribute less to the local measure of bars.

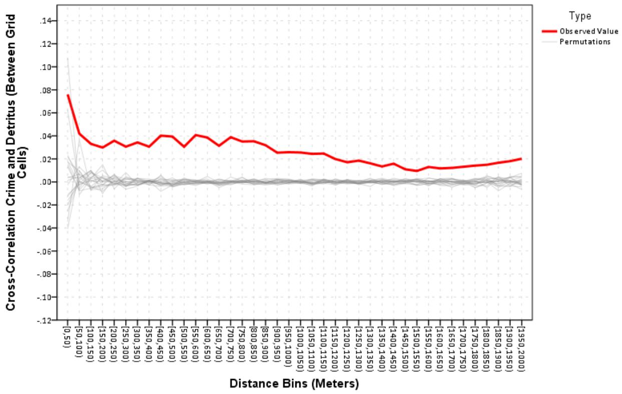
Table 20: Number of observation pairs used to calculate correlograms.

Distance Bin	Number Within	Number Between	Total
$[0, 50)$	14,508	1,323	15,831
$[50, 100)$	39,636	9,065	48,701
$[100, 150)$	53,478	22,758	76,236
$[150, 200)$	60,323	39,878	100,201
$[200, 250)$	61,276	62,611	123,887
$[250, 300)$	58,531	89,404	147,935
$[300, 350)$	51,989	118,053	170,042
$[350, 400)$	43,274	148,883	192,157
$[400, 450)$	31,227	182,937	214,164
$[450, 500)$	18,186	216,197	234,383
$[500, 550)$	6,796	247,987	254,783
$[550, 600)$	2,251	271,782	274,033
$[600, 650)$	547	293,852	294,399
$[650, 700)$	71	313,631	313,702
$[700, 750)$		332,197	332,197
$[750, 800)$		350,938	350,938
$[800, 850)$		368,242	368,242
$[850, 900)$		383,968	383,968
$[900, 950)$		401,763	401,763
$[950, 1000)$		418,931	418,931
$[1000, 1050)$		435,783	435,783
$[1050, 1100)$		453,037	453,037
$[1100, 1150)$		469,193	469,193
$[1150, 1200)$		483,587	483,587
$[1200, 1250)$		494,236	494,236
$[1250, 1300)$		486,604	486,604
$[1300, 1350)$		468,077	468,077
$[1350, 1400)$		437,564	437,564
$[1400, 1450)$		404,995	404,995
$[1450, 1500)$		378,433	378,433
$[1500, 1550)$		348,074	348,074
$[1550, 1600)$		313,405	313,405
$[1600, 1650)$		278,896	278,896
$[1650, 1700)$		239,271	239,271
$[1700, 1750)$		205,608	205,608
$[1750, 1800)$		174,683	174,683
$[1800, 1850)$		149,986	149,986
$[1850, 1900)$		124,591	124,591
$[1900, 1950)$		98,649	98,649
$[1950, 2000)$		82,815	82,815

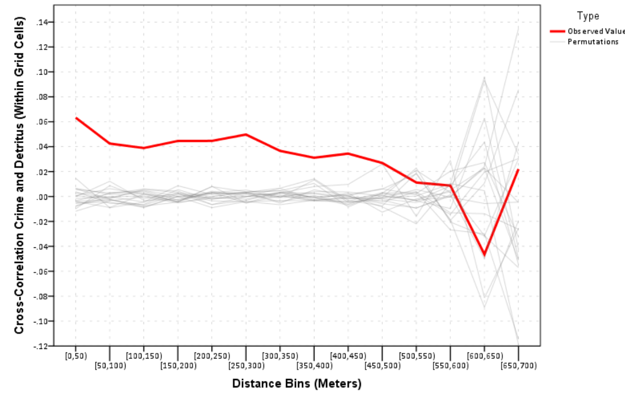
These correlograms can be extended to examine cross correlations between different variables and different spatial lags. So instead of calculating the covariance of a variable with itself at different lags, in Equation 9.20 one can replace the z_{i+h} term with a different variable, say y_{i+h}, and then calculate the covariance. Figure 35 presents several of these cross correlograms for the relationship between crime and detritus calls for service. The plots on the left display the within neighborhood correlations, and the plots on the right display the between neighborhood correlations. The figures in the top row show the marginal relationships (e.g. the cross correlations for crime and detritus calls for service), and the figures in the bottom row show the residuals from the model in Table 18 (i.e. regressing crime on the neighborhood dummy variables) and compares it to the residuals from regressing detritus 311 calls for service on all the same variables (minus detritus 311 calls for service). These residuals are then orthogonal to all of the other variables currently included in the models, and thus any residual correlation would have to be attributable to either spatial correlation between the two variables (or some other omitted variable in the model).

Figure 35 shows that the marginal relationships (in the top row) do not show any evidence of the neighborhood model. That is there is spatial auto-correlation both within and between neighborhoods, and those correlations decrease at further distances. The residual correlograms (in the bottom row) do not show the same pattern though, and have correlations within the set of 19 permutations.[8] Interestingly, there is evidence of the neighborhood over-differencing in the between neighborhood correlogram as I illustrated earlier. Within neighborhoods still have positive auto-correlation (as is shown in Figure 33), it is only on

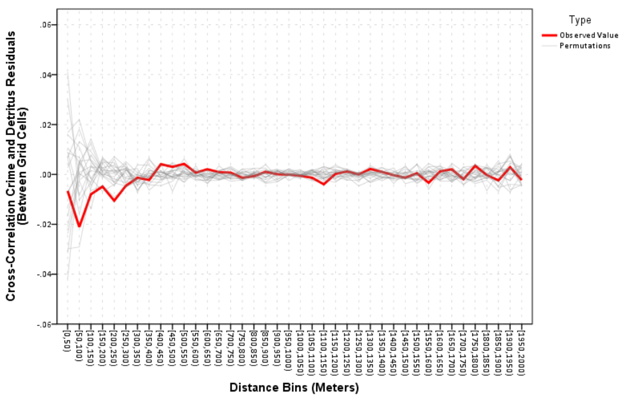
[8]It is unclear how the permutations should be conducted in the cross-correlation setting. Here I have permutated both of the original variables, although it may be reasonable to only permutate one of the variables.



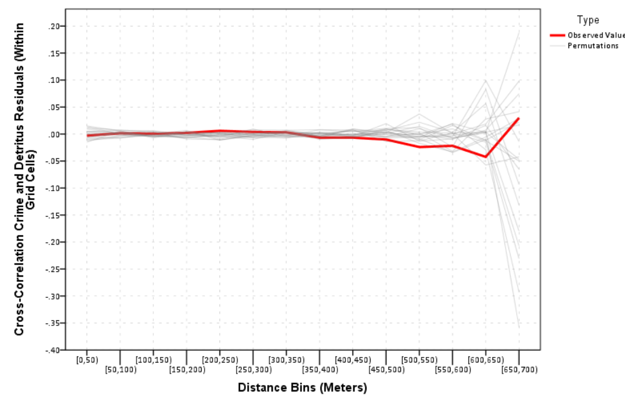
(a)



(b)



(c)



(d)

Figure 35: The cross correlograms on the left display the auto-correlations for the original crime data and detritus calls for service only between neighborhood grid cells. The auto-correlation plots on the right only displays auto-correlations for the same crime data within grid cells. The top row displays the cross-correlations for the original data series, and the bottom rows display the cross-correlations from the model residuals.

the borders of neighborhoods that this causes spurious negative auto-correlation. We can see some evidence of this negative auto-correlation in the plot on the lower right, but the observed correlations do not fall outside of the 19 permutations.

Figure 36 displays the same cross correlations for crime and infrastructure 311 calls for service with much the same results. The marginal variables do not show evidence of a neighborhood effects model, and the correlograms examining the partial residuals are all near zero. For the between neighborhood model the lag at [50, 100) is outside of the 19

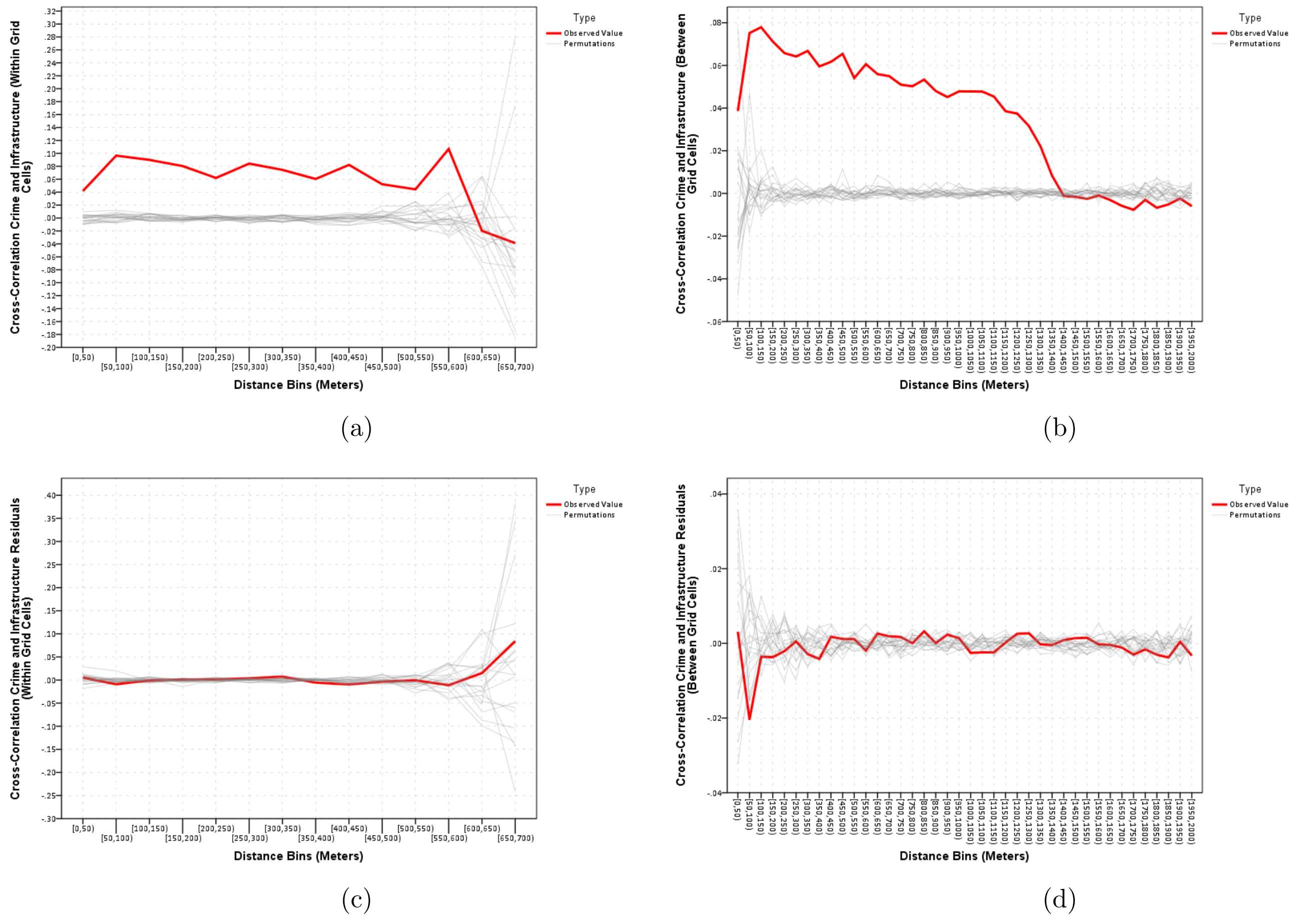


Figure 36: The cross correlograms on the left display the auto-correlations for the original crime data and infrastructure calls for service only between neighborhood grid cells. The auto-correlation plots on the right only displays auto-correlations for the same crime data within grid cells. The top row displays the cross-correlations for the original data series, and the bottom rows display the cross-correlations from the model residuals.

permutations, but this should only be taken as tentative evidence of negative auto-correlation given the variance of the spatial correlations observed in other plots. Again though it is indicative of the problem of neighborhood over-differencing previously discussed.

9.5 Conclusion

This chapter shows an example of how one can use a fixed effects model specification to identify local effects independent of neighborhood effects, as well as how one can distinguish between a spatial effects model and a neighborhood effects model using spatial correlograms. Here I find that the amount of detritus and infrastructure 311 calls for service have a consistent and small effect on the number of reported crimes for D.C. street units across multiple model specifications. So while this may be theoretically interesting, it is likely any policy implementation to reduce the underlying conditions that affect the number of 311 calls for service would be unlikely to have any perceivable effect on crime at small areas.[9]

I also show that the crime data at street units do not appear to reasonably follow a neighborhood effects model, and the introduction of neighborhood fixed effects appears to introduce an artefact of negative auto-correlation in the residuals. Because of this, I will not subsequently estimate any models using neighborhood fixed effects, as they do not appear to be a reasonable model for the crime data. This is not surprising, as a neighborhood model is very restrictive. It essentially states that units within a neighborhood are all similar (or monolithic) and this similarity is the same even if the street units are neighbors or are 700 meters away. As long as they are within the same neighborhood they should have the same covariance. This goes against Tobler's first law of geography (Tobler, 1970), and I find this model not likely to be reflective of any spatial data generating process in the social sciences.[10]

[9] I do not mean this to be an overly negative perspective on the relationship. La Vigne (1997) talks about how the Metro police are specifically taught to report physical problems with the subway in D.C. It would seem fairly easy to implement this above ground as well, and have police report infrastructure or detritus problems as part of there regular duties, with little added cost to their time. It is also possible if one evaluates such initiative city wide it may be possible to detect an effect on crime, but at small places it seems unlikely.

[10] I can imagine other situations in which monolithic neighborhoods may be more reasonable, such as

But, without the neighborhood fixed effects there still appears to be spatial autocorrelation in the residuals of the model. In the subsequent chapter I will include non-linear spatial terms as a function of the spatial coordinates of a street unit to attempt to account for this spatial auto-correlation in the model, along with several other predictors that vary at the street unit level to predict crime.

students within a classroom or a school. For spatial data though I believe monolithic neighborhoods to be unlikely.

Chapter 10

A General Model of Crime

Building on the findings from Chapter 8 and Chapter 9, I incorporate a series of additional place based covariates that predict the number of crimes at a street unit based on a variety of crime generator and attractor locations and include a set of non-linear spatial terms to account for the residual spatial auto-correlation. The motivation is simultaneously to identify whether the previous findings are spurious because of omitted variables in the model and to build a more reasonable model of crime at small places. More reasonable entails both including other crime generator locations that predict a large amount of human activity that appeared to be lacking in the prior models (e.g. a shopping mall), as well as explicit spatial modelling of the spatial trends in the data (e.g. crime was clustered in the center of the city compared to the periphery).

The first section of this chapter provides descriptive statistics for the variables that will be used in the model. Then follows the fitted negative binomial regression model, with and without non-linear spatial terms. This section also provides generic indicators of effect sizes in the form of incident rate ratios for arbitrary changes in the independent variables depending on their typical variation. The following section then discusses residual diagnostics, including the small amount of residual auto-correlation that still exists in the residuals, as well as discussion of outlier locations.

10.1 Descriptive Statistics

Table 21 provides univariate descriptive statistics for the variables that are included in the model excluding the non-linear restricted cubic spline terms for the area of the Thiessen polygon and the spatial coordinates. Included in the table are the mean and standard deviation, as well as the minimum, maximum, the sum of all areas, and the 25th, 50th and 75th percentiles of the distribution. These measures are for the total 21,506 street unit locations, and no missing data exist.

The logged areas (Thiessen polygon, road area, and sidewalk area) are calculated as continuous variables of the areas that intersect the Thiessen polygon. Areas for school, recreation area, university, public housing, or park are calculated as dummy variables. Calculated as continuous measures, these areas tended to be very bimodal in terms of percentage area. That is, the locations either tended to not intersect at all or were entirely covered by such locations. It is more likely that being close to such an area is predictive of crime, but I find the area being covered more by the associated generator area type is unlikely to additively add to the expected number of crimes. For this reason I recoded these areas into dummy variables.

I decided to split up alcohol licenses into their constituent types in this model, liquor stores, grocery or convenience stores, and bars or restaurants. The spatial lags of each of

Table 21: Univariate Descriptive Statistics (n = 21,506)

Variable	Mean	Std. Dev.	Min.	Max.	Sum	25th	50th	75th
Crime	1.5	4.4	0	249	32,440	0.0	0.0	1.0
Liquor Stores	0.0	0.1	0	2	211	0.0	0.0	0.0
Neighbor Liquor Stores	0.1	0.2	0	3	1,244	0.0	0.0	0.0
Grocery or Conv. Store	0.0	0.1	0	2	267	0.0	0.0	0.0
Neighbor Grocery Stores	0.1	0.3	0	3	1,529	0.0	0.0	0.0
Bars and Restaurants	0.0	0.4	0	20	1,071	0.0	0.0	0.0
Neighbor Bars and Restaurants	0.3	1.4	0	33	6,969	0.0	0.0	0.0
Detritus 311	4.3	7.1	0	83	91,527	0.0	1.0	6.0
Neighbor Detritus 311	28.3	30.2	0	270	609,347	5.0	19.0	41.0
Infrastructure 311	1.1	2.0	0	44	23,833	0.0	0.0	1.0
Neighbor Infrastructure 311	7.0	7.2	0	106	150,622	2.0	5.0	10.0
Litter Cans	0.2	0.7	0	12	5,265	0.0	0.0	0.0
Toxic Release Sites	0.0	0.0	0	4	19	0.0	0.0	0.0
Vacant Property	0.0	0.2	0	8	544	0.0	0.0	0.0
Green Sites	0.0	0.2	0	6	638	0.0	0.0	0.0
Trees	5.6	4.9	0	52	120,533	1.0	5.0	8.0
Street Lights	3.1	2.4	0	47	67,438	2.0	3.0	4.0
Bus Stops	0.2	0.5	0	9	3,482	0.0	0.0	0.0
Metro Entrance	0.0	0.1	0	3	90	0.0	0.0	0.0
Halfway House	0.0	0.0	0	1	9	0.0	0.0	0.0
HIV Clinic	0.0	0.1	0	3	122	0.0	0.0	0.0
Hospital	0.0	0.0	0	1	16	0.0	0.0	0.0
Library	0.0	0.0	0	1	28	0.0	0.0	0.0
Places of Worship	0.0	0.2	0	4	896	0.0	0.0	0.0
Police Station	0.0	0.0	0	1	15	0.0	0.0	0.0
Shopping Center	0.0	0.0	0	1	27	0.0	0.0	0.0
Sidewalk Caf	0.0	0.2	0	10	452	0.0	0.0	0.0
Wireless Hot Spot	0.0	0.1	0	2	326	0.0	0.0	0.0
Intersection	0.4	0.5	0	1	8,172	0.0	0.0	1.0
Log of Area	8.5	0.9	0	13	183,071	8.0	8.5	9.0
Log of Road Area	7.1	0.8	0	11	152,476	6.7	7.1	7.5
Log of Sidewalk Area	5.4	1.5	0	10	115,994	5.0	5.6	6.2
School Area	0.1	0.2	0	1	1,306	0.0	0.0	0.0
Recreation Area	0.0	0.1	0	1	466	0.0	0.0	0.0
University Area	0.0	0.2	0	1	607	0.0	0.0	0.0
Public Housing Area	0.0	0.1	0	1	401	0.0	0.0	0.0
Park Area	0.1	0.3	0	1	2,404	0.0	0.0	0.0

these measures it still included in the model, as well as the spatial lags for the 311 calls for service.

Table 22 provides the bivariate correlations for each of the same variables.

10.2 Model Fitting

Table 23 presents the model estimates for negative binomial regression model with the full set of explanatory variables. The first set of coefficient estimates are the linear predictors

Table 22: Bivariate Correlations

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
1	Crime																																					
2	Liquor Stores	0.07																																				
3	Neighbor Liquor Stores	0.11	0.00																																			
4	Grocery Stores	0.10	0.11	0.07																																		
5	Neighbor Grocery Stores	0.14	0.07	0.20	0.03																																	
6	Bars and Restaurants	0.25	0.17	0.16	0.09	0.09																																
7	Neighbor Bars and Restaurants	0.22	0.11	0.28	0.06	0.17	0.43																															
8	Detritus 311	0.13	-0.02	0.03	0.01	0.10	-0.04	-0.03																														
9	Neighbor Detritus 311	0.13	0.01	0.03	0.04	0.12	-0.03	-0.06	0.63																													
10	Infrastructure 311	0.15	0.03	0.08	0.04	0.10	0.10	0.13	0.34	0.26																												
11	Neighbor Infrastructure 311	0.23	0.05	0.13	0.06	0.18	0.14	0.22	0.34	0.50	0.37																											
12	Litter Cans	0.13	0.14	0.13	0.13	0.10	0.19	0.23	-0.06	-0.03	0.13	0.10																										
13	Toxic Release Sites	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	0.00	-0.01	0.01																									
14	Vacant Property	0.04	0.01	0.01	0.01	0.02	0.00	-0.01	0.14	0.14	0.06	0.07	-0.01	0.00																								
15	Green Sites	0.09	0.03	0.04	0.02	0.01	0.14	0.15	-0.03	-0.05	0.07	0.08	0.09	0.00	0.00																							
16	Trees	0.18	0.01	0.03	0.03	0.08	0.03	0.07	0.37	0.35	0.25	0.30	0.10	0.00	0.05	0.02																						
17	Street Lights	0.22	0.03	0.09	0.04	0.10	0.10	0.14	0.32	0.32	0.29	0.33	0.15	0.00	0.06	0.06	0.50																					
18	Bus Stops	0.06	0.07	0.05	0.06	0.04	0.05	0.06	-0.07	-0.05	0.06	0.00	0.40	-0.01	-0.02	0.03	0.06	0.10																				
19	Halfway House	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00																			
20	HIV Clinic	0.06	0.02	0.03	0.02	0.04	0.06	0.04	0.00	0.02	0.04	0.05	0.03	0.00	0.01	0.05	0.02	0.03	0.02	0.03																		
21	Hospital	0.02	0.00	-0.01	0.00	-0.01	0.01	0.00	-0.01	-0.01	0.00	-0.01	-0.01	0.00	0.00	0.00	-0.01	-0.01	0.00	0.00	0.10																	
22	Library	0.02	0.01	0.01	0.02	0.02	0.01	0.01	-0.01	-0.01	0.01	0.00	0.03	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00																
23	Metro Entrance	0.03	0.03	0.03	0.02	0.03	0.07	0.07	-0.02	-0.03	0.06	0.04	0.10	0.00	-0.01	0.03	-0.01	0.03	0.12	0.00	0.01	0.00	0.03															
24	Places of Worship	0.04	0.02	0.03	0.04	0.05	0.01	0.02	0.03	0.08	0.03	0.07	0.04	0.00	0.03	0.02	0.06	0.03	0.02	0.02	0.07	-0.01	0.00	-0.01														
25	Police Station	0.01	0.00	0.03	0.00	0.01	0.03	0.05	0.01	0.00	0.02	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01													
26	Shopping Center	0.09	0.08	0.03	0.03	0.04	0.09	0.03	-0.01	-0.01	0.02	0.02	0.04	0.00	0.00	0.01	0.00	0.03	0.04	0.00	0.01	0.00	0.00	0.07	0.00	0.05												
27	Sidewalk Caf	0.13	0.11	0.13	0.06	0.06	0.60	0.34	-0.03	-0.03	0.08	0.11	0.16	0.00	-0.01	0.10	0.03	0.07	0.05	0.00	0.02	0.00	0.01	0.06	0.00	0.01	0.02											
28	Wireless Hot Spot	0.06	0.00	0.02	0.00	0.04	0.02	0.02	-0.01	0.01	0.06	0.04	0.11	0.00	0.00	0.15	0.03	0.01	0.03	0.00	0.06	0.00	0.25	0.03	0.01	0.01	0.01	0.01										
29	Intersection	-0.11	-0.01	-0.03	0.01	-0.05	-0.04	-0.03	-0.15	-0.14	-0.01	-0.19	0.15	0.00	-0.03	-0.02	0.02	0.01	0.16	-0.02	-0.01	0.00	-0.02	0.01	-0.04	-0.01	-0.01	-0.02	-0.03									
30	Log of Area	0.14	0.00	0.02	0.00	0.03	0.02	0.03	0.14	0.10	0.11	0.16	0.02	0.04	0.02	0.11	0.29	0.38	0.05	0.00	0.02	0.03	0.00	0.03	0.01	0.01	0.03	0.01	0.08	-0.09								
31	Log of Road Area	0.17	0.04	0.10	0.04	0.08	0.06	0.09	0.06	0.04	0.13	0.15	0.12	0.03	0.01	0.10	0.26	0.43	0.15	0.00	0.04	0.03	0.01	0.04	0.04	0.02	0.04	0.03	0.07	0.12	0.69							
32	Log of Sidewalk Area	0.18	0.05	0.09	0.06	0.10	0.10	0.15	0.10	0.10	0.14	0.19	0.21	0.01	0.02	0.10	0.32	0.27	0.17	0.01	0.04	0.02	0.02	0.06	0.07	0.02	0.03	0.07	0.09	0.08	0.27	0.40						
33	School Area	0.05	-0.02	-0.01	0.00	0.00	-0.02	-0.02	0.00	0.04	0.02	0.04	0.09	0.00	0.02	0.05	0.08	0.03	0.05	0.00	0.00	-0.01	0.01	0.00	0.01	0.00	0.00	-0.01	0.21	0.02	0.09	0.06	0.13					
34	Recreation Area	0.02	-0.01	-0.02	-0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.24	0.00	0.00	0.02	0.05	0.02	0.00	0.01	0.00	0.00	0.00	0.00	-0.01	0.01	-0.01	0.00	0.10	0.00	0.07	0.02	0.08	0.18				
35	University Area	0.01	0.00	0.01	0.01	0.02	0.01	0.03	-0.06	-0.08	-0.01	-0.02	0.00	0.00	-0.01	0.02	-0.02	0.01	0.02	0.00	0.02	0.04	-0.01	0.01	-0.01	0.00	0.01	0.00	0.00	0.00	0.07	0.08	0.10	0.00	0.02			
36	Park Area	0.06	0.00	-0.01	0.00	0.00	-0.02	-0.03	0.01	0.00	0.00	0.00	0.02	0.00	0.00	-0.01	0.04	0.05	0.04	0.01	0.00	0.00	0.00	-0.01	0.01	0.02	0.00	-0.01	0.04	0.00	0.05	0.04	0.08	0.03	0.03	0.01		
37	Public Housing Area	-0.04	-0.03	-0.04	-0.02	-0.03	-0.04	-0.06	-0.09	-0.12	-0.06	-0.09	-0.01	0.01	-0.03	0.04	-0.08	0.00	0.03	-0.01	-0.01	0.00	0.00	-0.01	-0.03	-0.01	-0.01	-0.02	0.00	0.06	0.30	0.13	-0.06	-0.01	0.06	-0.03	-0.02	

Table 23: General Model of Crime Coefficient Estimates

Variable	**Model 1**			**Model 2**					
	B	Std. Err.		B	Std. Err.	Exp(B)	Low CI	High CI	
Liquor Store	.4	.23		.4	.24	1.5	.9	2.4	
Liq. Store Neigh.	.2	.07	**	.2	.07	1.2	1.0	1.4	**
Grocery Store	.4	.13	***	.3	.12	1.3	1.1	1.7	**
Groc. Store Neigh.	.2	.05	***	.1	.05	1.1	1.0	1.3	***
Bars	.1	.04	***	.1	.04	1.1	1.0	1.2	***
Bars Neighbor	.1	.02	***	.1	.02	1.1	1.1	1.1	***
Detritus 311	.0	.00	***	.0	.00	1.0	1.0	1.0	***
Detritus Neigh.	.0	.00	***	.0	.00	1.0	1.0	1.0	
Infra. 311	.0	.01	***	.0	.01	1.0	1.0	1.0	**
Infra. Neigh.	.0	.00	***	.0	.00	1.0	1.0	1.0	***
Litter Cans	.1	.03	***	.1	.03	1.1	1.1	1.2	***
Toxic Release Site	-.3	.16		-.3	.14	.7	.5	.9	**
Vacant Housing	.1	.06		.0	.05	1.0	.9	1.1	
Green Site	.0	.05		.0	.06	1.0	.9	1.2	
Trees on Street	.0	.00	***	.0	.00	1.0	1.0	1.0	***
Lights on Street	.1	.01	***	.1	.01	1.1	1.0	1.1	***
Bus Stop	.1	.04		.0	.04	1.0	1.0	1.1	
Halfway House	.2	.35		-.1	.35	.9	.5	1.8	
HIV Clinic	.3	.17		.2	.18	1.3	.9	1.8	
Hospital	1.1	.50	**	1.4	.61	4.2	1.3	14.0	**
Library	.3	.32		.5	.38	1.7	.8	3.6	
Metro Entrance	.0	.14		.1	.17	1.2	.8	1.6	
Place of Worship	.1	.06	**	.0	.05	1.0	.9	1.2	
Police Station	-.3	.42		-.2	.39	.8	.4	1.7	
Shopping Center	.4	.27		.8	.35	2.2	1.1	4.3	**
School	.4	.07	***	.2	.07	1.3	1.1	1.4	***
Recreation Area	-.1	.10		-.1	.11	.9	.8	1.1	
University Area	.1	.10		.1	.10	1.1	.9	1.4	
Public Housing Area	.6	.09	***	.5	.09	1.6	1.3	1.9	***
Park	.0	.08		-.1	.08	.9	.8	1.1	
Intersection	-.7	.05	***	-.7	.05	.5	.4	.5	***
Sidewalk Cafe	.0	.07		.0	.08	1.0	.9	1.2	
Wireless HotSpot	.2	.11	**	.2	.10	1.2	1.0	1.5	
Road Area (log)	.5	.05	***	.4	.05	1.5	1.4	1.7	***
Sidewalk Area (log)	.3	.02	***	.2	.02	1.3	1.2	1.3	***
Area Poly. (log)	-.5	.22	**	-.4	.17	.7	.5	1.0	**
Intercept	-1.9	1.54		7.9	2.75				***
Dispersion	2.9	.05		2.6	.05				

* is significant at .10 level, ** is significant at .05 level, and *** is significant at .01 level

without including the set of spatial spline terms, and the second model includes these spatial terms (effect estimates not shown). The restricted cubic spline estimates for the log of the area of the Thiessen polygon are not shown in either set of coefficients. The coefficients do not fluctuate greatly between the different model specifications. The display rounds the digits to either two or three significant digits, and so some effects are reported as having a standard error of zero, as these are rounded down below 0.05. Also this causes some estimates of the 95 percent confidence interval to appear to have zero length.

Figure 37 displays the exponentiated coefficients, incident rate ratios, and a 95 percent confidence interval for the count variables in the model that can be summarized by one term (Kastellec and Leoni, 2007). All terms are displayed as going from the value of zero to the value of one *except* for detritus calls for service, neighboring infrastructure 311 calls for service, trees, and street lights. The incident rate ratios for these variables are coded as going from 0 to 5. For neighboring detritus 311 calls the incident rate ratio is displayed for going from 0 to 20 calls for service in the neighboring areas. These values were chosen arbitrarily to reflect the more typical variance in the covariates so the incident rate ratios are more comparable. Confidence intervals for the 0-1 values are Wald confidence intervals provided directly by SPSS, the others are calculated using the delta method.

With an average of 1.5 crimes per street unit, an incident rate ratio of 2 can be interpreted as increasing the expected number of crimes at the typical street unit to 3. A more typical effect size of an incident rate ratio of 1.5 would increase the expected number of crimes to 2.25. So even with the largest effects the absolute terms with the total number of crimes increased (or decreased) still tends to be quite small in magnitude for the average number of crimes on a street. Going the opposite way, an incident rate ratio of 0.5 is the same relative decrease as an incident rate ratio of 2 is an increase. Because of this the ratios are plotted on a logarithmic scale (Galbraith, 1988). A decrease in the incidence of crime by a rate of 0.5 for a street with 1.5 crimes is 0.75 — so is cut in half instead of doubled. A decrease of 0.75 would result in an expected number of crimes to be 1.125.

The effects for the alcohol licenses, liquor stores, grocery stores, and bars, are all positive and similar in magnitude. Liquor stores have the largest point estimate for the effect, but the confidence interval is wide and covers the null effect of 1. The point effect of liquor store

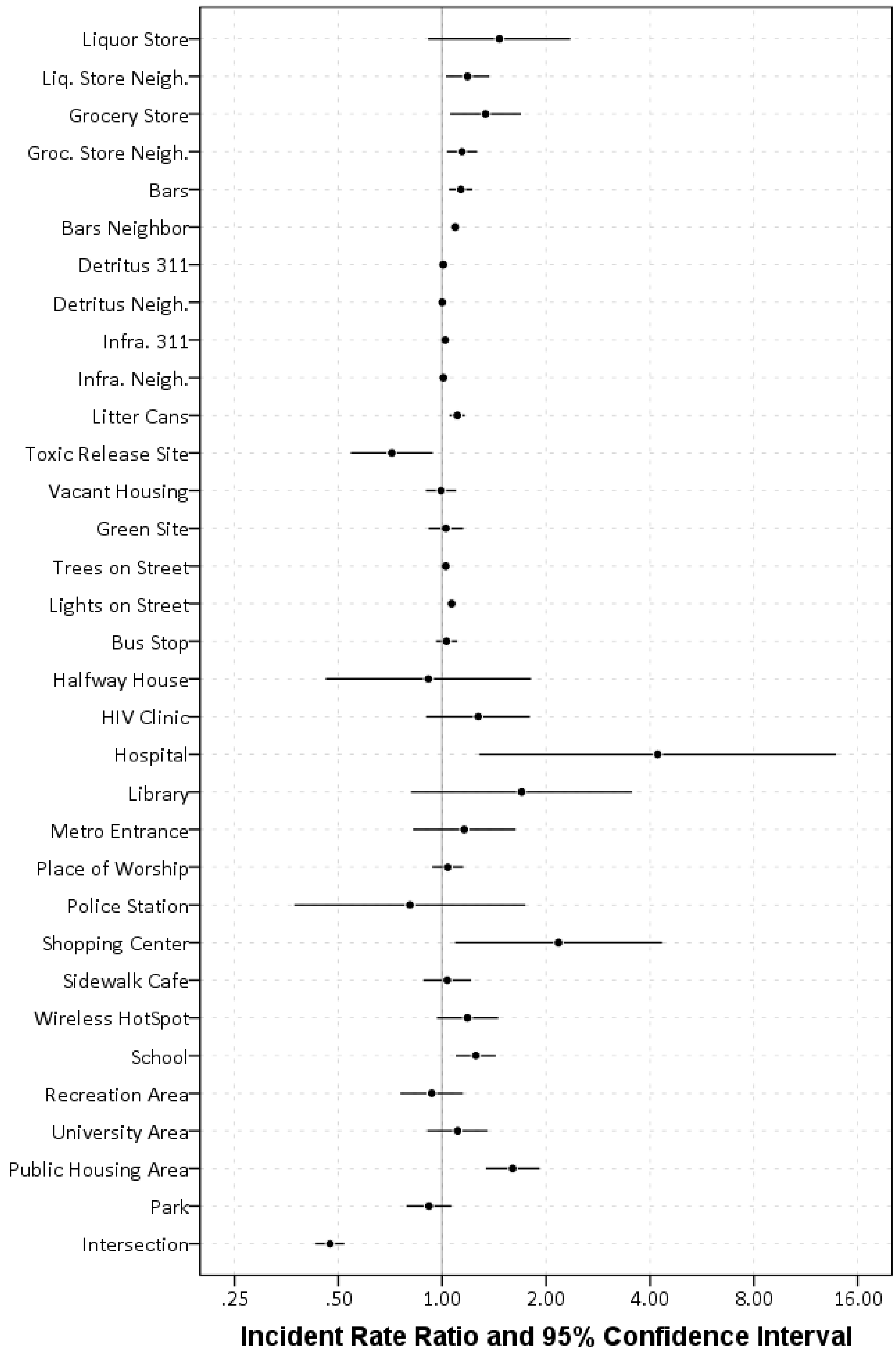


Figure 37: Incident rate ratios for model estimates predicting crime. All incident rate ratios are for going from 0 to 1 except for detritus related 311 calls for service, trees and stop lights. The incident rate ratio are going from 0 to 20 for the neighboring total of detritus 311 calls, and 0 to 5 for the other effects.

neighbors is smaller but the confidence interval does not cover 1. The total local effect of these licenses are smaller then when they were aggregated in previous models. The previous estimates of the local effect for all licenses was around 0.5 (for the linear predictor), and are 0.4, 0.3 and 0.1 for liquor, grocery and bars respectively, for an average effect of 0.27. In the case these components were orthogonal we would expect the average disaggregated effects added to be near the same size as the total effect, but in this sample they all have a slight positive correlation with one another so the reduction in effect sizes is not surprising (in addition to the other controls added into the model).

The neighbor effects for alcohol licenses are slightly smaller when disaggregating the different alcohol license types. In prior models the neighbor effect was around 0.22 in the linear predictor, and here are 0.2, 0.2 and 0.1 in the neighbor alcohol license effect for liquor stores, grocery stores and bars respectively. Again the neighboring effects appear to be *larger* than the local effects when considered how they diffuse to multiple adjacent regions. Given the standard errors of the local and neighbor effects the prior aggregation is not misleading. The effects are both in the same direction and in the same general order of size.

The average local effect of bars in this model is then around $\exp(0.3) \approx 1.3$, so if one added a bar to a location with an expected number of crimes being 1.5, the increase in crimes *on the local* street would be around half a crime, with the total expected to be close to 2 crimes. The increase in crimes on any *single* neighboring street is around $\exp(0.2) \approx 1.2$, and if the neighboring street has an expected number of 1.5 crimes the increase would be around 0.3 crimes for a total expected number of crimes being 1.8. Because this effect will diffuse to multiple streets though, typically around 6 to 8, this spatial diffusion will typically be larger combining all of the neighboring streets than the local effect. Going with a low estimate of the typical number of neighbors being 6, a typical increase in the total number of crimes when adding one liquor license will be slightly over 2 crimes combining both the local and the neighboring effect.[1]

The variables related to physical disorder, 311 calls for service, litter cans, toxic release sites, and vacant housing are all quite small and very close to 1 except for toxic release sites, which has a negative effect on crime. This is likely because the toxic release sites are industrial locations with little foot traffic. Although several of the coefficients for the 311 calls for service are estimated very precisely, they have nearly imperceivable effects on crime in this sample.

Measures of gentrification in this study (Wireless hot spots and coffee shops) have no obvious affect on crime; their confidence intervals cover 1. Some of the other variables that have been consistently found to affect crime, such as parks and bus stops, do not appear to affect crime in this cross-sectional sample. Bus stops are highly correlated with the inner parts of the city, and thus it is plausible the spatial trend terms removed the partial correlation in the subsequent model. The same could be said for coffee shops. The discretization of parks into multiple units may be partially responsible for the reason that parks are not correlated with crime when controlling for other factors. Shopping malls could be also be considered a proxy for gentrification, but they have a strong positive effect on crime. The generator effect of malls (e.g. they attract many of individuals) is surely much larger than any theoretical gentrification effect.

[1] As a robustness check to see if including intersections had any impact on the findings, I estimated interaction effects of intersections with local licenses, neighboring licenses, and local 311 calls for service (both Detritus and Infrastructure). All of the interactions were null *except* for intersection and neighboring licenses, which was positive. This is slight evidence that not including intersections slightly *decreases* the diffusion effects. This makes sense if intersections are more likely for crimes recorded outdoors, and so discarding intersections one would be discarding many crimes not at the local institution, but nearby.

Locations such as hospitals and police stations were included because these are often idiosyncratic locations where police record crimes as occurring. Hospitals are sometimes listed for assaults even if the assault occurred elsewhere and the victim was subsequently transported to the hospital. Police stations are similarly listed as the de facto location for certain crime complaints (e.g. if someone comes in person to file a report) although this occurs more often for minor crimes in the author's experience.

A street unit having a hospital on the premises is the largest effect in the model, with an increase in the incident rate ratio of over 4. The effect is quite imprecise though, and the confidence interval ranges from 1.3 to 14. Similarly the effect of a shopping center is quite large, with an incident rate ratio of over 2. Simply having a set of locations that predicts a large number of people in the area appear to be strong predictors of the number of crimes at small places.

10.3 Residual Diagnostics

As the same in prior presented models, the fit of the negative binomial model compared to the observed density in Figure 38 is quite good. There appears to be no need for zero inflated models, as the predicted number of zeroes given the negative binomial regression model fits the observed number of zeroes quite well. The tails of the distribution are predicted quite closely as well.

This section subsequently examines if any spatial autocorrelation still exists in the residuals of the model and the existence of outlier locations.

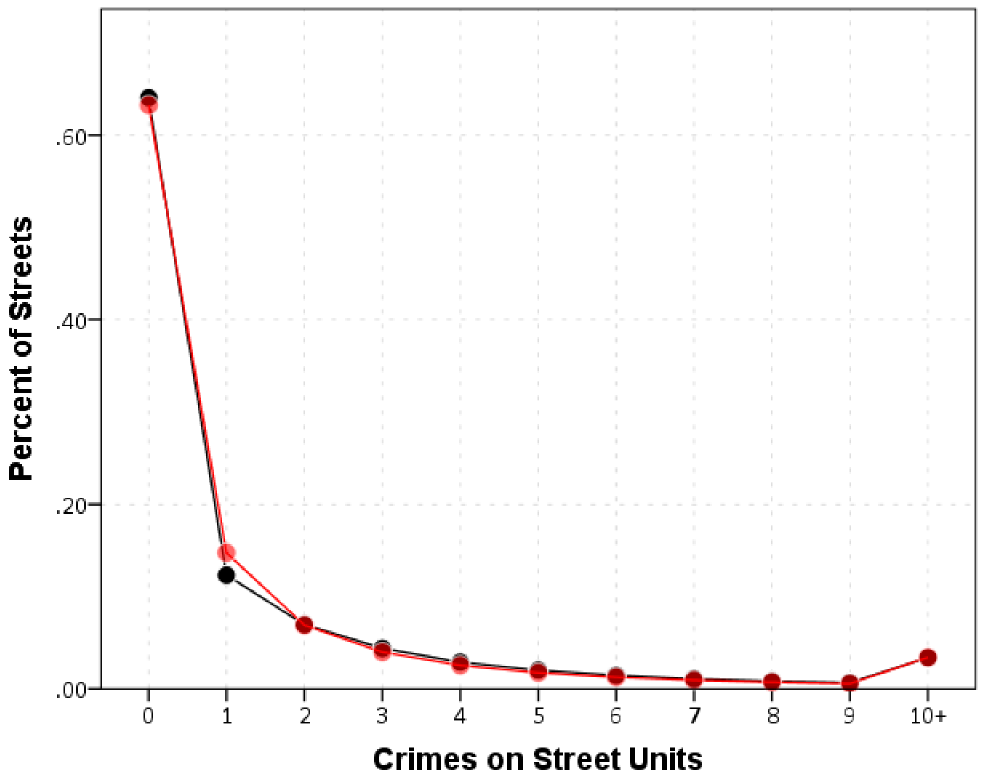


Figure 38: The fit of the negative binomial model is quite good. Only slightly under predicting zeroes and over predicting ones.

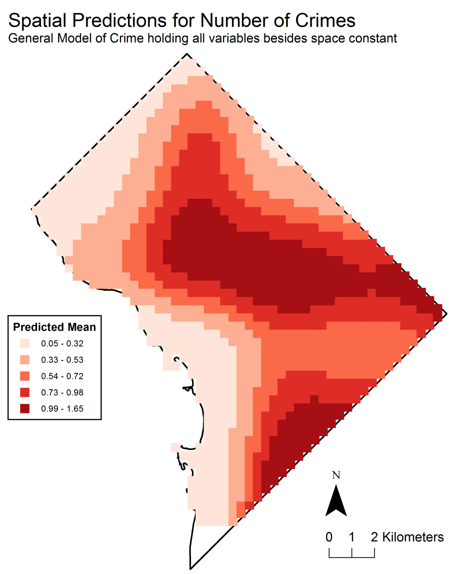
10.3.1 Residual auto-correlation

In the prior chapters, residual auto-correlation after fitting models was quite apparent, with Moran's I values between 0.05 and 0.15 depending on what types of residuals from the generalized linear model were examined. For this reason, fixed spatial terms were included in the model to detrend the data, and these terms consisted of restricted cubic splines of the projected x and y coordinates in meters as well as their interaction. This approach is more common in the natural sciences, but it has a nice analogue to the work of Shaw and McKay and the concentric rings model of crime emanating from the city center.

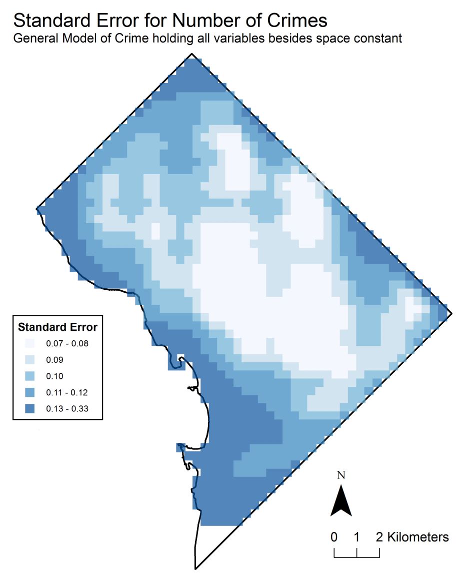
Figure 39 displays the ability of those terms to detrend the data. In the plot on the left are the mean predictions of crime over a regular grid holding covariates constant (for 0-1 or count covariates they are set to 0, for continuous covariates they are set to the mean). Comparing this to prior kernel density maps of crime shows a very similar pattern, and the terms are quite flexible enough to take the general trend of crime in particular areas of the city away. The map on the right displays the standard error of the predictions. The fixed spatial terms also have the advantage that they model the heteroscedasticity of the edge of the city explicitly.

Mapping the residuals of the model it is certainly an improvement over prior iterations. The residual map appears at first glance to be quite random. Presented in Figure 40 are the deviance residuals. Compared to previous maps, there are not strong clusters of under or over predictions in the model.

Examining the Moran's I value of the residuals using the same weights matrix for contiguity of the Thiessen polygons does reveal a slight amount of residual spatial auto-correlation



(a)



(b)

Figure 39: These plots display the predicted mean of crime and the standard error of those predictions holding all variables in the model at constant values. All count and 0-1 variables are set to zero, and all area variables are set to their means. The predicted surface is then a function of the non-linear spatial restricted cubic spline terms, as is the predicted value holding these other covariates constant. This approximates the general trend of crimes across the city very well, and explicitly allows the data at the edge of the city to be heteroscedastic.

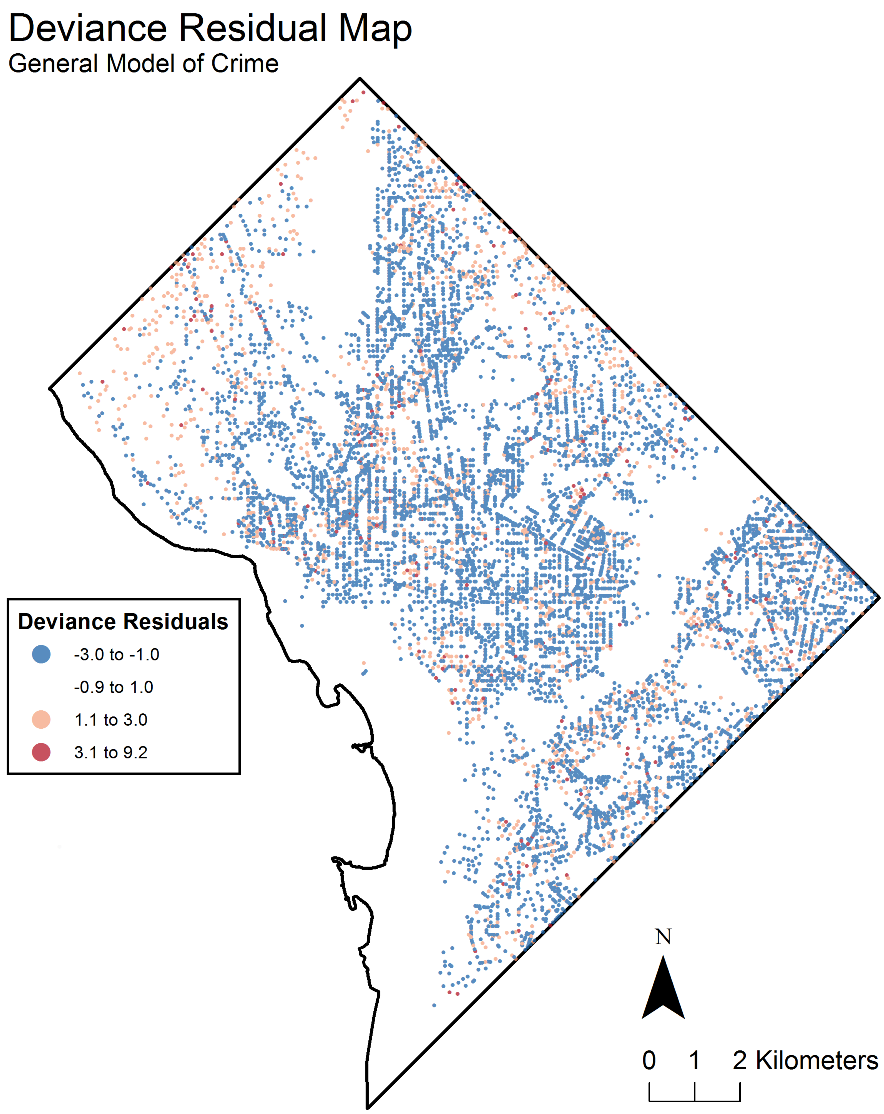


Figure 40: Deviance residuals for the general model of crime. As opposed to the prior models the residuals appear to be much more randomly distributed around the city.

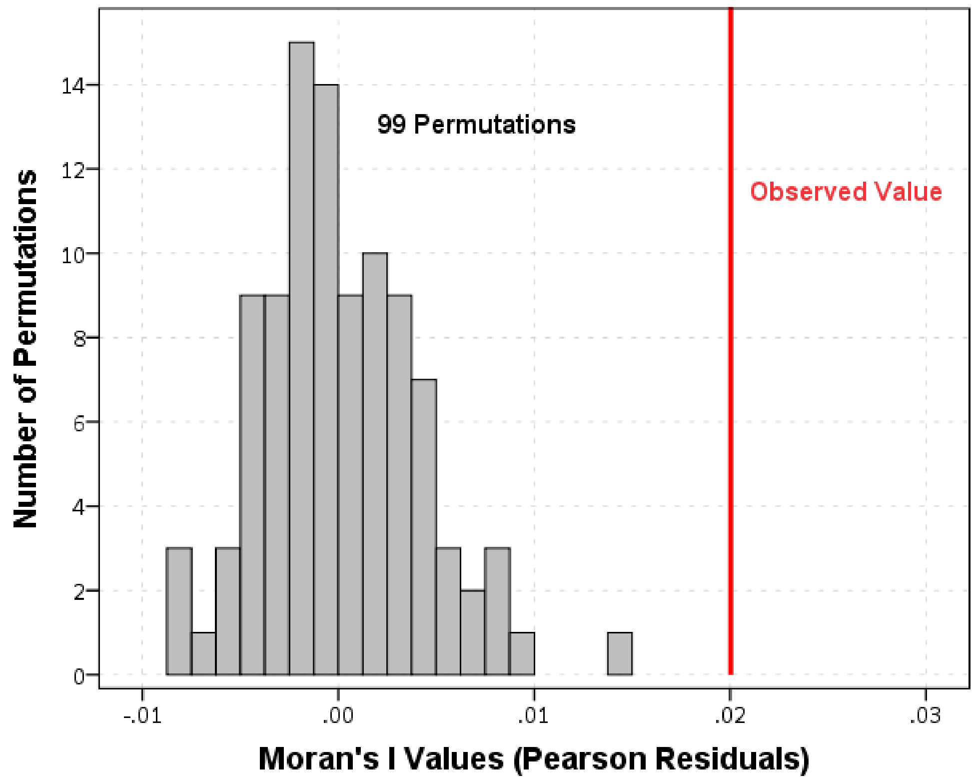


Figure 41: Moran's I values for Pearson residuals and for 99 permutations. The residual spatial autocorrelation is quite small, 0.02, but still significantly high compared to the null distribution.

still exists. Using the Pearson residuals and 99 permutations to generate a reference null distribution, a Moran's I value of 0.02 is observed, and the permutation distribution tends to be between -0.01 and 0.01, so the observed value is statistically significant at a 0.01 level. This test is displayed in Figure 41. Examination of the deviance residuals came to the same conclusion, and the observed value of Moran's I was slightly higher at 0.025.

Examining the correlogram at different distances, Figure 42 shows a similar story to that of the previous chapter. A negative auto-correlation is observed at very small distances of 100 meters or less, and then the auto-correlation swings back to positive and then generally declines for larger distances. Still the observed values are always outside of the 19 permutation grey lines until distance bins farther than $1,000$ meters.

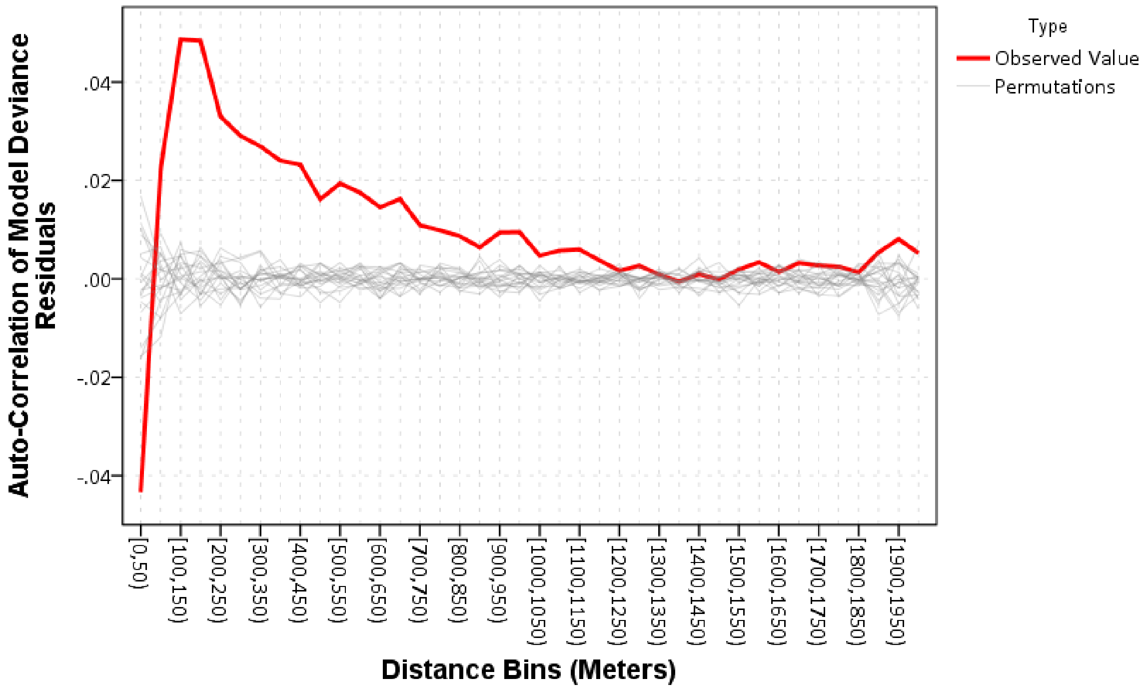


Figure 42: The spatial correlogram exhibits negative spatial auto-correlation at very short distances (under 100 meters) and then positive spatial auto-correlations at larger distance lags. The auto-correlations are smaller than the marginal crime totals, but are still unlikely to be generated by random spatial data having the same marginal distribution.

This suggests that the neighborhood differencing may not have been the culprit in the negative auto-correlation observed in the prior chapter. It is possible that the construction of the street units and the competitive process of assigning address coordinates is still not properly modelled. One may need to include spatial lags of the size of the neighbors as well as the size of the Thiessen polygon itself in the model. This correlogram also indicates that the spatial weights matrix only including contiguity neighbors is likely insufficient, as auto-correlation still exists at much higher distances of up to $1,000$ meters. The slow decay of the correlogram appears to be consistent with a moving average type spatial error.

10.3.2 Outliers

Comparing the outliers from the previous chapter in bars on crime, the model predicts the locations much more closely, but still tends to over predict the amount of crime at these particular locations. Again an online map is provided at `https://dl.dropbox.com/s/jxg2mjss6pvjam3/OutlierMap.html` where one can interactively see these 11 locations and their new predictions.

For the updated general model of crime seven new locations (and one of the prior locations which was greatly over-predicted) are included in a map of potential outlying locations. These locations were identified by examining model residuals, leverage and Cook's distance values. The current model did a much better job of predicting the higher crime locations over the previous iterations, and so including the wide set of generator locations likely increased the ability of the model to replicate reality. The spatial distribution of these outliers are dispersed throughout the city, unlike the prior model in which most were clustered in central D.C. Most of the outlier locations were now severely under predicted in their crime count. The findings are much more consistent with random outliers you might expect to find in 21,000 cases though, especially in terms of their spatial distribution.

10.4 Conclusion

This chapter presents an updated general model of crime at micro places. Many of the problematic aspects of the prior models, including poor predictions and spatial auto-correlation, were somewhat mitigated. The model still appears to have a slight amount of residual auto-correlation remaining, and the spatial correlogram shows that small distances still have a significant amount of negative auto-correlation and a positive auto-correlation at larger distances. This suggests that the overall auto-correlation may be the result of a mixture of negative and positive auto-correlation, and examination of the individual contributions to the global test statistic would be appropriate (Dray, 2011).

These model mis-specifications, as well as the potential for omitted variables, make interpreting the estimated effects in the model difficult. The different types of liquor licenses continue to be one of the larger effects on crime in the model, both locally and spatially. It is likely *some* of this effect is confounded with unobserved processes, such as liquor licenses self-selecting into locations that already have a larger amount of crime, but how much is unclear. A more appropriate research design might use other instruments to estimate the effect of bars on crime.

The effects of detritus and infrastructure 311 calls for service on crime were estimated as being statistically significant, but are incredibly small. These effects were similar in both models controlling for neighborhood effects as well as in the model presented in this chapter. The reason the effects are so precise is that calls for service are common throughout the city, so there is a great amount of variation. But, these effects appear inconsequentially small compared to other place based indicators of more people being in the area, such as a street unit being adjacent to a hospital or a public housing complex.

Other aspects of the built environment on crime have mixed results contrary to prior expectations. The null effect of parks is in particular surprising, and I suspect is partially a result of the conception of units of analysis. It is difficult to geocode locations within a park, and they are often assigned some arbitrary single location. With the discretization of street units, a park will be split up into multiple units, and so it may be that the one preferred geocoded location is high crime, but the rest of the locations are low crime.

Other measures of aspects of the built environment that should theoretically *decrease* crime, such as street lighting, trees, green sites, sidewalk cafes, and wireless hot spots, have either no discernible effect or slightly positive effects on crime. To appropriately identify the effect of these aspects of the built environment on crime will take more thought than a regression model that includes everything and the kitchen sink. Because these aspects of the built environment are very slowly changing, natural experiments are likely difficult to come by.

A unique contribution of the model presented in this chapter (in addition to the wide variety of aspects of the built environment) is the inclusion of non-linear spatial spline terms. While they do not greatly impact the estimated effects for most of the variables in the model, they do reduce the amount of spatial auto-correlation greatly as well as faithfully recreate the general trends in crime across the city. This provides a representation of crime that is a smooth function of space, as opposed to a neighborhood model of crime. A neighborhood model of crime implies that crime is monolithic within neighborhoods and discontinuous between neighborhoods, which is clearly inappropriate when examining the data.

Chapter 11

Conclusion and Future Goals

A recap of the empirical findings from the dissertation is quite simple:

- The number of liquor licenses at street units are a consistent predictor of Part 1 crimes. Adding one institution with a liquor license at a street unit will typically increase the yearly number of Part 1 crimes by around 2.
- Detritus and infrastructure 311 calls for service are associated with increased numbers of crimes on a street unit. The relationship is so small though to as be practically meaningless.

A hope of this dissertation though is to formalize *why* we want to use small units of analysis and *when* we **need** to use small units of analysis given the objectives of the study design. I was critical of prior work both suggesting one needed to use small units of analysis or that larger units of analysis were needed. The data analyses were vehicles to illustrate these points in applied examples.

Ultimately researchers in criminology will be interested in explaining not only why crime occurs, but also where crime occurs. I present a reductionist framework on why it is necessary to examine crime at small units of analysis, as otherwise we can not accurately represent or test our theories of crime at places. Simultaneously it also clarifies the role of aggregate level analysis, and how macro level analysis is not hopeless, just limited in the types of inferences one can make. It also has implications for how one interprets neighborhood level research, and how to *verify* a neighborhood level model one needs smaller units of analysis than the neighborhood.

11.1 Looking towards the Future

One of the grander claims in the dissertation is that crime at micro places do not conform to a neighborhood model of crime. That is, there appear to be no discrete boundaries where neighborhoods exist, and if one simply *assumes* neighborhood boundaries, one will find neighborhood effects even if there is a different data generating process. A spatial effects process as shown here in simulation and in empirical examples when aggregated up will look like a neighborhood effects model. Because of the prevalence of a neighborhood effects formulation, I demonstrate how one can examine between neighborhoods residual auto-correlation plots to determine if the data do not conform to the restrictive neighborhood model.

If one is concerned that a particular variable may be confounded with unobserved neighborhood effects I illustrated how using a fixed effects model can control for those omitted variables given you know the neighborhood boundaries. I did not however incorporate any demographic controls into the final analysis, such as the percentage of female headed households or the number of individuals below poverty at the micro place in the final model. It is demographically neutral. Other work has questioned the reasonableness of typical neighborhood boundaries (Hipp et al., 2011; Hipp and Boessen, 2013), and here I present a different formulation that accounts for general spatial trends by including non-linear spatial terms into the model. I believe the formulation I presented is a more realistic model of crime at micro-places, even absent these well established demographic predictors of crime. I suspect a similar view will not be held among many of my peers, as the relationship between demographic predictors of disadvantage and crime is so well established in macro level (Pratt and Cullen, 2005) and neighborhood level research (Sampson, 2012) it will appear strange not to include these factors in any model of crime.

There remains a large number of annoying methodological questions that need further investigation as to understand their impacts on the study of crime at micro places. While the prior chapters on how aggregation bias occurs shows why one wants to examine small units of analysis, this logic is unlikely to discriminate between different potential small units of analysis for many research questions, such as street segments and intersections used here or a fine grid such as is used in risk-terrain modelling (Kennedy et al., 2010). Using census boundaries with crime reports likely introduces serious error in assigning crime locations, as a large proportion of crime events will be geocoded to intersections that fall on the boundaries of census geographies. This problem is exacerbated if one is using crime already obfuscated to street midpoints, and in this case for the smallest census geography of a block nearly all crime falls on a border. For block groups this indeterminate amount is still incredibly large at around thirty percent in this sample. This makes the place for neighborhood covariates and crime even more uncertain, as allocation of demographic attributes to micro places will inevitably entail measurement error. Using street segments and intersections as a unit of analysis does not come without challenges though either, as these locations are not well defined for particular street layouts that occur in any non-perfect grid.

The number of potential covariates is immense as well. Prior work has examined land use, although different commercial establishments clearly show heterogeneity in the estimated effects presented here. Different types of liquor licenses (liquor stores, grocery and convenience stores, bars and restaurants) have approximately similar estimated effect sizes on crime in these models, but other commercial establishments are not as consistent predictors of crime. A street unit being adjacent to a hospital or a public housing complex predicts a fairly large increase in the rate of crime, and these types of institutions would not be typically captured if one was only modelling the generic land use classification. Simply obtaining all of these different predictors of crime in different study areas is quite a chore, although open data initiatives for many large cities make such data collection at least feasible.

In this analysis I limit the exploration of spatial effects to alcohol licenses and 311 calls for service. These spatial effects could easily be expanded to include all covariates of crime. It is a simple theoretical argument to suggest that these place based crime generators diffuse crime into nearby areas. It is also likely that the way that crime diffuses is not adequately represented in these models, and it may be more appropriate to consider places adjacent based on street connectivity than by adjacency of the Thiessen polygons (Groff, 2013).

Going forward while one may be able to reasonably predict crime at micro places, making inferences about the causes of crime is much more difficult. Aspects of the built environment are slowly changing, and make natural experiments to uniquely identify effects difficult. Here

I present some evidence that bars self-select into criminogenic locations by assuming that bars would have no effect on burglaries if there were no omitted variables in the model. As bars do have a slight effect on burglaries, it is likely the estimated total effect of bars on crime is over-stated. Other quasi-experimental designs will be needed to provide more reassurance that they observed effects are not spurious due to omitted variables.

Bibliography

Ahern, J., C. Margerison-Zilko, A. Hubbard, and S. Galea (2013). Alcohol outlets and binge drinking in urban neighborhoods: The implications of nonlinearity for intervention and policy. *American Journal of Public Health 103*(4), 81–87.

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology 20*, 93–114.

Andresen, M. A. (2011). The ambient population and crime analysis. *The Professional Geographer 63*(2), 193–212.

Andresen, M. A. and N. Malleson (2010). Testing the stability of crime patterns: Implications for theory and policy. *Journal of Research in Crime and Delinquency 48*(1), 58–82.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Anselin, L. (1995). Local indicators of spatial association- LISA. *Geographical Analysis 27*(2), 93–115.

Anselin, L. and D. Arribas-Bel (2013). Spatial fixed effects and spatial dependence in a single cross-section. *Papers in Regional Science 92*(1), 3–17.

Anselin, L. and W. K. Cho (2002). Spatial effects and ecological inference. *Political Analysis 10*(1), 276–297.

Bellair, P. E. and C. R. Browning (2010). Contemporary disorganization research: An assessment and further test of the systemic model of neighborhood crime. *Journal of Research in Crime and Delinquency 47*(4), 496–521.

Berk, R. A. (2005). Knowing when to fold 'em: An essay on evaluating the impact of CEASEFIRE, COMPSTAT, and EXILE*. *Criminology & Public Policy 4*(3), 451–466.

Berk, R. A. and J. MacDonald (2008). Overdispersion and poisson regression. *Journal of Quantitative Criminology 24*(3), 269–284.

Bernasco, W. and R. L. Block (2009). Where offenders choose to attack: A discrete choice model of robberies in Chicago. *Criminology 47*(1), 93–130.

Bernasco, W. and R. L. Block (2010). Robberies in Chicago: A block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. *Journal of Research in Crime and Delinquency 48*(1), 33–57.

Beyerlein, K. and J. R. Hipp (2005). Social capital, too much of a good thing? American religous traditions and community crime. *Social Forces 84*(2), 995–1013.

Blalock, Jr., H. M. (1972). *Causal inferences in nonexperimental research.* New York: W. W. Norton & Company, Inc.

Block, R. L. and C. R. Block (1995). Space, place, and crime: Hot spot areas and hot places of liquor-related crime. *Crime Prevention Studies 4*, 145–184.

Boggs, S. L. (1965). Urban crime patterns. *American Sociological Review 30*(6), 899–908.

Boots, B. N. (1986). *Voronoi (Thiessen) Polygons*. Norwich: Geo Books.

Bowers, K. (2013). Risky facilities: Crime radiators or crime absorbers? A comparison of internal and external levels of theft. *Journal of Quantitative Criminology, Online First*.

Bowers, K. J., S. D. Johnson, R. T. Guerette, L. Summers, and S. Poynton (2011). Spatial displacement and diffusion of benifits among geographically focused policing intiatives: A meta-analytical review. *Journal of Experimental Criminology 7*(4), 347–374.

Bowers, K. J., S. D. Johnson, and A. F. G. Hirschfield (2004). Closing off opportunities for crime: An evaluation of alley-gating. *European Journal of Criminal Policy and Research 10*(4), 285–308.

Braga, A. A. (2001). The effects of hot spots policing on crime. *The Annals of the American Academy of Political and Social Science 578*(1), 104–125.

Braga, A. A., A. V. Papachristos, and D. M. Hureau (2010). The concentration and stability of gun violence at micro places in Boston, 1980-2008. *Journal of Quantitative Criminology 26*(1), 33–53.

Branas, C. C., R. A. Cheney, J. M. MacDonald, V. W. Tam, T. D. Jackson, and T. R. Ten Have (2011). A difference-in-differences analysis of health, safety, and greening vacant urban space. *American Journal of Epidemiology 174*(11), 1296–1306.

Brandt, P. T., J. T. Williams, B. O. Fordham, and B. Pollins (2000). Dynamic modeling for persistent event-count time series. *American Journal of Political Science 44*(4), 823–843.

Brantingham, P. and P. Brantingham (1995). Criminality of place. *European Journal of Criminal Policy and Research 3*(3), 5–26.

Brantingham, P. J., G. E. Tita, M. B. Short, and S. E. Reid (2012). The ecology of gang territorial boundaries. *Criminology 50*(3), 851–885.

Brantingham, P. L. and P. J. Brantingham (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Pyschology 13*(1), 3–28.

Brewer, C. A. and L. Pickle (2002). Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Annals of the Association of American Geographers 92*(4), 662–681.

Briscoe, S. and N. Donnell (2003). Problematic licensed premises for assault in inner Sydney, New Castle, and Wollongong. *The Australian and New Zealand Journal of Criminology 36*(1), 18–33.

Britt, H. R., B. P. Carlin, T. L. Toomey, and A. C. Wagenaar (2005). Neighborhood level spatial analysis of the relationship between alcohol outlet density and criminal violence. *Environmental and Ecological Statistics 12*(4), 411–426.

Brown, B. B. and I. Altman (1991). Territoriality and Residential Crime: A conceptual framework. In P. J. Brantingham and P. L. Brantingham (Eds.), *Environmental Criminology*, pp. 55–76. Thousand Oaks, CA: Waveland Press.

Browning, C. R., R. A. Byron, C. A. Calder, L. J. Krivo, M. Kwan, J. Lee, and R. D. Peterson (2010). Commercial density, residential concentration, and crime: Land use patterns and violence in neighborhood context. *Journal of Research in Crime and Delinquency 47*(3), 329–357.

Buis, M. L. (2010). Direct and indirect effects in a logit model. *The Stata Journal 10*(1), 11–29.

Bushway, S. D. and D. McDowall (2006). Here we go again - Can we learn anything from aggregate-level studies of policy interventions. *Criminology & Public Policy 5*(3), 461–470.

Bushway, S. D., G. Sweeten, and D. B. Wilson (2006). Size matters: standard errors in the application of null hypothesis significance testing in criminology and criminal justice. *Journal of Experimental Criminology 2*(1), 1–22.

Cahill, M. and G. Mulligan (2007). Using geographically weighted regression to explore local crime patterns. *Social Science Computer Review 25*(2), 174–193.

Cerd, M., M. Tracy, S. F. Messner, D. Vlahov, K. Tardiff, and S. Galea (2009). Misdemeanor policing, physical disorder, and gun-related homicide. *Epidemiology 20*(4), 533–541.

Chamlin, M. B. and J. K. Cochran (1995). Assessing Messner and Rosenfeld's institutional anomie theory: A partial test. *Criminology 33*(3), 411–429.

Clare, J., J. Fernendez, and F. Morgan (2009). Formal evaluation of the impact of barriers and connectors on residential burglars' macro-level offending location choices. *The Australian and New Zealand Journal of Criminology 42*(2), 139–158.

Clarke, P., J. Ailshire, R. Melendez, M. Bader, and J. Morenoff (2010). Using Google Earth to conduct a neighborhood audit: Reliability of a virtual audit system. *Health & Place 16*(6), 1224–1229.

Clarke, R. V. (1992). Introduction. In R. V. Clarke (Ed.), *Situational Crime Prevention: Successful Case Studies*, pp. 3–36. Albany, NY: Harrow and Heston Publishers.

Cohen, D., S. Spear, R. Scribner, P. Kissinger, K. Mason, and J. Wildgen (2000). "Broken windows" and the risk of gonorrhea. *American Journal of Public Health 90*(2), 230–236.

Cohen, D. A., S. Inagami, and B. Finch (2008). The built environment and collective efficacy. *Health & Place 14*(2), 198–208.

Cohen, J., W. L. Gorr, and P. Singh (2003). Estimating intervention effects in varying risk settings: Do police raids reduce illegal drug dealing at nuisance bars? *Criminology 41*(2), 257–292.

Cohen, L. E. and M. Felson (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review 44*(4), 588–608.

Cressie, N. (1996). Change of support and the modifiable areal unit problem. *Geographical Systems 3*, 159–180.

Cressie, N. A. (1993). *Statistics for Spatial Data.* New York, NY: J. Wiley.

Davies, G. (2006). *Crime, neighborhood, and public housing.* New York: LFB Scholarly Publishing.

Deane, G., S. F. Messner, T. D. Stucky, K. McGeever, and C. Kubrin (2008). Not 'Islands, Entire of Themselves': Exploring the spatial context of city-level robbery rates. *Journal of Quantitative Criminology 24*(4), 363–380.

Demotto, N. and C. P. Davies (2006). A GIS analysis of the relationship between criminal offenses and parks in Kansas City, Kansas. *Cartography and Geographic Information Science 33*(2), 141–157.

Desmond, S. A., G. Kikuchi, and K. H. Morgan (2010). Congregations and crime: Is the spatial distribution of congregations associated with neighborhood crime rates? *Journal for the Scientific Study of Religion 49*(1), 37–55.

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *American Journal of Public Health 88*(2), 216–222.

Dogan, M. and S. Rokkan (1969). *Quantitative Ecological Analysis in the Social Sciences.* Cambridge, MA: MIT Press.

Donovan, G. H. and J. P. Prestemon (2012). The effect of trees on crime in Portland, Oregon. *Environment and Behavior 44*(1), 3–30.

Dray, S. (2011). A new perspective about Moran's coefficient: Spatial autocorrelation as a linear regression problem. *Geographical Analysis 43*(2), 127–141.

Durrleman, S. and R. Simon (1989). Flexible regression models with cubic splines. *Statistics in Medicine 8*(5), 551–561.

Eck, J. and E. Maguire (2000). Have changes in policing reduced violent crime? An assessment of the evidence. In A. Blumstein and J. Wallman (Eds.), *The Crime Drop in America*, pp. 207–265. New York, NY: Cambridge University Press.

Eck, J. E., R. V. Clarke, and R. T. Guerette (2007). Risky facilities: Crime concentration in homogenous sets of establishments and facilities. *Crime Prevention Studies 21*, 225–264.

Eizenberg, E. (2013). *From the ground up: Community gardens in New York City and the Politics of Spatial Transformation*. Burlington, VT: Ashgate.

Enders, W. (2010). *Applied Econometric Time Series*. Hoboken, NJ: John Wiley & Sons Inc.

Executive Office of the President (2014). Big data: Seizing opportunities, preserving values. Available at `http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf`.

Fagan, J. and G. Davies (2000). Street stops and broken windows: Terry, race, and disorder in New York City. *Fordham Urban Law Journal 28*(2), 457–504.

Fagan, J., G. Davies, and A. Carlis (2012). Race and selective enforcement in public housing. *Journal of Empirical and Legal Studies 9*(4), 697–728.

Farrington, D. P. and B. C. Welsh (2004). Measuring the effects of improved street lighting on crime: A reply to Dr. Marchant. *The British Journal of Criminology 44*(3), 448–467.

Feinberg, R. A. and H. Wainer (2011). Extracting sunbeams from cucumbers. *Journal of Computational and Graphical Statistics 20*(4), 793–810.

Felson, M., R. Berends, B. Richardson, and A. Veno (1997). Reducing pub hopping and related crime. *Crime Prevention Studies 7*, 115–132.

Felson, M. and E. Poulsen (2003). Simple indicators of crime by time of day. *International Journal of Forecasting 19*(4), 595–601.

Firebaugh, G. (1978). A rule for inferring individual relationships from aggregate data. *American Sociological Review 43*(4), 557–572.

Forsyth, A. J. and N. Davidson (2010). Community off-sales provision and the presence of alchohol-related detritus in residential neighborhoods. *Health & Place 16*(2), 349–358.

Foster, S. D., H. Shimadzu, and R. Darnell (2012). Uncertainty in spatially predicted covariates: Is it ignorable? *Journal of the Royal Statistical Society: Series C 61*(4), 637–652.

Franklin, F. A., T. A. Laveist, D. W. Webster, and W. K. Pan (2010). Alcohol outlets and violent crime in Washington D.C. *Western Journal of Emergency Medicine 11*(3), 283–290.

Freedman, D. A. (2006). On the so-called "Huber sandwich estimator" and "robust standard errors". *The American Statistician 60*(4).

Galbraith, R. F. (1988). A note on graphical presentation of estimated odds ratios from several clinical trials. *Statistics in Medicine 7*(8), 889–894.

Gau, J. M. and T. C. Pratt (2008). Broken windows or window dressing? Citizens' (in)ability to tell the difference between disorder and crime. *Criminology & Public Policy 7*(2), 163–194.

Gault, M. and E. Silver (2008). Spuriousness of mediation? Broken windows according to Sampson and Raudenbush (1999). *Journal of Criminal Justice 36*(3), 240–243.

Gehlke, C. E. and K. Biehl (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association 29*(185), 169–170.

Gelman, A. and J. Hill (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, NY: Cambridge University Press.

Gelman, A., D. K. Park, S. Ansolabehere, P. N. Price, and L. C. Minnite (2001). Models, assumptions and model checking in ecological regressions. *Journal of the Royal Statistical Society: Series A 164*(1), 101–118.

Giles, D. (2013). Robust standard errors for nonlinear models. Econometrics Beat: Dave Giles' Blog. Obtained on 6/12/2014. Available at `http://davegiles.blogspot.com/2013/05/robust-standard-errors-for-nonlinear.html`.

Gotway, C. A. and L. J. Young (2002). Combining incompatible spatial data. *Journal of the American Statistical Association 97*(458), 632–648.

Graif, C. and R. J. Sampson (2009). Spatial heterogeity in the effects of immigration and diversity on neighborhood homicide rates. *Homicide Studies 13*(3), 242–260.

Griffiths, E. and J. M. Chavez (2004). Communities, street guns and homicide trajectories in Chicago, 1980-1995: Merging methods for examining homicide trends across space and time. *Criminology 42*(2), 941–978.

Groff, E. R. (2008). Adding the temporal and spatial aspects of routine activities: A further test of routine activity theory. *Security Journal 21*(1-2), 95–116.

Groff, E. R. (2011). Exploring 'near': Characterizing the spatial extent of drinking place influence on crime. *Australian & New Zealand Journal of Criminology 44*(2), 156–179.

Groff, E. R. (2013). Quantifying the exposure of street segments to drinking places nearby. *Journal of Quantitative Criminology, Online First*.

Groff, E. R. and B. Lockwood (2014). Criminogenic facilities and crime across street segments in Philadelphia: Uncovering evidence about the spatial extent of facility influence. *Journal of Research in Crime and Delinquency 51*(3), 277–314.

Groff, E. R. and E. S. McCord (2012). The role of neighborhood parks as crime generators. *Security Journal 25*(1), 1–24.

Groff, E. R., D. Weisburd, and S. Yang (2010). Is it important to examine crime trends at a local "Micro" level?: A longitudinal analysis of street to street variability in crime trajectories. *Journal of Quantitative Criminology 26*(1), 7–32.

Gruenewald, P. J. (2007). The spatial ecology of alcohol problems: Niche theory and assortative drinking. *Addiction 102*(6), 870–878.

Gruenewald, P. J., B. Freisthler, L. G. Remer, E. A. LaScala, and A. J. Treno (2006). Ecological models of alcohol outlets and violent assaults: Crime potentials and geospatial analysis. *Addiction 101*(5), 666–677.

Grunfeld, Y. and Z. Griliches (1960). Is aggregation necessarily bad? *The Review of Economics and Statistics 42*(1), 1–13.

Guerette, R. T. and K. J. Bowers (2009). Assessing the extent of crime displacement and diffusion of benifits: A review of situational crime prevention evaluations. *Criminology 47*(4), 1331–1368.

Haberman, C. P., E. R. Groff, and R. B. Taylor (2013). The variable impacts of public housing community proximity on nearby street robberies. *Journal of Research in Crime and Delinquency 50*(2), 163–188.

Hammond, J. L. (1973). Two sources of error in ecological correlations. *American Sociological Review 38*(6), 764–777.

Hannan, M. T. and L. Burstein (1974). Estimation from grouped observations. *American Sociological Review 39*(3), 374–392.

Harcourt, B. E. and J. Ludwig (2006). Broken windows: New evidence from New York City and a five-city social experiment. *The University of Chicago Law Review 73*(1), 271–320.

Hardin, J. (1996). Fixed-, between-, and random-effects in xtreg. StataCorp LP. Obtained on 5/12/2014. Available at `http://www.stata.com/support/faqs/statistics/xtreg-and-effects/`.

Harrell, Jr., F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis.* New York: Springer-Verlag.

Harrower, M. and C. Brewer (2003). ColorBrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal 40*(1), 27–37.

Hauser, R. M. (1970). Context and consex: A cautionary tale. *American Journal of Sociology 75*(4), 645–664.

Hauser, R. M. (1974). Contextual analysis revisited. *Sociological Methods & Research 2*(3), 365–375.

Haywood, J., P. Kautt, and A. Whitaker (2009). The effects of 'Alley-Gating' in an English town. *European Journal of Criminology 6*(4), 361–381.

Hinkle, J. C., D. Weisburd, C. Famega, and J. Ready (2014). The problem is not just sample size: The consequences of low base rates in policing experiments in smaller cities. *Evaluation Review, Online First.*

Hipp, J. R. (2007). Block, tract, and levels of aggregation: Neighborhood structure and crime and disorder as a case in point. *American Sociological Review 72*(5), 659–680.

Hipp, J. R. and A. Boessen (2013). Egohoods as waves washing across the city: A new measure of "neighborhoods". *Criminology 51*(2), 287–327.

Hipp, J. R., R. W. Faris, and A. Boessen (2011). Measuring 'neighborhood': Constructing network neighborhoods. *Social Networks 34*(1), 128–140.

Hodges, J. S. and B. J. Reich (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician 64*(4), 325–334.

Imai, K., L. Keele, and D. Tingley (2010). A general approach to causal mediation analysis. *Pyschological Methods 15*(4), 309–334.

Jackson, C. K. and E. G. Owens (2011). One for the road: Public transportation, alcohol consumption, and intoxicated driving. *Journal of Public Economics 95*(1-2), 106–121.

Jacobs, B. A. (2010). Serendipity in robbery target selection. *The British Journal of Criminology 50*(3), 514–529.

Jargowsky, P. A. (2005). The Ecological Fallacy. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement*, pp. 715–722. New York, NY: Elsevier.

Jennings, J. M., A. J. Milam, A. Greiner, C. D. M. Furr-Holden, F. C. Curriero, and R. J. Thornton (2014). Neighborhood alcohol outlets and the association with violent crime in one mid-Atlantic city: The implications for zoning policy. *Journal of Urban Health 91*(1), 62–71.

Johnson, S. D. and K. J. Bowers (2010). Permeability and burglary risk: Are cul-de-sacs safer? *Journal of Quantitative Criminology 26*(1), 89–111.

Johnson, S. D., L. Summers, and K. Pease (2009). Offender as forager: A direct test of the boost account of victimization. *Journal of Quantitative Criminology 25*(2), 181–200.

Kang, E. L., D. Liu, and N. Cressie (2009). Statistical analysis of small-area data based on independence, spatial, non-hierarchical, and hierarchical models. *Computational Statistics & Data Analysis 53*(8), 3016–3032.

Kastellec, J. P. and E. L. Leoni (2007). Using graphs instead of tables in Political Science. *Perspectives on Politics 5*(4), 755–771.

Keizer, K., S. Linderberg, and L. Steg (2008). The spreading of disorder. *Science 322*(5908), 1681–1685.

Kelling, G. L. and W. H. Sousa (2001). Do police matter? An analysis of the impact of New York City's police reforms. *Manhattan Institute for Police Research 22*, 1–25.

Kennedy, L. W., J. M. Caplan, and E. Piza (2010). Risk clusters, hotspots, and spatial intelligence: Risk terrain modelling as an algorithm for police resource allocation strategies. *Journal of Quantitative Criminology 27*(3), 339–362.

Kent, J. and M. Leitner (2009). Utilizing land cover characteristics to enhance journey-to-crime estimation models. *Crime Mapping: A Journal of Research and Practice 1*(1), 33–54.

King, G. (1996). Why context should not count. *Political Geography 15*(2), 159–164.

King, G. (1997). *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton, New Jersey: Princeton University Press.

Kinney, J. B., P. L. Brantingham, K. Wuschke, M. G. Kirk, and P. J. Brantingham (2008). Crime attractors, generators and detractors: Land use and urban crime opportunities. *Built Environment 34*(1), 62–74.

Krivo, L. J. and R. D. Peterson (1996). Extremely disadvantaged neighborhoods and urban crime. *Social Forces 75*(2), 619–648.

Kubrin, C. E. and R. Weitzer (2003). New directions in social disorganization theory. *Journal of Research in Crime and Delinquency 40*(4), 374–402.

Kuo, F. E. and W. C. Sullivan (2001). Environment and crime in the inner city: Does vegetation reduce crime? *Environment and Behavior 33*(3), 343–367.

Kurland, J., S. D. Johnson, and N. Tilley (2014). Offenses around stadiums: A natural experiment on crime attraction and generation. *Journal of Research in Crime and Delinquency 51*(1), 5–28.

Kutner, M. H., C. J. Nachtsheim, and J. Neter (2004). *Applied Linear Regression Models*. New York, NY: McGraw-Hill Irwin.

La Vigne, N. G. (1997). Safe transport: Security by design on the Washington Metro. *Crime Prevention Studies 6*, 163–197.

Land, K. C. and G. Deane (1992). On the large-sample estimation of regression models with spatial- or network-effects terms: A two-stage least squares approach. *Sociological Methodology 22*, 221–248.

Land, K. C., P. L. McCall, and L. E. Cohen (1990). Structural covariates of homicide rates: Are there any invariances across time and social space? *The American Journal of Sociology 95*(4), 922–963.

Langbein, L. I. and A. J. Lichtman (1978). *Ecological Inference*. Beverly Hills, CA: Sage Publications, Inc.

Larson, R. C. (1975). What happened to patrol operations in Kansas City? A review of the Kansas City preventive patrol experiment. *Journal of Criminal Justice 3*(4), 267–297.

Lesage, J. and R. K. Pace (2010). The biggest myth in spatial econometrics. *Available at SSRN 1725503*.

Lesage, J. P. and R. K. Pace (2009). *Introduction to spatial econometrics*. Boca Raton, FL: CRC Press.

Levine, N. and K. E. Kim (1998). The location of motor vehicle crashes in Honolulu: A methodology for geocoding intersections. *Computers, Environment and Urban Systems 22*(6), 557–576.

Levine, N., M. Wachs, and E. Shirazi (1986). Crime at bus stops: A study of environmental factors. *Journal of Architectural and Planning Research 3*(4), 339–361.

Levitt, S. D. (2004). Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not. *The Journal of Economic Perspectives 18*(1), 163–190.

Li, W. and J. D. Radke (2012). Geospatial data integration and modeling for the investigation of urban neighborhood crime. *Annals of GIS 18*(3), 185–205.

Lipton, R., X. Yang, A. Braga, J. Goldstick, M. Newton, and M. Rura (2013). The geography of violence, alcohol outlets, and drug arrests in Boston. *American Journal of Public Health 103*(4), 657–664.

Livingston, M. (2008). A longitudinal analysis of alcohol outlet density and assault. *Alcoholism: Clinical and Experimental Research 32*(6), 1074–1079.

Loftin, C. (1986). Assaultive violence as a contagious social process. *Bulletin of the New York Academy of Medicine 62*(5), 550–555.

Loftin, C. and S. K. Ward (1983). A spatial autocorrelation model of the effects of population density on fertility. *American Sociological Review 48*(1), 121–128.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage Publications.

Long, J. S. and J. Freese (2006). *Regression models for categorical dependent variables using Stata*. College Station, TX: Stata Press.

Loukaitou-Sideris, A. (1999). Hot spots of bus stop crime: The importance of environmental attributes. *Journal of the American Planning Association 65*(4), 395–411.

Lugo, W. (2008). Alcohol and crime: Beyond density. *Security Journal 21*(4), 229–245.

Lum, C. (2008). The geography of drug activity and violence: Analyzing spatial relationships of non-homogenous crime event types. *Substance Use & Misuse 43*(2), 179–201.

Lum, C. (2011). Violence, drug markets and racial composition: Challenging stereotypes through spatial analysis. *Urban Studies 48*(13), 2715–2732.

Lum, C., C. S. Koper, and C. W. Telep (2010). The evidence-based policing matrix. *Journal of Experimental Criminology 7*(1), 3–26.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–542.

Massey, D. S. and N. A. Denton (1988). The dimensions of residential segregation. *Social Forces 67*(2), 218–315.

Matthews, R. (1992). Developing more effective strategies for curbing prostitution. In R. V. Clarke (Ed.), *Rational Choice and Situational Crime Prevention*, pp. 89–98. Albany, NY: Harrow and Heston Publishers.

Mayhew, P. (1991). Crime in Public View: Surveillance and Crime Prevention. In P. J. Brantingham and P. L. Brantingham (Eds.), *Environmental Criminology*, pp. 119–134. Thousand Oaks, CA: Waveland Press.

McCord, E. S. and J. H. Ratcliffe (2007). A micro-spatial analysis of the demographic and criminogenic environment of drug markets in Philadelphia. *Australian & New Zealand Journal of Criminology 40*(1), 43–63.

McCord, E. S. and J. H. Ratcliffe (2009). Intensity value analysis and the criminogenic effects of land use features on local crime patters. *Crime Patterns and Analysis 2*(1), 17–30.

McDowall, D., R. McCleary, E. E. Meidinger, and R. A. Hay, Jr. (1980). *Interrupted time series analysis.* Thousand Oaks, CA: Sage Publications.

Mears, D. P. and A. S. Bhati (2006). No community is an island: The effects of resource deprivation on urban violence in spatially and socially proximate communities. *Criminology 44*(3), 509–548.

Messner, S. F., L. Anselin, R. D. Baller, D. F. Hawkins, G. Deane, and S. E. Tolnay (1999). The spatial patterning of county homicide rates: An application of exploratory spatial data analysis. *Journal of Quantitative Criminology 15*(4), 423–450.

Messner, S. F., S. Galea, K. J. Tardiff, M. Tracy, A. Bucciarelli, T. M. Piper, V. Frye, and D. Vlahov (2007). Policing, drugs, and the homicide decline in New York City in the 1990s. *Criminology 45*(2), 385–414.

Mohler, G. O., M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association 106*(493), 100–108.

Morenoff, J. D., R. J. Sampson, and S. W. Raudenbush (2001). Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology 39*(3), 517–558.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46*(1), 69–85.

Murray, R. K. and D. W. Roncek (2008). Measuring diffusion of assaults around bars through radius and adjacency techniques. *Criminal Justice Review 33*(2), 199–220.

Negreiros, J. a. G., M. O. T. Painho, F. J. Aguilar, and M. A. Aguilar (2010). A comprehensive framework for exploratory spatial data analysis: Moran location and variance scatterplots. *International Journal of Digital Earth 3*(2), 157–186.

Newman, G. (1997). Introduction: Towards a Theory of Situational Crime Prevention. In G. Newman, R. V. Clarke, and S. G. Shoham (Eds.), *Rational Choice and Situational Crime Prevention*, pp. 1–24. Brookfield, Vermont: Ashgate Publishing.

Newman, G., R. V. Clarke, and S. G. Shoham (1997). *Rational Choice and Situational Crime Prevention*. Brookfield, Vermont: Ashgate Publishing.

Newman, O. (1972). *Defensible Space: Crime Prevention Through Urban Design*. New York, NY: Macmillan.

Oakes, J. M. (2004). The (mis)estimation of neighborhood effects: Causal inference for a practicable social epidemiology. *Social Sciences & Medicine 58*(10), 1929–1952.

Oberwittler, D. and P. H. Wikstrom (2009). Why small is better: Advancing the study of the role of behavioral contexts in crime causation. In D. Weisburd, W. Bernasco, and G. J. N. Bruinsma (Eds.), *Putting Crime in its Place: Units of Analysis in Geographic Criminology*, pp. 35–59. New York, NY: Springer.

Oppenshaw, S. (1984). *The Modifiable Areal Unit Problem*. Norwich: Geo Books.

Osgood, D. W. (2000). Poisson-based regression analysis of aggregate crime rates. *Journal of Quantitative Criminology 16*(1), 21–43.

Osgood, D. W. and J. M. Chambers (2000). Social disorganization outside the metropolis: An analysis of rural youth violence. *Criminology 38*(1), 81–116.

Ostrom, Jr., C. W. (1990). *Time series analysis: Regression techniques*. Newbury Park, CA: Sage Publications.

Papachristos, A. V., C. M. Smith, M. L. Scherer, and M. A. Fugiero (2011). More coffee, less crime? The relationship between gentrification and neighborhood crime rates in Chicago, 1991 to 2005. *City & Community 10*(3), 215–240.

Paternoster, R., R. Brame, P. Mazerolle, and A. Piquero (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology 36*(4), 859–866.

Patuelli, R., D. A. Griffith, M. Tiefelsdorf, and P. Nijkamp (2011). Spatial filtering and eigenvector stability: Space-time models for German unemployment data. *International Regional Science Review 34*(2), 253–280.

Pearl, J. (2000). *Causality: models, reasoning and inference*, Volume 29. Cambridge Univ Press.

Perkins, D. D., A. Wandersman, R. C. Rich, and R. B. Taylor (1993). The physical environment of street crime: Defensible space, territoriality and incivilities. *Journal of Environmental Pyschology 13*(1), 29–49.

Peterson, R. D. and L. J. Krivo (2009). Segregated spatial locations, race-ethnic composition, and neighborhood violent crime. *The Annals of the American Academy of Political and Social Science 623*(1), 93–107.

Pirkis, J., M. J. Spittal, G. Cox, J. Robinson, Y. T. Cheung, and D. Studdert (2013). The effectiveness of structural interventions at suicide hotspots: A meta-analysis. *International Journal of Epidemiology 42*(2), 541–548.

Platt, J. R. (1964). Strong inference. *Science 146*(3642), 347–353.

Poulsen, E. and L. W. Kennedy (2004). Using dasymetric mapping for spatially aggregated crime data. *Journal of Quantitative Criminology 20*(3), 243–262.

Pratt, T. C. and F. T. Cullen (2005). Assessing macro-level predictors and theories of crime: A meta-analysis. *Crime and Justice 32*, 373–450.

Preacher, K. J. and A. F. Hayes (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers 36*(4), 717–731.

Pridemore, W. A. and T. H. Grubesic (2012). A spatial analysis of the moderating effects of land use on the association between alcohol outlet density and violence in urban areas. *Drug and Alcohol Review 31*(4), 385–393.

Pridemore, W. A. and T. H. Grubesic (2013). Alcohol outlets and community levels of interpersonal violence: Spatial density, outlet type, and seriousness of assault. *Journal of Research in Crime and Delinquency 50*(1), 132–159.

Radil, S. M., C. Flint, and G. E. Tita (2010). Spatializing social networks: Using social network analysis to investigate geographies of gang rivalry, territoriality, and violence in Los Angeles. *Annals of the Association of American Geographers 100*(2), 307–326.

Ratcliffe, J. H. (2012). The spatial extent of criminogenic places: A changepoint regression of violence around bars. *Geographical Analysis 44*(4), 302–320.

Ratcliffe, J. H., T. Taniguchi, and R. B. Taylor (2009). The crime reduction effects of public cctv cameras: A multi-method spatial approach. *Justice Quarterly 26*(4), 746–770.

Ratcliffe, J. H. and T. A. Taniguchi (2008). Is crime high around drug-gang street corners?: Two spatial approaches to the relationship between gang set spaces and local crime levels. *Crime Patterns and Analysis 1*(1), 23–46.

Ratcliffe, J. H., T. A. Taniguchi, and E. R. Groff (2011). The Philadelphia foot patrol experiment: A randomized controlled trial of police patrol effectiveness in violent crime hotspots. *Criminology 49*(3), 795–831.

Raudenbush, S. W. and R. J. Sampson (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology 29*, 1–41.

Rengert, G. F. (1996). *The Geography of Illegal Drugs*. Boulder, CO: Westview Press.

Rengert, G. F. (2004). The journey to crime. In G. Bruinsma, H. Elffers, and J. W. de Keijser (Eds.), *Punishment, places and perpetrators: Developments in criminology and criminal justice research*, pp. 169–181. Portland, OR: Willan Publishing.

Reynald, D. M. and H. Elffers (2009). The future of Newman's Defensible Space Theory. *European Journal of Criminology 6*(1), 25–46.

Ridgeway, G. (2006). Assessing the effect of race bias in post-traffic stop outcomes using propensity scores. *Journal of Quantitative Criminology 22*(1), 1–29.

Ridgeway, G. (2007). *Analysis of racial disparities in the New York Police Department's stop, question, and frisk practices*. Santa Monica, CA: Rand Coorporation.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review 15*(3), 351–357.

Roethlisberger, F. J. (2001). The Hawthorne Experiments. In W. E. Natemeyer and J. T. M. (Eds.), *Classics of Organizational Behavior*, pp. 29–40. Long Grove, IL: Waveland Press.

Roncek, D. W. (1975). Density and crime: A methodological critique. *American Behavioral Scientist 18*(6), 843–860.

Roncek, D. W. (1981). Dangerous places: Crime and residential environment. *Social Forces 60*(1), 74–96.

Roncek, D. W., R. Bell, and J. M. A. Francik (1981). Housing projects and crime: Testing a proximity hypothesis. *Social Problems 29*(2), 151–166.

Roncek, D. W. and A. W. Lobosco (1983). Effect of high schools on crime in their neighborhoods. *Social Science Quarterly 64*(3), 598–613.

Roncek, D. W. and P. A. Maier (1991). Bars, blocks, and crime revisited: Linking the theory of routine activities to the empiricism of "Hot Spots". *Criminology 29*(4), 725–753.

Rosenfeld, R., R. Fornango, and E. Baumer (2005). Did CEASEFIRE, COMPSTAT, and EXILE reduce homicide. *Criminology & Public Policy 4*(3), 419–449.

Rosenfeld, R., R. Fornango, and A. F. Rengifo (2007). The impact of order-maintenance policing on New York City homicide and robbery rates: 1988-2001. *Criminology 45*(2), 355–384.

Rossmo, D. K. (1995). Strategic crime patterning: Problem-oriented policing and displacement. In C. R. Block, M. Dabdoub, and S. Fregly (Eds.), *Crime Analysis Through Computer Mapping*, pp. 1–14. Washington, DC: Police Executive Research Forum.

Roth, R. E., A. W. Woodruff, and Z. F. Johnson (2010). Value-by-alpha maps: An alternative technique to the cartogram. *The Cartographic Journal 47*(2), 130–140.

Sampson, R. J. (2010). Gold standard myths: Observations on the experimental turn in quantitative criminology. *Journal of Quantitative Criminology 26*(4), 489–500.

Sampson, R. J. (2012). *Great American city: Chicago and the enduring neighborhood effect.* Chicago, IL: University of Chicago Press.

Sampson, R. J. and S. W. Raudenbush (1999). Systematic social observation of public spaces: A new look at disorder in Urban neighborhoods. *American Journal of Sociology 105*(3), 603–651.

Sampson, R. J. and S. W. Raudenbush (2004). Seeing disorder: Neighborhood stigma and the social construction of "Broken Windows". *Social Pyschology Quarterly 67*(4), 319–342.

Sampson, R. J., S. W. Raudenbush, and F. Earls (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science 277*(5238), 918–924.

Savitz, D. A. (2012). Commentary: A niche for ecologic studies in environmental epidemiology. *Epidemiology 23*(1), 53–54.

Savitz, N. V. and S. W. Raudenbush (2009). Exploiting spatial dependence to improve measurement of neighborhood social processes. *Sociological Methodology 39*(1), 151–183.

Shadish, W. R., T. D. Cook, and D. T. Campbell (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston, MA: Houghton Mifflin Company.

Shaw, C. R. and H. D. McKay (1969). *Juvenile Delinquency and Urban Areas: A Study of Rates of Delinquency in Relation to Differential Characteristics of Local Communities in American Cities. Revised Edition.* Chicago, IL: University of Chicago Press.

Sherman, L. W. and D. Weisburd (1995). General deterrent effects of police patrol in crime hot spots: A randomized, controlled trial. *Justice Quarterly 12*(4), 625–648.

Short, M. B., P. J. Brantingham, A. L. Bertozzi, and G. E. Tita (2010). Dissipation and displacement of hotspots in reaction-diffusion models of crime. *Proceedings of the National Academy of Sciences 107*(9), 3961–3965.

Short, M. B., M. D'Orsogna, P. J. Brantingham, and G. E. Tita (2009). Measuring and modeling repeat and near-repeat burglary effects. *Journal of Quantitative Criminology 25*(3), 325–339.

Skogan, W. G. (1990). *Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods.* New York, NY: The Free Press.

Smith, W. R., S. G. Frazee, and E. L. Davison (2000). Furthering the integration of routine activity and social disorganization theories: Small units of analysis and the study of street robbery as a diffusion process. *Criminology 38*(2), 489–524.

Snijders, T. A. B. and R. J. Bosker (2012). *Multilevel analysis: An introduction to basic and advances multilevel modeling.* London: Sage.

Socia, K. M. (2011). The policy implications of residence restrictions on sex offender housing in Upstate NY. *Criminology & Public Policy 10*(2), 351–389.

Stucky, T. D. and J. R. Ottensman (2009). Land use and violent crime. *Criminology 47*(4), 1223–1264.

Taniguchi, T. A., J. H. Ratcliffe, and R. B. Taylor (2011). Gang set space, drug markets, and crime around drug corners in Camden. *Journal of Research in Crime and Delinquency 48*(3), 327–363.

Taniguchi, T. A., G. F. Rengert, and E. S. McCord (2009). Where size matters: Agglomeration economies of illegal drug markets in Philadelphia. *Justice Quarterly 26*(4), 670–694.

Taniguchi, T. A. and C. Salvatore (2012). Exploring the relationship between drug and alcohol treatment facilities and violent and property crime: A socioeconomic contingent relationship. *Security Journal 25*(2), 95–115.

Taylor, B., C. Koper, and D. Woods (2011). A randomized controlled trial of different policing strategies at hot spots of violent crime. *Journal of Experimental Criminology 7*(2), 149–181.

Taylor, R. B. (1997). Social order and disorder of street blocks and neighborhoods: Ecology, microecology, and the sytematic model of social disorganization. *Journal of Research in Crime and Delinquency 34*(1), 113–155.

Taylor, R. B. (2001). *Breaking away from broken windows: Baltimore neighborhoods and the nationwide fight against crime, grime, fear, and decline.* Boulder, CO: Westview Press.

Taylor, R. B. and S. Gottfredson (1986). Environmental design, crime, and prevention: An examination of community dynamics. *Crime & Justice 8*, 387–416.

Taylor, R. B., B. A. Koons, E. M. Kurtz, J. R. Greene, and D. D. Perkins (1995). Street blocks with more nonresidential land use have more physical deterioration. *Urban Affairs Review 31*(1), 120–136.

Tcherni, M. (2011). Structural determinants of homicide: The big three. *Journal of Quantitative Criminology 27*(4), 475–496.

Theil, H. (1965). *Linear aggregations of economic relations*. Amsterdam: North-Holland Publishing Company.

Tilley, N. (1997). Realism, Situational Rationality and Crime Prevention. In G. Newman, R. V. Clarke, and S. G. Shoham (Eds.), *Rational Choice and Situational Crime Prevention*, pp. 95–114. Brookfield, Vermont: Ashgate Publishing.

Tita, G. E., J. Cohen, and J. Engberg (2005). An ecological study of the location of gang "Set Space". *Social Problems 52*(2), 272–299.

Tita, G. E. and G. Ridgeway (2007). The impact of gang formation on local patterns of crime. *Journal of Research in Crime and Delinquency 44*(2), 208–237.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit Region. *Economic Geography 46*, 234–240.

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association 74*(367), 519–530.

Tolnay, S. E., G. Deane, and E. M. Beck (1996). Vicarious violence: Spatial effects on southern lynchings, 1890-1919. *American Journal of Sociology 102*(3), 788–815.

Townsley, M. (2009). Spatial autocorrelation and impacts on criminology. *Geographical Analysis 41*(4), 452–461.

Treno, A. J., P. J. Gruenewald, L. G. Remer, F. Johnson, and E. A. LaScala (2008). Examining multi-level relationships between bars, hostility and aggression: Social selection and social influence. *Addiction 103*(1), 66–77.

Waller, L. A., L. Zhu, C. A. Gotway, D. M. Gorman, and P. J. Gruenewald (2007). Quantifying geographic variations in associations between alcohol distribution and violence: A comparison of geographically weighted regression and spatially varying coefficient models. *Stochastic Environmental Research and Risk Assessment 21*(5), 573–588.

Wang, F. (2005). Job access and homicide patterns in Chicago: An analysis at multiple geographic levels based on scale-space theory. *Journal of Quantitative Criminology 21*(2), 195–217.

Weeks, J. R., A. Getis, A. G. Hill, S. Agyei-Mensah, and D. Rain (2010). Neighborhoods and fertility in Accra, Ghana: An AMOEBA-based approach. *Annals of the Association of American Geographers 100*(3), 558–578.

Weisburd, D., W. Bernasco, and G. J. Bruinsma (2009). *Putting crime in its place: Units of analysis in geographic criminology*. New York, NY: Springer.

Weisburd, D., S. D. Bushway, C. Lum, and S. Yang (2004). Trajectories of crime at places: A longitudinal study of street segments in the city of Seattle. *Criminology 42*(2), 283–322.

Weisburd, D., E. R. Groff, and S. Yang (2012). *The criminology of place: Street segments and our understanding of the crime problem*. New York, NY: Oxford University Press.

Weisburd, D., E. R. Groff, and S. Yang (2014). Understanding and controlling hot spots of crime: The importance of formal and informal social controls. *Prevention Science 15*(1), 31–43.

Weisburd, D., N. A. Morris, and E. R. Groff (2009). Hot spots of juvenile crime: A longitudinal study of arrest incidents at street segments in Seattle, Washington. *Journal of Quantitative Criminology 25*(4), 443–467.

Weisburd, D., N. A. Morris, and J. Ready (2008). Risk-focused policing at places: An experimental evaluation. *Justice Quarterly 25*(1), 163–200.

Weisburd, D. and A. R. Piquero (2008). How well do criminologists explain crime? Statistical modeling in published studies. *Crime and Justice 37*(1), 453–502.

Weisburd, D., L. A. Wyckoff, J. Ready, J. E. Eck, J. C. Hinkle, and F. Gajewski (2006). Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benifits. *Criminology 44*(3), 549–592.

Wheeler, A. P. (2012). The moving home effect: A quasi experiment assessing effect of home location on the offence location. *Journal of Quantitative Criminology 28*(4), 587–606.

Wheeler, D. C. and L. A. Waller (2009). Comparing spatially varying coefficient models: A case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests. *Journal of Geographical Systems 11*(1), 1–22.

White, G. F., R. R. Gainey, and R. A. Triplett (2012). Alcohol outlets and neighborhood crime: A longitudinal analysis. *Crime & Delinquency, Online First*.

Wilcox, P. and J. E. Eck (2011). Criminology of the unpopular. *Criminology & Public Policy 10*(2), 473–482.

Wilcox, P., N. Quisenberry, D. T. Cabrera, and S. Jones (2004). Busy places and broken windows? Toward defining the role of physical structure and process in community crime models. *Sociological Quarterly 45*(2), 185–207.

Wilson, J. Q. and G. L. Kelling (1982). The police and neighborhood safety: Broken windows. *Atlantic Monthly 127*, 29–38.

Wood, D. (1991). In Defense of Indefensible Space. In P. J. Brantingham and P. L. Brantingham (Eds.), *Environmental Criminology*, pp. 77–95. Thousand Oaks, CA: Waveland Press.

Wortley, R. and M. McFarlane (2011). The role of territoriality in crime prevention: A field experiment. *Security Journal 24*(2), 149–156.

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics 5*(3), 161–215.

Wyant, B. R., R. B. Taylor, J. H. Ratcliffe, and J. Wood (2011). Deterrence, firearm arrests, and subsequent shootings: A micro-level spatio-temporal analysis. *Justice Quarterly 29*(4), 524–545.

Yang, S. (2010). Assessing the spatial-temporal relationship between disorder and violence. *Journal of Quantitative Criminology 26*(1), 139–163.

Young, L. J., C. A. Gotway, J. Yang, G. Kearney, and C. DuClos (2009). Linking health and environmental data in geographical analysis: It's so much more than centroids. *Spatial and Spatio-temporal Epidemiology 1*(1), 73–84.

Zhu, L., D. Gorman, and S. Horel (2006). Hierarchical Bayesian spatial models for alchohol availability, drug "hot spots" and violent crime. *International Journal of Health Geographics 5*(1).

Zhu, L., D. M. Gorman, and S. A. Horel (2004). Alcohol outlet density and violence: A geospatial analysis. *Alcohol and Alcoholism 39*(4), 369–375.

Zimbardo, P. G. (1973). A field experiment in auto shaping. In C. Ward (Ed.), *Vandalism*, pp. 85–90. London: The Architectural Press.