Social Science Computer Review

Using Geographically Weighted Regression to Explore Local Crime Patterns

Meagan Cahill and Gordon Mulligan

The online version of this article can be found at:

Published by:

$SAGE

Additional services and information for *Social Science Computer Review* can be found at:

Email Alerts: http://ssc.sagepub.com/cgi/alerts

Subscriptions: http://ssc.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://ssc.sagepub.com/content/25/2/174.refs.html

Using Geographically Weighted Regression to Explore Local Crime Patterns

Meagan Cahill
The Urban Institute

Gordon Mulligan
University of Arizona

The present research examines a structural model of violent crime in Portland, Oregon, exploring spatial patterns of both crime and its covariates. Using standard structural measures drawn from an opportunity framework, the study provides results from a global ordinary least squares model, assumed to fit for all locations within the study area. Geographically weighted regression (GWR) is then introduced as an alternative to such traditional approaches to modeling crime. The GWR procedure estimates a local model, producing a set of mappable parameter estimates and *t*-values of significance that vary over space. Several structural measures are found to have relationships with crime that vary significantly with location. Results indicate that a mixed model— with both spatially varying and fixed parameters—may provide the most accurate model of crime. The present study demonstrates the utility of GWR for exploring local processes that drive crime levels and examining misspecification of a global model of urban violence.

Keywords: geographically weighted regression; crime; Portland, Oregon

Ecological studies of crime have long demonstrated the tendency of criminal events to cluster in space. The search for ecological covariates of crime has been aided in recent decades by the development of multivariate statistical techniques and guided by ecological theories, especially social disorganization and routine activities theories. Led by Land, McCall, and Cohen's (1990) study of homicide covariates, many ecological studies are driven by the search for structural covariates of crime that are "invariant" over space and time. The approach, however, fails to recognize the possibility of important local differences between predictor variables and crime levels, assuming processes between the two operate identically over space (i.e., assuming the processes are stationary). Criminological researchers have, in the past decade, begun to recognize the importance of considering the nonstationarity of spatial processes and turned more attention to local studies of crime.

Both substantively and empirically, an exploration of the spatial patterns of crime in any study is warranted. A theoretical argument can be made that causal processes driving crime

174

activity may vary over space; that is, predictor variables may operate differently in different locations, even within an urban area (Baller, Anselin, Messner, Deane, & Hawkins, 2001; Fotheringham, Charlton, & Brunsdon, 2001). This may be especially relevant in policy studies, where there is growing recognition that understanding the context of crime—the where and when of criminal events—is key to understanding how crime can be controlled and prevented. Crime studies that highlight local variations—local contexts of crime—will likely have more relevance to real-world policy applications. Empirically, if these variations in causal processes do exist and are not accounted for, the statistical model will be inaccurate (Baller et al., 2001).

Finally, exploring spatial data in ecological studies of crime can be useful even if the existence of local processes is not theoretically supported. Recognizing that localized trends in spatial data can affect the accuracy of a global model by reducing its explanatory power in some areas provides an impetus for exploration of the spatial patterns. Assuming a global model does exist, an exploration of spatial patterns in the data can help determine whether a global model is misspecified—whether the model is missing important predictor variables or if a spatial term should be included in the model—which would improve the accuracy of the global model in explaining crime levels across the study area. The present study examines covariates of violent crime in Portland, Oregon. In doing so, this study is not unique; examples of this type of research abound in the literature (e.g., Ackerman, 1998; Cahill & Mulligan, 2003; Harries, 1995; Messner & Tardiff, 1985; Sampson & Groves, 1989). Instead, the present study emphasizes the possible spatial variation in crime measures and their covariates by presenting a local analysis of crime using geographically weighted regression (GWR) and comparing the results to a global ordinary least squares (OLS) model. The GWR method estimates parameters for all sample points in the data set, taking into account the nonstationarity of relationships. The results demonstrate the utility of such an analysis for exploring local processes that drive crime levels and examining misspecification of a global model of urban violence. Before discussing both the global OLS model and the GWR results, however, a brief review of the theoretical perspectives employed is presented.

Crime and Communities Perspective

Ecological research is founded on the idea that understanding the characteristics of places—including physical and social measures—that affect the number of targets and offenders in an area is necessary to an understanding of the causes of crime. Theoretically, studies of this nature have been informed by two somewhat different perspectives: (a) social control-disorganization theory and (b) routine activities theory. Although the two schools of thought are closely related, an important distinction can be made. Social control-disorganization theory focuses on the ability (or lack thereof) of residents of some geographic unit (e.g., a neighborhood) to come together to achieve a common goal, such as reducing predatory crime (Sampson, 1997, 1999). Alternatively, routine activities theory focuses on the presence of opportunities for crime in an area, as shaped by residents' daily activities and determined by the spatial and temporal intersection of three key elements: suitable targets, motivated offenders, and the lack of any capable guardians (Cohen & Felson, 1979; Miethe & Meier, 1994). Wilcox, Land, and Hunt (2003) integrate the two theories into a

"multicontextual opportunity theory" that recognizes the context of crime as essential to an understanding of crime patterns.

The opportunity framework considers both individual and structural influences on criminal opportunity. Social structures affect criminal opportunity by influencing the routine activities of residents and visitors and by affecting the sociodemographic makeup of places, such as income and education levels, family stability, employment patterns, and age and demographic structures. The structural characteristics and routine activity patterns of individuals and areas in turn influence the crime profile of a place, which can vary over time with changes in structure and activity.

Taking a classical stance on motivation, Wilcox et al. (2003) assume all individuals in a bounded locale (place) to be motivated offenders. Thus, at the individual level, exposure and proximity to other individuals in a particular place is assumed to increase risk of individual victimization. At the environmental level, exposure to the motivated offender population is a function of population density (i.e., the higher the population density, the higher one's exposure to motivated offenders). The authors also distinguish between those motivated offenders who reside in an area ("resident motivated offender") and those who come to the area for other reasons (e.g., work, school, shopping, recreation; "ephemeral motivated offenders"). These populations can be thought of as functions of land use in the area. Furthermore, under this conceptualization, targets can be objects or individuals. Capable guardians are also individuals, but guardianship is affected not only by the simple number of people in a place but also by the ability of the population in that place to effect social control and prevent crime. In particular, disadvantage can decrease the level of social control operating in an area by restricting the ability of residents to mobilize resources. The mobilization of resources plays an important role in the exertion of social control and the ability of residents to organize to address problems, including violence.

In light of the GWR application in the present study, only the environmental contexts of crime in Portland will be considered. Thus, although not providing a complete test of theory, this study will nonetheless contribute to an understanding of ecological aspects of criminal events. The main postulates of the theory are operationalized into several structural measures, discussed below.

Crime Data and Structural Measures

Violent crime data (including homicide, sexual assault, robbery, and aggravated assault) were collected from the Portland Bureau of Police for the years 1998 to 2002. The location and date of each reported crime was collected, and those data were geocoded and aggregated to the census block group level. Frequencies of crime for each category were averaged over the 5 years in the study period to control for anomalous years when there may have been an unexplained spike or fall in crime rates. The log of violent crime rates is used as the dependent variable in both the OLS and GWR models.

The spatial distribution of violence rates is shown in Figure 1.[1] The highest rates are clustered in the downtown area and along the northern edge of the city. In addition, there are pockets of high rates on the eastern side of the city. The map of violence rates in Figure 1 also shows some key elements in the geography of Portland. Namely, there are several main

Figure 1
Average Violence Rates per 100,000 Persons by Block Group,
1998 to 2002, Portland, Oregon
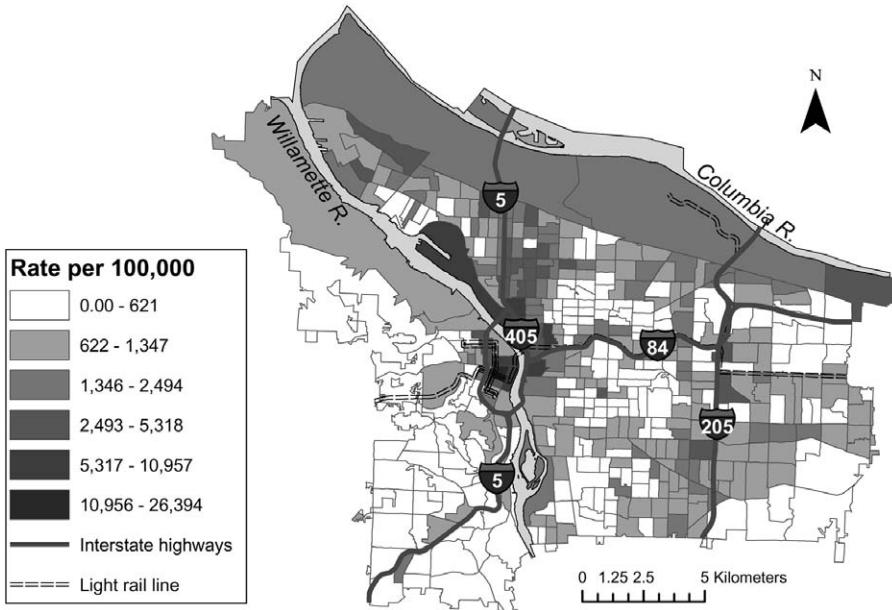
highways, the Willamette River, and a light-rail system that serve to divide the city into five distinct sectors (northwest, west, south, east, and north). The city's downtown is located on the western banks of the Willamette River and encircled by Interstate Highways 5 and 405. A light-rail system generally follows Interstate Highway 84, cutting west-east across the center of the city. Higher levels of violence appear to flank the interstate highways throughout the city, even through areas that are highly affluent, such as the southwestern tip of the city, where higher violence levels are clustered along Interstate 5.

Based on the 2000 census and the theoretical framework, several structural measures were examined for use in the regression modeling at the block group level. A measure of disadvantage used here was proposed by Massey (2001) in his discussion of the growing economic segregation of both poor and affluent families. Sociologists have recently turned more attention to concentrated affluence, investigating the idea that affluence is more significant than simply being "not disadvantaged." Instead, much as Wilson (1987) argued that living in a disadvantaged neighborhood compounded the effects of disadvantage, living in affluent neighborhoods can compound the effects of affluence (Brooks-Gunn, Duncan, Kato, & Sealand, 1993; Massey, 1996). Affluent neighborhoods may produce protective characteristics based on access to and mobilization of various resources (Morenoff, Sampson, & Raudenbush, 2001). Massey's (2001) essay on the neighborhood effects literature suggested that concentrated disadvantage and affluence represent two ends of a continuum and thus are highly

(negatively) correlated and should not be included in statistical models as separate measures. To accommodate this, he proposed the Index of Concentration at the Extremes (ICE), calculated as the number of affluent families minus the number of poor families, divided by the total number of families. For this research, affluence is defined as families with incomes greater than $50,000, the median income level in Portland, and poor is defined as families with incomes less than $15,000, an approximate poverty threshold. The use of $50,000 as the lower bound of the affluence measure follows the previous research of Morenoff and colleagues (2001) and is similar to the threshold used by Kubrin and Stewart (2006) in their work on recidivism in Multnomah County, Oregon. The index has a theoretical range of –1 to 1, where –1 identifies areas where all families are poor and 1 identifies areas where all families are affluent. A 0 value identifies areas with an equal share of poor and affluent families.

Other measures of the level of social control, or guardianship, in an area are residential stability, racial heterogeneity, proportion of single-person households, and level of family disruption (Wilcox et al., 2003). Stability is measured as the percentage of residents who lived in the same residence in 1995. The heterogeneity index used here "takes into account both the relative size and number of groups in the population" (Sampson & Groves, 1989, p. 784) and is equal to $1 - \Sigma p_i^2$, where p_i is the proportion of each racial group in the population. The index has a theoretical range of 0 to 1, where 1 indicates maximum heterogeneity. In practice, however, the upper limit of the index is equal to $1 - 1/n$, where n is the number of racial groups. For this research, seven racial groups defined by the U.S. census are included, and the practical upper limit of the index is 0.86. Family disruption (or lack thereof) is measured by the proportion of married families in the area.

Other structural measures suggested by the opportunity framework indirectly measure the availability of targets and offenders. These are population density (persons per square kilometer), a land-use measure that indicates the percentage of land put toward commercial or multiple uses (i.e., residential and commercial), and a dummy variable that indicates whether there is a light-rail stop in the area. These measures are postulated to influence the number of motivated offenders in an area: Population density increases the number of potential offenders, commercial or multiple use areas increase the number of visitors to an area, and the existence of a light-rail stop in the area can mean that the area is more easily accessible and has more visitors. It should be noted, however, that population density can be differently interpreted within the same theoretical framework; it can be understood to increase the number of guardians in an area and thus have a negative relationship with crime rates. The evidence in the literature regarding population density has been mixed, but the variable included here as GWR will be applied in an exploratory manner, allowing a more thorough examination of the relationship between density and violent crime.

Descriptive statistics for the violence and structural measures are shown in Table 1.

A Global Model of Violence in Portland

A multivariate model was developed to estimate average levels of violence in Portland during the 1998 to 2002 period. The model was developed at the block group level using OLS regression. The model is considered to be global as one parameter is estimated for each variable included in the model. The relationships between predictor variables and the

Table 1
Descriptive Statistics for Violence and Structural Measures

	Minimum	Maximum	M	SD
Violence rate	0.00	26,394.16	1,093.78	1,708.97
Heterogeneity index	0.06	0.74	0.34	0.16
ICE	–0.54	1.00	0.38	0.25
Married families	26.53	100.00	74.10	14.96
Multiple land use	0.00	86.00	12.60	15.57
Population density	0.03	11.13	2.71	1.37
Residential stability	5.84	79.34	46.13	13.15
Single-person households	8.33	91.66	32.02	14.10

Note: $N = 448$. ICE = index of concentration at the extremes.

violence measure are assumed to be the same at all locations within Portland. The results of the model are shown in Table 2. The table includes both unstandardized and standardized parameter estimates and collinearity statistics. Table 2 shows that 5 of 8 parameter estimates are significant, and all but one estimate are in the expected direction. Two measures of guardianship—heterogeneity and single-person households—were positively related to violence rates. These measures are hypothesized to foster violent crime by impeding a neighborhood's ability to mobilize resources for addressing crime problems and to develop social control. Two other guardianship measures, residential stability and percentage of married families, were negatively related to crime rates as they indicate areas where residents might be more invested in their neighborhood and more able to mobilize resources. These relationships are all as expected given an opportunity framework, ceteris paribus. The measure of concentrated affluence, ICE, was positively related to crime, indicating that areas with higher affluence have higher levels of violence. This result is highly unusual, and it is not clear why it occurred. The ramifications of this finding, however, are important as they indicate that, in Portland, areas with higher levels of affluence are experiencing high levels of violence. This finding is opposite to what is predicted by opportunity theory. The results of the GWR analysis will be used to explore this finding and explain why this result was obtained, helping to identify whether certain areas with a positive relationship between ICE and violent crime are influencing the global model.

The parameter estimate for population density in this model was significantly negative, in line with the interpretation of population density as a measure of guardianship. The coefficients for the multiple land use and light-rail stop measures were both significantly positive, supporting both as a measures of increased targets and offenders.

The results of the global model are encouraging, and the model explained 36% of the variance in violence rates, a result generally on par with criminal justice studies. However, this relatively low level of explanatory power is an indication that the global model may not be properly specified; GWR allows an opportunity to increase the explanatory power of the model by incorporating important spatial relationships. The variance inflation factors (VIFs) indicate that collinearity among the coefficients is low. Although the results of the model are promising, more than 60% of the variance in the violence measure is unexplained. There are several reasons for this level of unexplained variance. Some obvious determinants

Table 2
Ordinary Least Squares Regression Model, Log of Violent Crime Rate

	Parameter Estimates			
	Unstandardized	Standardized	t	VIF
Intercept	6.81	—	10.21	—
Heterogeneity index	4.03	0.38	7.31	1.86
ICE	0.78	0.12	2.23	1.95
Light-rail stop	0.40	0.06	1.52	1.26
Married families	−0.03	−0.24	−4.84	1.72
Multiple land use	0.02	0.21	4.12	1.83
Population density	−0.10	−0.08	−1.97	1.28
Residential stability	−0.01	−0.08	−1.42	2.17
Single-person households	0.01	0.07	1.18	2.20
Adj. R^2	.361			
SE of the estimate	1.332			

Note: VIF = variance inflation factor; ICE = index of concentration at the extremes.

of violence that are missing from this model could improve the results. Specifically, the measures employed here may be too indirect and may not provide the best measures of targets, offenders, and guardians. Individual data collected from residents, whether averaged and included in an aggregate form or incorporated into a multilevel model, would improve measures of routine activities of residents and available targets, for instance, and better capture their effect on the three main elements of criminal opportunity. Furthermore, it is likely that local variations in the relationships between the predictor variables and the violence measures do exist, and failing to include this variation can reduce the explanatory power of the model. The effects of either type of misspecification (failing to include appropriate predictors in the model or failing to model spatial patterns) are compounded by the fact that the global OLS model is masking any variations in relationships between the independent and dependent variables. In any case, GWR can be utilized in an exploratory manner to examine the model's performance over space and speculate on possible improvements, whether those are including other measures in a global model or deciding that an explicitly spatial model is more appropriate.

GWR

One of the problems with estimating global regression models for spatial data is that variations over space that might exist in the data are suppressed. In the example given above, the relationship between a violence measure and violence predictors is assumed to be equal at every point in the study area, Portland. To explore whether this is an accurate representation of violence in the study area, a GWR model is useful. The starting point for development of a GWR model is the basic linear regression model,

$$y_i = a_0 + \Sigma_k \beta_k x_{ik} + \varepsilon_i$$

Calibration of the model in Equation 1 results in one parameter estimate for each variable included. A variation of the traditional linear regression model shown in Equation 1 was developed by Brunsdon, Fotheringham, and Charlton (1996), called GWR. Instead of estimating one parameter for each independent variable, GWR estimates local parameters. A parameter is estimated for each data location in the study area. The GWR model is thus expressed as,

$$y_i = a_{0i} + \Sigma_k \beta_{ki} x_{ik} + \varepsilon_i$$

where β_{ki} is the value of β_k at point i (Brunsdon et al., 1996; Fotheringham, Brunsdon, & Charlton, 2002; Fotheringham et al., 2001). For the present study, then, for each variable in the model, a parameter is estimated for each block group in Portland. GWR thus allows a "continuous surface of parameter values" (Fotheringham et al., 2001, p. 52) that can be mapped for visual inspection.

In a GWR model, parameters are estimated using a weighting function based on distance so that locations closest to the estimation point have more influence on the estimate. The GWR parameter estimates are solved using the following equation, given in matrix form,

$$\hat{b}_i = (X_T W_i X)^{-1} X_T W_i y$$

where $\hat{b}_i$ is the estimate of b_i, the location-specific parameter, and W_i is an n by n spatial weighting matrix "whose off-diagonal elements are zero and whose diagonal elements denote the geographical weighting of observed data *for point i*" (Fotheringham et al., 2001, p. 52), as shown below,

$$W_i = \begin{vmatrix} W_{i1} & 0 & 0 & \cdots & 0 \\ 0 & W_{i2} & 0 & \cdots & 0 \\ 0 & 0 & W_{i3} & \cdots & 0 \\ . & . & . & \cdots & 0 \\ 0 & 0 & 0 & \cdots & W_{in} \end{vmatrix}$$

In the spatial weighting matrix, w_{in} is a weight of the data in block group n for estimation of the model around point i; for the present study, the point of estimation is each block group centroid. The weights can be calculated using a variety of methods; for this research, a Gaussian weighting function is employed. According to this weighting scheme, the weight, or influence, of other data on point i decreases with increased distance from point i. The decreasing weight follows a Gaussian curve. The use of a continuous weighting function such as this dictates that locations closest to the point of estimation are more strongly weighted in the calibration of the model. The weighting function employed here takes the form,

$$W_{ij} = \exp\left(\frac{d_{ij}^2}{h^2}\right)$$

where d_{ij} is the Euclidean distance between a block group i, where the model is being calibrated, and another data point j, which in this case is another block group, and h "is a

bandwidth that affects the distance-decay of the weighting function" (Fotheringham et al., 2001, p. 52). Although previous work with GWR has shown that the use of different continuous weighting functions does not have much influence on the model (Fotheringham et al., 2001; Fotheringham et al., 2002), selection of the bandwidth can significantly affect the model calibration. If the bandwidth is too large, the spatial variation will be low and the model at each point will tend toward the global model. If the bandwidth is too small, the number of data points used in estimation may become too low and result in instability in the parameter estimates. Here, the optimal bandwidth is determined through an iterative process to minimize the Akaike information criterion. For a more complete description of bandwidth calibration, see Brunsdon et al. (1996) and Fotheringham et al. (2002).

The local regression model for this study was calibrated using GWR software developed by Fotheringham, Brunsdon, and Charlton (2003). Along with providing the parameter estimates, R^2 values, and t-statistics for each parameter at each data point (block group centroid), the software performs a Monte-Carlo test for assessing the spatial variation in the relationships between the violence measure and explanatory measures. The test is an iterative process that involves randomly rearranging the data to different locations, recalculating parameter estimates and variances, and comparing those variances with the original model's variance (where the data were in the correct location). The result is a p value for each parameter that indicates whether significant spatial variation in the relationship between that parameter and the violence measure exists. There are other tests for spatial variation, but the Monte-Carlo process provides the most robust results. This type of test is important for determining whether a local regression model is indeed appropriate.

A Local Model of Violence in Portland

In the context of the present study, the application of GWR is warranted for several reasons. The OLS model, although promising, left more than 60% of the variance in the violence measure unexplained. Furthermore, one parameter estimate (ICE) had a counterintuitive direction. GWR offers an avenue of spatial data exploration in a regression-modeling framework. GWR also allows a speculation on whether the relationships between violence and the criminal opportunity measures are inherently spatial and can only be accurately modeled if space is explicitly accounted for, indicating directions for future work modeling crime in Portland and, more generally, modeling violent crime in urban areas.

The same independent and dependent variables used in the above OLS model were used to develop a GWR model. The optimal bandwidth was found to be 2.28 kilometers. This bandwidth distance is very similar to the optimal bandwidth found in Malczewski and Poetz's (2005) study of burglary using GWR. Together, these results support the idea that crime is a local occurrence and that local contexts should be considered in studies of crime, in line with routine activity theory's postulate that crime is a function of the daily activities of individuals.

The results indicate that the GWR model represents a statistically significant improvement over the global model at the .01 level. Where the global R^2 was .361, the average local R^2 produced by the GWR model was .86, representing a large improvement in explained variance. The Monte-Carlo tests for significant spatial variation provide evidence for the impor-

Figure 2
**t-Values for Heterogeneity Index Parameter Estimates**
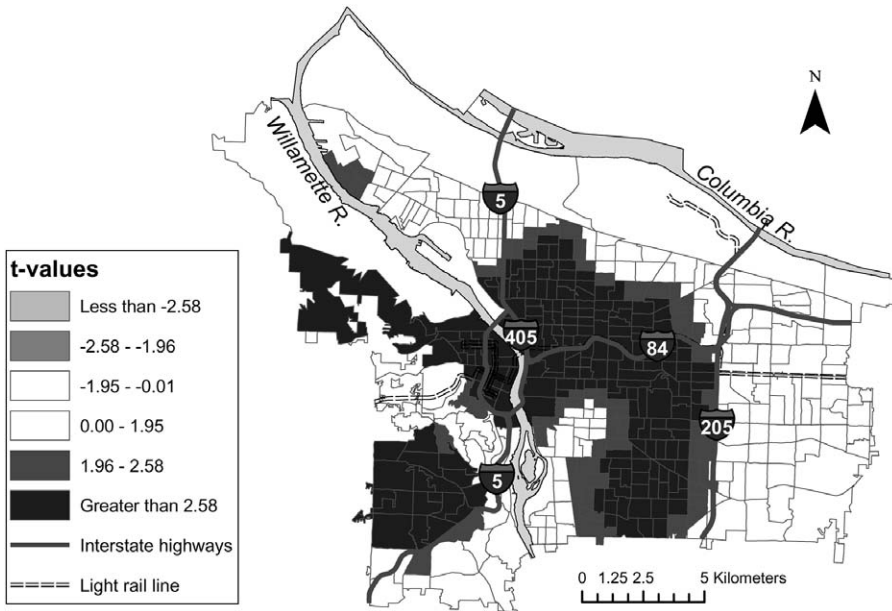
tance of exploring spatiality in statistical models. The tests revealed that the relationships between the violence measure and 4 of the 8 independent measures display significant variation at the 5% level across the city of Portland. Those measures include the heterogeneity index, ICE, population density, and single-person households. Although the resulting set of parameter estimates for each variable can be mapped for visual inspection, perhaps more important are the patterns of t-values for the parameter estimates, revealing which areas have statistically significant estimates. Maps of t-values for the four statistically significant variables are provided in Figures 2 to 5. For all four maps, lighter colors indicate areas of statistically significant negative values, whereas darker colors indicate areas of statistically significant positive values.

First, the map of t-values for the heterogeneity index, shown in Figure 2, loosely follows the pattern of actual values for this parameter. There are few negative parameter estimates, lending support to the OLS model that provided a positive estimate for racial heterogeneity. The lowest 25% of values of the heterogeneity index actually have the highest parameter estimates; 17% of the estimates for the lowest 25% of heterogeneity values are greater than the global estimate. The variable appears to have the greatest positive influence on the model where heterogeneity is lowest.

The measure of disadvantage, ICE, had a positive parameter estimate in the OLS model, a counterintuitive result that was difficult to explain in the opportunity framework. The map of t-values for the measure, shown in Figure 3, reveals that the highly significant positive
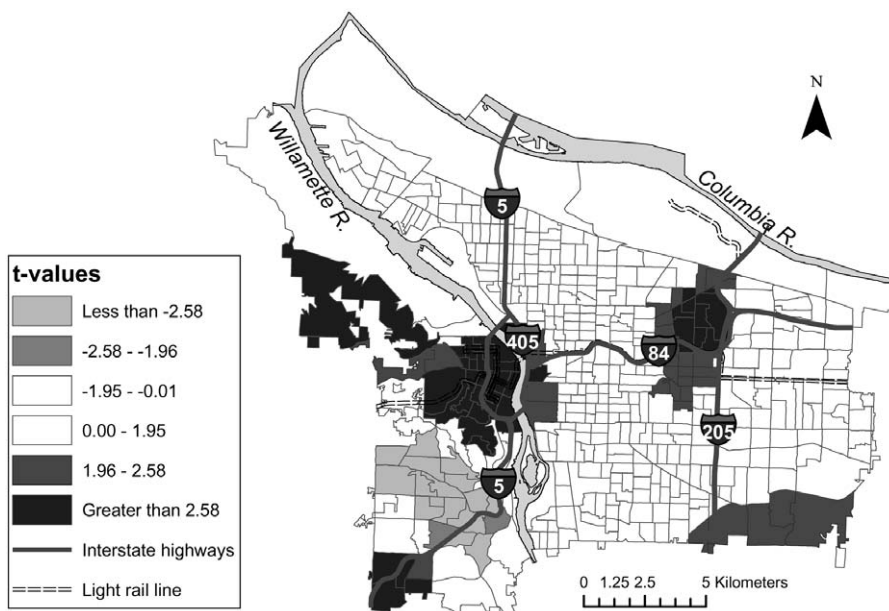
Figure 3
_t_-Values for Index of Concentration at the Extremes Parameter Estimates
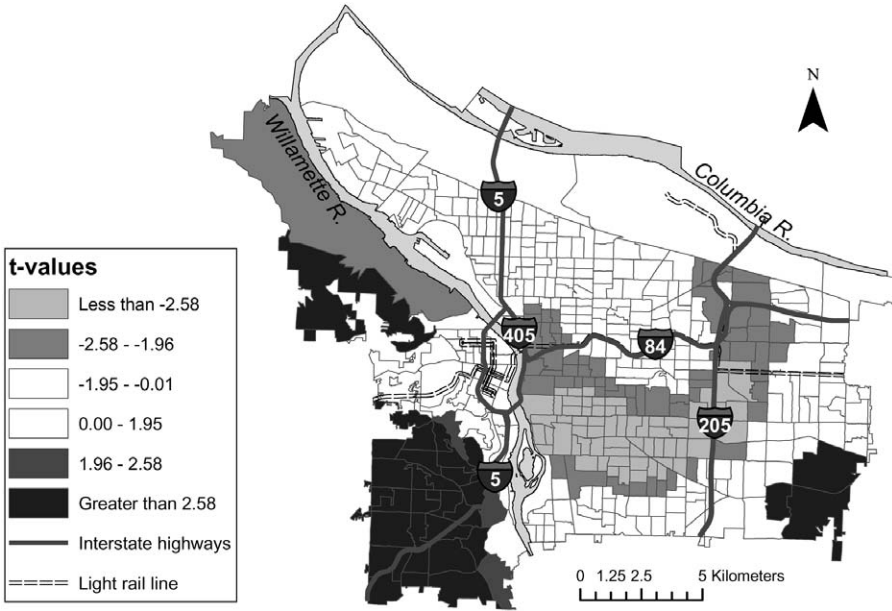
Figure 4
_t_-Values for Population Density Parameter Estimates
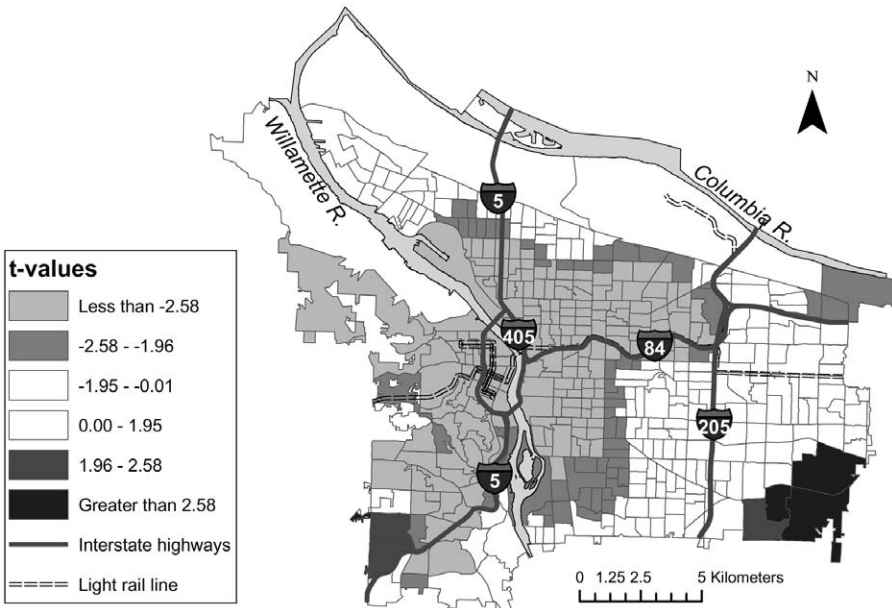
Figure 5
_t_-Values for Single-Member Household Parameter Estimates

values are mainly clustered in the southwest corner of the city, with some highly significant positive values on the eastern edge of the city. These are some of the most affluent areas of the city. Clusters of significant negative values are located in the center of town, and there are actually more block groups with significantly negative values than block groups with positive values; only 10% of the cases have significantly positive parameter estimates, whereas 30% have significantly negative parameter estimates. The 10% with significantly positive estimates have an average violence rate only slightly higher than the overall mean violence rate. The absolute values of the parameter estimates are much larger on the positive side, with the upper 9% of cases having a larger absolute value than the largest negative value. In addition, of those block groups in the upper 25% of ICE values, 20% had values in the upper 25% of parameter estimate values. High ICE values in many cases correspond with large positive coefficients for the ICE measure. The results also suggest that the global model may not capture some other relationship that could help explain the lack of correspondence between parameter estimates and concentrated affluence. A few extreme cases may be affecting the results of the global model, or the ICE measure may behave differently in relation to crime in different areas. It appears, however, that, overall, the theorized negative relationship between affluence and violence holds and that some extremely affluent areas of Portland have violence rates higher than expected but about average compared to the rest of the city. These affluent areas are likely affecting the results of the global model; these results highlight the importance of considering local relationships in crime studies.

Conflicting evidence exists regarding the relationship between population density and

Table 3
Descriptive Statistics for Parameter Estimates in
Geographically Weighted Regression Model

	M	*SD*
(Intercept)	6.41	2.64
Heterogeneity index	2.64	3.03
ICE	0.13	3.51
Light-rail stop	0.26	0.82
Married families	–0.01	0.02
Multiple land use	0.02	0.02
Population density	–0.17	0.31
Residential stability	–0.01	0.04
Single-person households	0.01	0.03

Note: $N = 448$. ICE = index of concentration at the extremes.

crime. In the global model, the measure had significantly negative parameter estimate, but the GWR results indicate that the parameter estimates do vary significantly across space. Figure 4 reveals that most significant estimates are negative, with a large cluster of these cases dominating the center of town. Less than 5% of the block groups have a significantly positive parameter estimate; the population density in these block groups is below the city-wide average and violence rates are in the lowest 15% of all block groups citywide. The results indicate that although spatial variation of the parameter estimates may be significant, the vast majority of cases support the findings in the global model of an inverse relationship between population density and violence rates.

The pattern of significant parameter estimates for single-person households (Figure 5) generally follows the pattern of actual values for that measure. The single-person-household parameter estimates are grouped into distinct clusters of positive and negative values. Estimates are positive in the downtown area, where the actual values for the variable are highest. Estimates are negative in the southwest corner of the city, a residential and affluent area where the average violence rate and average number of single-member households are lower than the citywide averages. This supports the finding in the OLS model that single-person households serve to increase the number of targets and reduce guardianship, therefore contributing to higher levels of crime.

The GWR results for the most part supported the results of the OLS model, even while providing more insight into structural influences on violence in Portland. Table 3 provides the mean parameter estimate values for each measure to allow some comparison between the global and local models; all average estimates in the local GWR model have the same sign as their counterparts in the global model. The GWR results did, however, identify at least one measure, ICE, that should be more closely investigated in an attempt to determine why the patterns of actual values and parameter estimates do not coincide. The *t*-value maps also allowed visual inspection of areas where specific measures have a strong influence in the model (where the estimates are largest, or absolute values are highest). In addition, the *t*-value maps maps allowed local variations not captured by the OLS model to be identified. In several instances, both positive and negative values were estimated for a single

measure. This highlights the importance of considering local context when modeling urban violence. The spatial significance tests provided further support for the application of GWR modeling when studying crime.

GWR Clusters

The exploratory utility of GWR parameters can be extended by clustering together locations with similar parameter values for all variables (i.e., where whole models of violence are similar). This synthesizes the often huge amount of output created by the GWR model and aids interpretation of multiple parameter estimate maps. In the present study, a hierarchical clustering method was applied to the block groups based on the eight parameter estimates and the intercept. Experimentation with a range of clusters (between 4 and 9) revealed that the optimal choice in terms of number of clusters was 6. When producing 5 or fewer clusters, a large cluster with more than 90% of the block groups was extracted. With more than 6 clusters, however, block groups were still not divided into evenly sized groups. With six clusters, the data clustered into four relatively small groups and two large groups that dominate the center and eastern parts of the city. Even when extracting a larger number of clusters from the data, this large group remained, whereas other groups were split into even smaller groups. To characterize each cluster group, descriptive statistics for actual variable values—to describe each area's structural characteristics—and parameter estimates—to characterize the model of violent crime in each area—are provided in Table 4. Although significance tests are not appropriate for parameter estimates reported this way, the average values are useful for describing models in different parts of the city and can be compared to a reference model—the global OLS model. The table also provides information on the different components of violence to examine whether different types of violence are dominant in different areas. Each of the six clusters was named based on the average violence and structural measures to facilitate the discussion of cluster characteristics. Also included in the table is the number of block groups in each cluster (*n*). Together, the structural characteristics and violent crime models allow an investigation of like areas in the city and an exploration of areas where the relationship between violence and structural measures is similar.

The six clusters, shown in Figure 6, are fairly geographically coherent, even though latitude and longitude variables were not included in the clustering calculations. Discriminant analysis confirmed that, overall, approximately 65% of the cluster members were correctly identified based on their location; that is, 65% of the members in all clusters were geographically near other members of the same cluster. By group, 65% of the members in the largest two groups, termed Midtown–High Violence and Downtown and I-205, were accurately assigned based on location, whereas only 40% of the members in the group termed Transitions were accurately assigned based on location.

The largest cluster, with 294 block groups, Midtown–High Violence, dominates the city east of the Willamette River, with a scattering of block groups west of the Willamette. This cluster has a very high overall violence rate. Recalling the geographic pattern of violence rates shown in Figure 1, Midtown–High Violence contains most of the high-violence areas, but the average violence measure is slightly muted because the group is so large and also

Table 4
Descriptive Statistics for Violence and Structural Measures and Parameter Estimates by Cluster Group

Structural Measures	Affluence–Low Violence		Residential–Low Violence		Border Families		Transitions		Midtown–High Violence		Downtown and I-205	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Violence rate	80.00	58.93	195.02	161.31	275.39	387.29	340.18	291.77	1,025.52	1,005.87	1,488.88	2,832.17
Heterogeneity index	0.17	0.06	0.18	0.07	0.25	0.11	0.19	0.09	0.35	0.17	0.34	0.12
ICE	0.50	0.27	0.62	0.16	0.41	0.25	0.64	0.21	0.39	0.25	0.32	0.24
Light-rail stop	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.23	0.15	0.36
Married families	90.95	8.54	74.06	13.13	92.31	11.46	77.50	14.01	72.14	15.14	75.93	13.28
Multiple land use	1.59	1.60	3.97	3.37	9.38	16.05	9.65	12.51	10.70	12.29	18.62	21.01
Population density	1.94	0.67	1.88	0.53	1.65	1.13	1.67	0.39	2.70	1.13	3.01	1.84
Residential stability	53.65	11.31	50.06	18.86	47.83	10.03	52.48	15.34	47.44	12.32	41.76	14.14
Single-person households	30.27	8.76	32.61	5.55	26.83	9.97	31.60	9.47	30.23	11.16	37.00	19.49
Parameters												
Intercept	−2.44	3.32	6.47	1.84	−4.62	5.12	9.40	1.71	7.33	0.88	5.48	0.77
Heterogeneity index	22.61	5.27	16.18	2.98	5.29	4.56	6.41	2.32	1.69	0.81	2.97	0.54
ICE	14.14	2.46	9.85	1.06	14.77	2.50	6.30	1.88	−0.78	0.99	−0.72	1.08
Light-rail stop	0.74	1.98	3.21	0.52	−2.16	1.94	2.35	0.70	0.33	0.42	0.03	0.33
Married families	0.02	0.01	−0.01	0.01	0.02	0.05	−0.02	0.01	−0.01	0.02	0.00	0.01
Multiple land use	0.10	0.02	0.08	0.01	0.07	0.03	0.06	0.01	0.02	0.01	0.02	0.01
Population density	−0.32	0.55	−0.86	0.19	0.44	1.02	−0.61	0.27	−0.20	0.22	−0.09	0.17
Residential stability	−0.11	0.03	−0.12	0.02	−0.09	0.10	−0.11	0.03	−0.01	0.02	0.00	0.01
Single-person households	−0.06	0.05	−0.09	0.01	0.08	0.06	−0.06	0.03	0.00	0.02	0.02	0.02
n	5		5		13		10		294		121	

Note: ICE = index of concentration at the extremes.

Figure 6
Geographically Weighted Regression Clusters
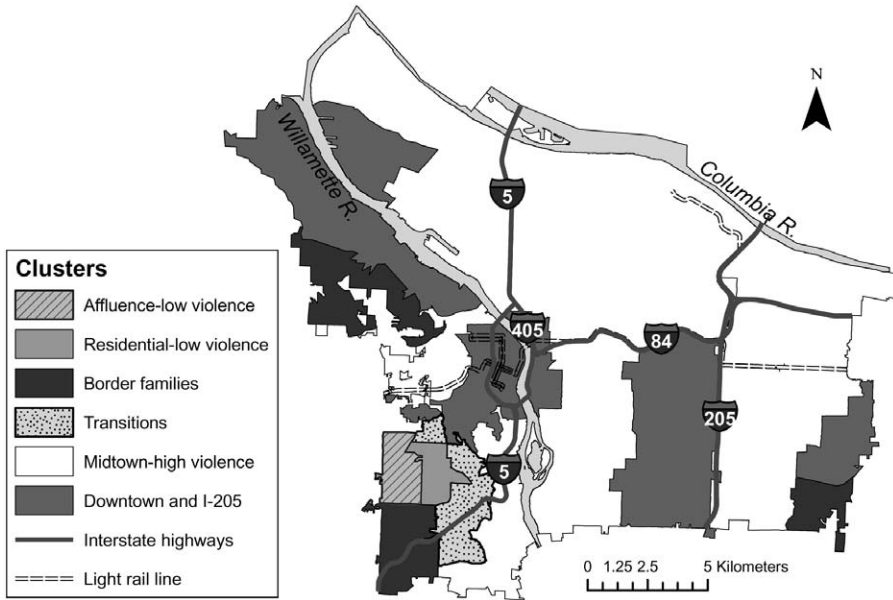
contains block groups with little violence. Because it is so large, the Midtown–High Violence cluster appears to be the average group, with midrange values on the structural measures. The second largest cluster with 121 block groups, Downtown and I-205, has the highest average violence rate. The cluster is not geographically coherent and is made up of small pockets of block groups across the city. The cluster captures all of Portland's downtown, where violence levels are the highest in the city, and also captures a section of block groups in the eastern half of the city along the Interstate 205 corridor. Recalling the pattern of violence rates shown in Figure 1, this Interstate 205 area has higher-than-average violence rates. In addition to having high levels of violence, the Downtown and I-205 cluster contains areas with the highest levels of concentrated poverty, racial heterogeneity, multiple land uses, and population densities, combined with low residential stability. The Downtown and I-205 cluster supports the postulates of opportunity theory in that higher levels of violence are found in areas with lower levels of guardianship, higher levels of targets and offenders, and higher levels of disadvantage.

In terms of the violence model, some of the smallest parameter estimates are found in the Midtown–High Violence and Downtown and I-205 clusters. One interesting finding is that these two clusters have negative values on the ICE measure; this finding reiterates the finding described above that the ICE parameter estimates are actually in the expected negative direction for most of the block groups. A small number of extremely affluent block groups, those grouped into the other four clusters, are influencing the performance of the ICE measure in the global model. The average models for the Midtown–High Violence and

Downtown and I-205 clusters reveal that, in general, the violence model is less sensitive to changes in structural measures where theory indicates that violence should be higher, that is, where poverty, densities, multiple land use, and heterogeneities are average to above average and stability, married families, and affluence are below average.

The rest of the city is divided into other small low-violence groups. Two small clusters in the southwestern tip of the city, Affluence–Low Violence and Residential–Low Violence, have the lowest average violence rates. These are more affluent areas of the city and experience very low violence rates, as shown in Figure 1. The area is also not divided by any major highways, which are characterized by higher levels of violence in Portland. These two highly affluent clusters have some of the largest average parameter estimates, especially for the heterogeneity index and the ICE measure; the parameter estimates on the ICE measure are also positive. This result indicates that in these areas small changes in racial heterogeneity or in levels of affluence have a greater effect on the level of crime than in the two largest, high-violence clusters. Again, the characteristics of these clusters support opportunity theory in that they have low levels of violence, fewer indications of targets and offenders, and more indications of guardianship, including affluence.

The cluster termed Border families is split into three parts in the southwestern, northwestern, and eastern parts of the city, characterized by average violence rates with a high number of married families. The cluster termed Transitions geographically and structurally separates the two high-violence clusters from the three low to average violence clusters. The Transitions cluster experiences average levels of violence, affluence, density, and racial heterogeneity. This pattern of clustering, with several small clusters of low to average violence levels, may be masking some of the variation of violence across the city, but because the clusters were created on the parameter values, the grouping indicates that in the lower violence areas of Portland, violence levels similarly respond to changes in structural characteristics.

Another noteworthy result is that the signs for 5 of the 8 average parameter estimates are both positive and negative across clusters. That is, one cluster may have a negative value for a specific coefficient, whereas another cluster has a positive value for the same coefficient. The heterogeneity index and multiple land use terms are consistently positive across clusters, and the residential stability term is consistently negative across all clusters. In most cases, however, the range of means across clusters is small, and the switch from negative to positive values represents a very small change in absolute value. Also noteworthy is the number of average parameters that are close to zero; the multiple land use, married families, residential stability, and single-person household terms all display both very low average parameter estimates and very low ranges of values. This indicates that although statistically significant, these measures have a small affect on violence levels relative to the other structural measures in the model. The average parameter values allowed a consideration of groups with similar values on all the parameters, highlighting some of the main differences in structural measures and violence across the city of Portland.

Discussion

The application of GWR to a model of violence rates and its comparison to an OLS base model has yielded several striking results. Theoretically, the OLS model, although not as robust

as hoped, did provide support for the criminal opportunity theory. Five of the 8 measures of the elements of opportunity—targets, offenders, and guardians—were significant, and 7 were in the expected direction. The ICE measure was the only variable with a counterintuitive result—a positive coefficient. The GWR results, however, provided insight to the model and revealed that most areas did indeed have a negative parameter estimate; areas with positive estimates tended to have extremely high levels of affluence. A smaller number of areas were affluent with a positive parameter estimate; 20% of block groups in the upper quartile of ICE values also had values in the upper quartile of parameter estimate values. High ICE values in many cases correspond with large positive coefficients for the ICE measure. One approach to further exploring this issue is to recalculate the ICE measure for all block groups using a higher income value to indicate affluence. Previous researchers have used family incomes of or close to $50,000 as the affluence threshold (Kubrin & Stewart, 2006; Morenoff et al., 2001); however, using only the top quartile—$75,000 and higher—might be a more accurate representation of affluence in Portland. Although the ICE result is still not completely explained, the OLS model masked important variation in the parameter, and the GWR results allowed some speculation as to the reason behind the counterintuitive result. The GWR results allow the researcher to focus an investigation on those areas where the model is not performing as expected.

Other GWR results for the most part strengthened the OLS findings. However, the spatial significance tests revealed that 4 of 8 parameters demonstrated significant variation over space; that is, the relationship between those parameters and the violence measure varied across the study area. One way to statistically model this result is to develop a mixed model where some parameters are allowed to vary over space and are estimated using the GWR methods described above, whereas other parameters in the model are fixed. The fixed parameters would have only one estimate, a global estimate that assumed the relationship between that measure and violence to be equal across space. This type of model would also allow dummy location variables, such as a "west of the Willamette River" variable, that may help explain some of the ICE-violence relationship to be included in the model without creating uninterpretable results.

The hierarchical clustering exercise resulted in six geographically coherent groups with similar overall models based on the GWR parameter estimates. The average values for each parameter within each group showed that the largest (positive or negative) parameter estimates were found in the Affluence–Low Violence and Residential–Low Violence clusters, where the average violence rates were the lowest of the six clusters. Smaller average parameter estimates (positive or negative) clustered in the Midtown–High Violence and Downtown and I-205 clusters, which surprisingly had the highest overall violence rates.

Although in support of opportunity theory, the OLS model was not as robust as was hoped, explaining only 36% of the variance in Portland's violent crime rates. The results, especially those for the ICE and heterogeneity variables, indicate that measures not included in the above models could improve the performance of both models. These measures were included as proxies for control or guardianship, and thus reconsidering the measures of guardianship employed in the present study could afford a stronger model. In addition, several of the measures, although significant, had very low parameter estimates, especially single-person households, married families, and population density. These guardianship measures might be too indirect, and the model might be improved with more

direct measures of control. Introducing alternate or additional measures of opportunity may also improve the model. Specifically, opportunity theory draws from routine activity theory; a measure of individual behavior that better captures aggregate routine activities in each area would likely improve the model.

Conclusion

Generally, the results support the application of GWR in this context as the results provided insight into the spatial patterns of parameter relationships. The GWR modeling exercise thus demonstrated the efficacy of this method for descriptive purposes—for exploring spatial relationships between predictor variables and the dependent variable. In addition, the spatial significance Monte-Carlo tests strengthened the argument for at least considering space in studies of violence, if not explicitly including it; here, most of the variables did indeed have locally varying relationships with the violence measure.

GWR can be useful in different types of crime studies. Here, applied in a test of opportunity theory, the exploration of space can help account for differences between communities not captured by standard measures and thus explain causes of crime in different areas. GWR can also be particularly useful in policy studies. Different interventions for crime reduction or prevention may be appropriate in different areas; local attitudes toward types of interventions can vary across an urban area and affect the success of an intervention. Alternatively, GWR can be used to evaluate the success of an intervention already in place by determining areas where the intervention was more successful and why. The method is thus applicable in a range of contexts within the field of criminology.

Note

1. All figures referenced in the text are available in color at http://meagan.cahill.googlepages .com/sscorefigures.

References

Ackerman, W. V. (1998). Socioeconomic correlates of increasing crime rates in smaller communities. *Professional Geographer, 50*(3), 372-387.

Baller, R. D., Anselin, L., Messner, S., Deane, G., & Hawkins, D. F. (2001). Structural covariates of U.S. county homicide rates: Incorporating spatial effects. *Criminology, 39*(3), 561-590.

Brooks-Gunn, J., Duncan, G., Kato, P., & Sealand, N. (1993). Do neighborhoods influence child and adolescent behavior? *American Journal of Sociology, 99*, 353-395.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis, 28*(4), 281-298.

Cahill, M., & Mulligan, G. (2003). The determinants of crime in Tucson, Arizona. *Urban Geography, 24*(7), 582-610.

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

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. West Sussex, UK: John Wiley.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression software* (version 2.0.5). Newcastle upon Tyne, UK: Newcastle University, Department of Geography.

Fotheringham, A. S., Charlton, M., & Brunsdon, C. (2001). Spatial variations in school performance: A local analysis using geographically weighted regression. *Geographical & Environmental Modelling, 5*(1), 43-66.

Harries, K. D. (1995). The ecology of homicide and assault: Baltimore City and County, 1989-91. *Studies on Crime and Crime Prevention, 4*(1), 44-60.

Kubrin, C. E., & Stewart, E. A. (2006). Predicting who reoffends: The neglected role of neighborhood context in recidivism studies. *Criminology, 44*(1), 165-197.

Land, K. C., McCall, P. L., & Cohen, L. (1990). Structural covariates of homicide rates: Are there any invariances across time and space? *American Journal of Sociology, 95*, 922-963.

Malczewski, J., & Poetz, A. (2005). Residential burglaries and neighborhood socioeconomic context in London, Ontario: Global and local regression analysis. *Professional Geographer, 57*(4), 516-529.

Massey, D. S. (1996). The age of extremes: Concentrated affluence and poverty in the twenty-first century. *Demography, 33*, 395-412.

Massey, D. S. (2001). The prodigal paradigm returns: Ecology comes back to sociology. In A. Booth & A. Crouter (Eds.), *Does it take a village? Community effects on children, adolescents, and families* (pp. 41-48). Mahwah, NJ: Lawrence Erlbaum.

Messner, S. F., & Tardiff, K. (1985). The social ecology of urban homicide: An application of the "routine activities" approach. *Criminology, 23*(2), 241-267.

Miethe, T. D., & Meier, R. F. (1994). *Crime and its social context*. Albany: State University of New York Press.

Morenoff, J. D., Sampson, R. J., & Raudenbush, S. W. (2001). Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology, 39*(3), 517-560.

Sampson, R. J. (1997). The embeddedness of child and adolescent development: A community-level perspective on urban violence. In J. McCord (Ed.), *Violence and childhood in the inner city* (pp. 31-77). Cambridge, UK: Cambridge University Press.

Sampson, R. J. (1999). What "community" supplies. In R. F. Ferguson & W. Dickens (Eds.), *Urban problems and community development* (pp. 241-292). Washington, DC: Brookings Institution.

Sampson, R. J., & Groves, W. B. (1989). Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology, 94*(4), 774-802.

Wilcox, P., Land, K. C., & Hunt, S. A. (2003). *Criminal circumstance: A dynamic multicontextual criminal opportunity theory*. New York: Aldine de Gruyter.

Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. Chicago: University of Chicago Press.

Meagan Cahill is a research associate in The Urban Institute's Justice Policy Center. Her current research focuses on the geography and community context of crime and the evaluation of crime reduction and prevention programs. She may be contacted at MCahill@ui.urban.org.

Gordon Mulligan is a professor emeritus in geography and regional development at The University of Arizona. His interests lie in the areas of economic geography and regional science. He may be contacted at mulligan@email.arizona.edu.