METHODOLOGICAL CONSIDERATIONS ON THE SPATIAL AND TEMPORAL ANALYSIS OF VIOLENT CRIME PATTERNS AND THE IDENTIFICATION OF SOCIAL AND ENVIRONMENTAL CORRELATES OF CRIME

METHODOLOGICAL CONSIDERATIONS ON THE SPATIAL AND TEMPORAL ANALYSIS OF VIOLENT CRIME PATTERNS AND THE IDENTIFICATION OF SOCIAL AND ENVIRONMENTAL CORRELATES OF CRIME

By LI HE, B.Sc., M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

DOCTOR of PHILOSOPHY (2016) McMaster University
(School of Geography and Earth Sciences) Hamilton, Ontario, Canada

TITLE	Methodological considerations on the spatial and temporal analysis of violent crime patterns and the identification of social and environmental correlates of crime
AUTHOR	Li He Bachelor of Geographic Information System (Northwest University, 2009) Master of Science in Remote Sensing Technology and Application (Nanjing Normal University, 2012) Master of Science in Earth System Science (The Chinese University of Hong Kong, 2011)
SUPERVISOR	Dr. Antonio Páez, Professor School of Geography and Earth Sciences
NUMBER OF PAGES	xii, 156

ABSTRACT

Violent crime which is referred to as crime against persons has been recognized as the most serious species of crime. Violent crime imposes monetary costs, psychological costs, and social stability disruption. Incarceration strategies used in several societies have been inefficient in deterring, incapacitating, and rehabilitating offenders of violent crime. Additional approaches to preventing and reducing violent crime are needed.

This dissertation aims to address three objectives. First, empirically investigate the temporal stability of socio-economic covariates of violent crime. Second, investigate the persistence of violent crime hot spots over time, and the socio-economic factors that correlate with said persistence. Third, develop a Google Street View-based environmental audit approach to quickly and systematically collect environmental data, and explore the role of physical features of built environment in inducing and deterring violent crime.

The results of model parameter analysis indicate that it is unnecessary to average crime over multiple years if model parameters are stable across years. The results of hot spots persistence analysis suggest that a combination of social disorganization theory and routine activity theory provides an applicable framework for understanding the temporal dimension of violent crime hot spots. By identifying the factors that contribute to the persistence of hot spots of crime, insights gained from the results can help to inform focused crime prevention practices. The results obtained from Poisson regression with spatial filtering and GSV-based environmental audit provide a series of environmental correlates of violent crime, and provide both theoretical and practical implications for several theories of crime and crime prevention efforts.

ACKNOWLEDGEMENTS

First and foremost I would like to express my sincere gratitude to my academic supervisor Professor Dr. Antonio Páez for his continuous guidance, encouragement, and support of my entire Ph.D. study. His firm guidance laid critical foundation for the successful completion of this dissertation. His immense knowledge and expertise significantly helped in developing my future career path. Thanks Dr. Desheng Liu and Dr. Shiguo Jiang for their kindly sharing of the crime dataset, which lays the basis for the works in this dissertation.

I would like to thank my committee members Dr. Niko Yiannakoulias and Dr. Darren Scott for serving as my committee members and guiding me through all four years. I would also like to thank my previous committee member Professor Dr. Pavlos Kanaroglou who has recently passed away. I gained inspirations in many ways from his expertise.

My sincere thank also goes to my colleges and friends for their help and support: Dr. MD Moniruzzaman, Kristina Cimaroli, Mario Reyes, Victor Weigang Tang, Charles Burke, Wei Lu, Celenna Ciuro, Yue Fu, Keqi Wei, Eric Zhao, and Felix Chan.

Last but not the least; I would like to thank my beloved family: my parents, my mother-in-law and father-in-law, my wife, and my sister, for your remote and continuous sacrifices, encouragement, and supports during the entire duration of my Ph.D. study in Canada.

PREFACE

This dissertation consists of five chapters which refer to an introduction, three research papers, and one conclusion. The first research paper that has been published on a peer-reviewed leading journal investigates the temporal stability of socio-economic covariates of violent crime. The second paper that has been published on another peer-reviewed top tier international journal investigates the persistence of violent crime hot spots over time, and the socio-economic factors that correlate with said persistence. The third paper has been submitted to a peer-reviewed leading international journal and is currently under review. The third paper developed a new procedure to collect environmental data and explore the role of physical features of urban residential environment in reducing and inducing violent crime. All three papers focus on the methodological considerations on the spatial and temporal analysis of violent crime patterns and the identification of social and environmental correlates of crime. Violent crime data which has been used in three papers were generated from the same crime database. Therefore, there is certain degree of overlap between three papers in theories of crime and description of crime data and census data used for analysis.

For all three articles, the dissertation author conducted literature review, census data collection, data cleaning, spatial and quantitative analysis of crime data, model estimation, result interpretation, and manuscript writing. Dr. Antonio Páez is co-authored in three papers for his continuous guidance in advancing the research from conceptualization through completion and publication. More specifically, Dr. Páez provides professional contribution in developing research questions and methods, critical appraisal, manuscript editing and revision. Dr. Desheng Liu and Dr. Shiguo Jiang are co-authored as provider of the crime data and professional feedback on manuscripts. The said three papers are as follows:

Chapter 2:

He, L., Páez, A., Liu, D., & Jiang, S. (2015). Temporal stability of model parameters in crime rate analysis: An empirical examination. Applied Geography, 58, 141-152.

Chapter 3:

He, L., Páez, A., & Liu, D. (2016). Persistence of crime hot spots: An ordered probit analysis. Geographical Analysis, DOI: 10.1111/gean.12107.

Chapter 4:

He, L., Páez, A., & Liu, D. (2016). Built environment and violent crime: An environmental audit approach using Google Street View. Computers, Environment and Urban Systems. Under Review.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 Justification of research topic

The FBI's violent crime category consists of four offenses: murder and nonnegligent manslaughter, forcible rape, robbery, and aggravated assault. Compared to crime against property, violent crime that is referred to as crime against persons has been recognized as the most serious species of crime (Elliott and Ageton, 1980). The incidence of violent crime is a major issue of public concern in all societies; violent crime imposes monetary costs (e.g., property loss, medical costs, insurance, law enforcement, the judiciary and corrections) (Andresen, 2006b), and also disrupts social stability – in some cases to the point of becoming a major obstacle to development (Fajnzylber et al., 2002). Violent crime is increasingly recognized as a public health issue (Mair and Mair, 2003; Sparks, 2011; Dustmann and Fasani, 2016). High incidence of violent crime within neighborhoods is a powerful psychosomatic stressor for people living there (Sparks, 2011), and result in many psychological costs (e.g., safety and victimization, mental distress, fear of crime, and anxiety) (Andresen, 2006b; Dustmann and Fasani, 2016). Current incarceration strategies used in several societies have been inefficient in deterring, incapacitating, and rehabilitating offenders of violent crime. Thus, additional approaches to preventing and reducing violent crime are needed (Mair and Mair, 2003).

Researchers have observed that violent crime exhibit noticeable variations across neighborhoods of U.S. cities (Eck and Weisburd, 1995; Sampson et al., 1997). Geographical concentration of crime events forms so called crime “hot spots”. Over the last several decades, a large body of studies has emerged and sought explanations of criminal activity in general characteristics of the social structure and social control on which offenders and victims are embedded. They have consistently shown that socio-economic deprivation, residential mobility, family disruption, and ethnic heterogeneity lead to community social disorganization, which in turn, increases violent crimes (Sampson, 1985; Cahill and Mulligan, 2003; Andresen, 2006a, 2006b; Charron, 2009; Porter and Purser, 2010). More recently attention has been drawn to the role of “places” [1] (Eck and Weisburd, 1995) in inhibiting or encouraging criminal activity. Of particular interest is the role of fine-scale physical features of built environment in understanding disproportionate concentration of violent crimes (Sampson and Raudenbush, 2004; Groff et al., 2009).

The topic of the geography of violent crime is still challenging for several reasons. First, researches have noticed temporal stability (or instability) of crime patterns, however, there’s a lack of investigation in the current literature regarding the implications of the said temporal stability (or instability) for crime modeling practice. Second, the temporal persistence of crime hot spots is recognized as a valuable

[1] In contrast to neighborhoods, criminologists have defined “places” as specific locations within the larger social environment. A place is a small area, and usually a street corner, address, building, or street segment (Eck and Weisburd, 1995).

indicator of consistent problem areas. The current literature has focused on relatively short time periods (e.g., days, weeks, or seasons) which potentially suffer from high variance in risk estimates. The current literature has not adequately addressed the mechanisms that perpetuate or interrupt persistent crime hot spots. Third, recent studies empirically support the role of physical features of built environment in inducing and hindering violent crime (Brantingham and Brantingham, 1993; Anderson et al., 2012). Particularly, studies of the broken window theory have provided evidence that physical disorder is an environmental correlate of crime. However, current studies are limited by, on the one hand, the difficulty involved in collecting fine-scale quantitative environmental data; on the other hand, conventional environmental audit approach that is costly, time-consuming, and burdensome.

1.2 Research questions and objectives

Based on the above considerations, the primary objective of this dissertation is to investigate the social and environmental factors that influence violent crime based on crime hot spot identification and characterization. The following three specific objectives will be addressed in three chapters:

1) To investigate the spatial and temporal stability of violent crime patterns and its implications for modeling practice.

This study contributes to the literature of crime analysis by (1) empirically investigating the temporal stability of model parameters in crime rate analysis, and (2) assessing the appropriateness of using single-year or multi-year average crime data in spatial analysis of crime. This will be the foundation for the second and third objectives by providing a rationale for the use of different modeling approaches.

2) To illustrate a modeling approach useful to examine the persistence of crime hot spots over longer time periods, and identify the neighborhood structural correlates that perpetuate or interrupt the presence of crime hot spots.

Whereas in objective 1 the outcome is evidence of spatial and temporal stability or lack thereof, as part of this objective the correlates that determine the persistence of crime hot spots.

Aside from the identified socio-economic correlates, recent studies on built environment and violent crime provides evidences that the environmental context can induce or inhibit violent crime (Sampson and Raudenbush, 2004; Spicer et al., 2012). Of particular interests are fine-scale physical characteristics of built environments. Hence, this is addressed in the third objective.

3) To investigate the role of fine-scale physical features of residential built environment in inducing and deterring violent crime. A Google Street View-based environmental audit approach is developed to enable quick and systematic collection of environmental characteristics associated with violent crime.

Based on addressing the above three objectives, this dissertation contribute to the literature of crime analysis by providing a comprehensive understanding of hot spots of violent crime. To be specific, geographical concentration of violent crime can be correlated to general characteristics of the social context and built environment in which offenders and victims are embedded.

1.3 Dissertation contents

The remainder of this dissertation is organized as follows.

Chapter 2 investigates the temporal stability of model parameters in crime rate analysis. The objective of this chapter is to investigate the spatial and temporal stability of violent crime patterns and its implications for modeling practice in Chapter 3 and 4. A difference of means test was used to detect significant fluctuation in crime data. Keeping in mind that social-economic indicator may explain and moderate this variation; this study argues that the criteria to support the averaging of crime data across years are not only the presence of year-by-year volatility, but also the temporal stability of relationships between socio-economic characteristics and crime. To verify this, a seemingly unrelated regression is used to jointly estimate the relationship between five single years (1998-2002) and five-year average violent crime rates and socio-demographic variables at the block group level. The Wald statistic test is used to test the temporal stability of estimated model parameters. With regard to the model estimation and Wald test results, if parameters of one group are not significantly different from those of other groups, we can conclude that the crime characteristics are stable across the period considered, and if the average group's performance is worse than other single year groups, there is no basis for using the average measure. Otherwise, the crime characteristics are heterogeneous in the study region, and if the average group performs better than other groups, using average measure may help to reduce the volatility and fluctuation within the data.

Chapter 3 investigate the persistence of hot spots of violent crime. Whereas in Chapter 2 the outcome is evidence of spatial and temporal stability of model parameters or lack thereof, as part of Chapter 3 the neighborhood structural correlates that perpetuate or interrupt the presence of crime hot spots. Nine years' worth of crime data and two decennial census datasets were used. The study first used the measure of cluster membership (being part of a hot spot) to capture the magnitude of the hot spot persistence over several years. An ordered probit regression was then used to explore the socio-economic correlates of cluster membership. The identified neighborhood characteristics provide insights to inform crime reduction initiatives.

Chapter 4 examines the relationship between built environment and violent crime, through an environmental audit approach using Google Street View. Together with Chapter 2 and 3, this chapter provides a comprehensive understanding of hot spots of violent crime from perspectives of both social context and physical features of built environment. This study first addressed the question of "where" to conduct environmental audit by using Poisson regression model with spatial filtering. After controlling for relevant socio-economic correlates, target sites can be determined by analyzing residual spatial pattern as retrieved by a spatial filter. Google Street View (GSV) was used to systematically identify physical features associated to violent crime, after socio-economic correlates have been controlled for in the model. A desktop data collection tool was developed to facilitate virtual audit data collection. Results obtained from the Poisson estimation and GSV-based environmental audit give credence to social disorganization theory, routine activity theory, and broken window theory. Results can help to provide practical insights for Crime Prevention Through Environmental Design (CPTED) initiatives and inform focused crime prevention efforts.

The dissertation concludes with a discussion of contributions, limitations, and future directions in Chapter 5.

Chapter 2 Temporal stability of model parameters in crime rate analysis

2.1 Introduction

A frequent task in spatial modeling of crime is the selection of a dependent variable, either single-year or multi-year average of crime. A number of studies prefer to use a single year and corresponding census data in regression analysis (e.g., Chamlin and Cochran, 2004; Andresen, 2006a; Bjerk, 2010). The purported reason for this, in addition to data availability, is that a gap between the year of census data and crime data may hide potential relationships between crimes and socio-economic characteristics (Andresen, 2006a). Meanwhile, others prefer to smooth crime data by taking the average of crime rates (or counts) over multiple years, in order to reduce the influence of volatility and fluctuation in the year-to-year occurrence of crime. This is, by far, the most common practice; see for example, Sampson (1985, 1987), Messner and Golden (1992), Peterson et al. (2000), Browning et al. (2000), Cahill and Mulligan (2003, 2007), Ye and Wu (2011), Light and Harries (2012), and Erdogan et al. (2013); for a review see Krivo and Peterson (1996). Admittedly, crime patterns may vary over multiple years, and the location of crime events can also markedly shift over an area's landscape from year to year. In this case, taking average of crimes may help to reduce volatility, wash off the potential effect of outliers, and demonstrate the mean characteristics of crime over a specific time period. To date, however, there is no evidence to support either approach (i.e. single-year or averaging). Lacking a statistical examination of the data, it is unclear to what extent the analysis of these averaged and smoothed secondary data would be useful for revealing the association between crimes and correlates of crime. Averaged crime data may be different with any single-year data, which does not represent any real crime process and may fail to provide meaningful information for understanding the spatial and temporal pattern of crime. This question is of interest and has implications for researchers to make appropriate modeling decisions.

Of particular interest is the potential temporal instability of crime patterns. Criminologists have long been aware of this issue, and within the literature, a frequent strategy used to statistically examine this temporal fluctuation is the difference of means test (e.g., Novak et al., 1999; Cahill and Mulligan, 2003). For instance, Cahill and Mulligan (2003) employed a five-year average because the single-year patterns were significantly different (higher or lower) from the five-year average. Admittedly, the difference of means test can test for differences between means from several separate groups of crime data. However, in regression analysis of crime, the appropriateness of using single-year or multi-year average crime data should not be purely determined by the appearance of annual fluctuations in crime data itself, but the temporal stability of crime process and characteristics, in other words, the temporal stability of relationships between socio-economic characteristics and crime. This is to say, yearly variation of crime data within a certain range may be accounted for by socio-economic characteristics derived from decennial census data. Therefore, it is necessary to examine the extent to which claims of socio-economic invariance are sensitive to the temporal dynamics of crime data.

This study attempts to fill this gap in the literature by empirically investigating the temporal stability of model parameters in crime rate analysis, and further assessing the appropriateness of using single-year or multi-year average data in spatial analysis of crime. The case study is violent crime in Columbus, Ohio, using socio-economic data derived from the U.S. Census 2000. We employ seemingly unrelated regressions (SUR) (Zellner, 1962; Davidson and MacKinnon, 1993) to jointly estimate the relationship between 5 single years (1998-2002) and five-year average violent crime rates and socio-demographic variables at the block group level. The Wald statistic test is then used to test the temporal stability of estimated model parameters. With regard to the SUR and Wald test results, if parameters of one group are not significantly different from those of other groups, we can conclude that the crime characteristics are stable across the period considered, and if the average group's performance is worse than other single year groups, there is no basis for using the average measure. Otherwise, the crime characteristics are heterogeneous in the study region, and if the average group performs better than other groups, using average measure may help to reduce the volatility and fluctuation within the data.

2.2 Context and data sources

2.2.1 Crime data

Our analysis relies on two sets of data, that is, crime data of the city of Columbus and decennial census data. With a population of 711,470 in 2000, Columbus is the capital and the largest city in the state of Ohio, and the 15th largest city in the U.S. Previous research by Browning et al. (2010) indicated that the city of Columbus has characteristics that reflect a wide range of places in the United States from the perspective of population composition, economic functions, commercial and industrial functions. Detailed daily crime data comes from the Columbus Division of Police, and covers a period of 9 years (from 1994 to 2002). Information provided by the dataset includes location of crime incident, date and time when crime happened, date when crime was cleared, Uniform Crime Reporting (UCR) code, and type of crime, among other attributes. Over 160 disaggregated types of crime are available in this large dataset. Criminal offenses are grouped in two main categories in criminology (Ellis and Walsh, 2000), that is, violent crime and property crime. Following FBI's UCR Program, violent crime is composed of four offenses: murder and non-negligent manslaughter, forcible rape, robbery, and aggravated assault. In our dataset, there are in excess of 4,000 violent crimes each year.

The dataset provides an accuracy of 95 per cent success rate for the geocoding of violent crime. Remaining unmatched points are addresses erroneously coded by police personnel. With such a high rate of successful matches and the randomness of unmatched points, there are no reasonable concerns about bias in the analysis (Ratcliffe, 2004).

In crime modeling, a large gap between census data and crime data may lead to a failure to explain crime patterns, and the potential relationships between crimes and neighborhood socio-demographic and economic characteristics may be biased (Andresen, 2006a). The most relevant census data for this crime dataset is Census 2000. Therefore, to reduce the magnitude of this gap, the final dataset is selected to include 5 years of data (1998, 1999, 2000, 2001, 2002). The reason of using 5 years of data is because our data is limited to 2002, so five years includes, in addition to census year data, the two preceding and subsequent years. There are 22,987 violent crimes in total over 5 years: 4,099 in 1998, 4,415 in 1999, 4,647 in 2000, 4,944 in 2001, and 4,882 in 2002. The violent crime rates in Columbus, which are calculated by dividing the number of violent crimes by the total residential population of the study area, have steadily increased from 6.87 per 1,000 population in 1998, to 7.39 per 1,000 population in 1999, 7.78 per 1,000 population in 2000 and 8.28 per 1,000 population in 2001. In 2002 the violent crime rate in Columbus was 8.17 per 1,000 population.

2.2.2 Census data

The second dataset is obtained from two institutions. One is the US Census Bureau, which provides Summary File 3 (SF3). The SF3 consists of Census 2000 general demographic, social, economic and housing characteristics compiled from a sample of approximately 19 million housing units (about 1 in 6 households) that received the Census 2000 long-form questionnaire. The SF3 provides independent variables at census block group level. The other data source is the National Historical Geographic Information System (Minnesota Population Center, 2011), which was launched in 2007 and is maintained by the Minnesota Population Center at the University of Minnesota. It is a historical GIS project to create and freely disseminate a database incorporating all available aggregate census information for the U.S. between 1790 and 2010. We use it to create supplementary socio-economic variables.

2.2.3 Spatial unit of analysis

There are numerous ways to create zoning systems (Páez and Scott, 2004), therefore, the modifiable areal unit problem (MAUP) has always been a typical issue in spatial analysis when aggregate geographical data is used. Likewise, the choice of spatial unit (scale) can significantly affect the result of crime analysis (Langford and Unwin, 1994; Andresen and Malleson, 2013). The MAUP can affect indices created from aggregate areal data (Páez and Scott, 2004). Appropriate spatial unit can help reducing the effect of MAUP and ecological fallacy in crime data and socio-economic indicators (Andresen, 2006a; Andresen and Malleson, 2011). Smaller census unit is preferable to large unit because data can be less aggregated and averaged (Oberwittler and Wikström, 2009). Frequently used geographic units in spatial crime analysis are:

census tract (e.g., Browning et al., 2010), census block group (e.g., Cahill and Mulligan, 2003), census block (e.g., Bernasco et al., 2013), dissemination area (e.g., Law and Quick, 2013), enumeration area (e.g., Andresen, 2006a, 2006b), face block (e.g., Schweitzer et al., 1999), street block and more micro spatial units such as street segment (e.g., Weisburd et al., 2012). Census tract and census block are respectively the largest and smallest census unit in the U.S. Previous research suggests that crime analysis carried out at census tract level may obscure much of the geographic variability that captures violent crime (Cahill and Mulligan, 2003). But census data are not available at census block level. Face block, street block and street segment are commonly used for studying micro built environmental characteristics.

Consequently, in this research we use census block group (BG) which is between the level of census tract and the census block and is the smallest geographical unit for which the U.S. Census Bureau tabulates and publishes sample data. Typically, a BG has a population of 600 to 3,000 people. In this research, the study area does not cover the whole city and the reason is twofold. The first is the contingency of BGs. There are some "holes" (suburbs) in BGs in Columbus. Crime records in these BGs are generally not maintained by Columbus Division of Police which is our crime data source, thus they are excluded from the study area. The other reason is that the U.S. census sets a threshold (100 people in a specific geographic unit) for data availability, which means that census data is only available for those census units having a population more than 100 (U.S. Census Bureau, 2003). In our dataset, four randomly distributed BGs have to be excluded because data are not available for them. As a result, those continuous BGs with available census data are selected as the study area, and there are finally 567 BGs in the study area, out of a total of 571.

2.3 Methods

2.3.1 Exploratory spatial data analysis

One of the most common and important tasks in Exploratory Spatial Data Analysis (ESDA) is examining spatial autocorrelation in the data (Rangel et al., 2010). Social science variables are usually positively spatially autocorrelated due to the way phenomena are geographically organized (Griffith, 2003). Therefore, it is necessary to explore the presence of spatial autocorrelation among our variables first. The global Moran's I statistic introduced by Moran (1950) is the most popular measure of spatial autocorrelation, and it takes the form (see Griffith and Amrhein, 1991):

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2} \tag{1}$$

where *n* is the total number of spatial features, X_i and X_j are values for feature *i* and *j*, $\bar{X}$ is the mean of *X*, $(X_i - \bar{X})$ indicates the deviation of *i*'s feature value X_i from its mean $\bar{X}$, and w_{ij} are the elements of the spatial weight matrix, indicating the spatial weight between feature *i* and *j*.

Moran's I represents a linear association between a feature value and the spatially weighted average of neighboring values (Ratcliffe, 2010). The expected value of

Moran's I statistic is $-1/(n-1)$, which tends to zero as the sample size increases. An I coefficient larger than $-1/(n-1)$ indicates positive spatial autocorrelation, and less than $-1/(n-1)$ indicates negative spatial autocorrelation, and a Moran's I approaching $-1/(n-1)$ indicates an absence of spatial autocorrelation (Griffith and Amrhein, 1991). Hypothesis testing against the null hypothesis of spatial independence can be evaluated by a z-score and p-value. The z-score values associated with a 95 percent confidence level are -1.96 and +1.96 standard deviations, indicating that if the computed z-score fall between -1.96 and +1.96, the p-value is larger than 0.05, and the spatial autocorrelation is not statistically significant, the observed pattern could very likely be the result of random spatial processes.

2.3.2 Seemingly unrelated regression

We use multiple equation models to estimate yearly relationships between socio-economic characteristics and violent crime, and use a testing method to investigate the stability of these relationships across years. Although we could estimate models for each single year separately and test stability of parameters of each model, it is possible that the models will not be independent due to cross-equation error correlations. Therefore, it is appropriate to take these correlations into account. Zellner's seemingly unrelated regression (Zellner, 1962) is used to estimate models of violent crime rate as a function of a host of social disorganization variables.

As one type of multiple equation regression, seemingly unrelated regression (SUR) is defined as a generalization of a linear regression model that consists of several regression equations. In SUR, a set of equations that have their own dependent and independent variables have cross-equation error correlation, i.e. sharing interdependent unmeasured causes (Light and Harris, 2012). The SUR is performed in two steps. Firstly, ordinary least squares (OLS) is performed to estimate the variance and covariance of cross-equation error term. Secondly, generalized least squares (GLS) is used in the equation system for cross-equation error correlation. Error term's variance and covariance obtained from the residuals of previous OLS regression are used to estimate model parameters. Compared to OLS, SUR is able to provide more efficient and precise estimates of model parameters and standard errors by comprehensively employing all the information in the equation system and considering the interdependent unmeasured causes which can similarly influence violent crime in different years (Zellner, 1962). In addition, it provides a unified framework to test the hypothesis of parameter volatility across equations.

SUR is defined in the following way. Assume that there are M equations in the system. Then, the i'th equation is as follows:

$$y_i = X_i\beta_i + \varepsilon_i\ ,\ i = 1, \dots, M \quad (2)$$

where i represents the equation number, the i'th equation has a single dependent variable y_i which is an $R \times 1$ vector (R is the observation index), X_i is an $R \times k_i$ design matrix, where k_i is the number of regressors, β_i is a $k_i \times 1$ vector of the regression coefficients, and ε_i is an $R \times 1$ vector which indicates random error terms. All these M vector equations can be stacked, and the SUR system then becomes (Zellner, 1962):

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix} \quad (3)$$

Equations (2) and (3) can be written more compactly as:

$$y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \Omega) \quad (4)$$

where y is an $MR \times 1$ vector, X is an $MR \times k$ matrix, β is a $k \times 1$ vector and $k = \sum_{i=1}^{M} k_i$, and ε is an $MR \times 1$ vector of disturbances.

The model assumes that the error terms ε have cross-equation correlations, that is:

$$E(\varepsilon_{is}\varepsilon_{jt}) = \begin{cases} \sigma_{ij} & \text{if } s = t, \\ 0 & \text{else} \end{cases}$$

where ε_{is} and ε_{jt} are the error terms of the s'th observation of the i'th equation and the t'th observation of the j'th equation.

The $MR \times 1$ disturbance vector can be assumed to have a variance-covariance matrix as follows:

$$\Omega = E(\varepsilon\varepsilon') = E\left[\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_M\end{bmatrix}\begin{bmatrix}\varepsilon_1' & \cdots & \varepsilon_M'\end{bmatrix}\right] = E\begin{bmatrix}\varepsilon_1\varepsilon_1' & \varepsilon_1\varepsilon_2' & \cdots & \varepsilon_1\varepsilon_M'\\ \varepsilon_2\varepsilon_1' & \varepsilon_2\varepsilon_2' & \cdots & \vdots\\ \vdots & \vdots & \ddots & \vdots\\ \varepsilon_M\varepsilon_1' & \cdots & \cdots & \varepsilon_M\varepsilon_M'\end{bmatrix}$$

$$= \begin{bmatrix}\sigma_{11}I_R & \sigma_{12}I_R & \cdots & \sigma_{1M}I_R\\ \sigma_{21}I_R & \sigma_{22}I_R & \cdots & \vdots\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{M1}I_R & \cdots & \cdots & \sigma_{MM}I_R\end{bmatrix} = \sum \otimes I_R \quad (5)$$

where I_R is an R-dimensional identity matrix, $\otimes$ denotes the Kronecker product, and $\sum$ is:

$$\sum \equiv \begin{bmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M}\\ \sigma_{21} & \sigma_{22} & \cdots & \vdots\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{M1} & \cdots & \cdots & \sigma_{MM}\end{bmatrix} \quad (6)$$

The elements of matrix $\sum$ are estimated using residuals from the OLS regression:

$$\hat{\sigma}_{ij} = \frac{\hat{\varepsilon}_i'\hat{\varepsilon}_j}{R} \quad (7)$$

The GLS estimator of β in equation (4) is given by:

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \quad (8)$$

And its variance-covariance matrix is:

$$V(\hat{\beta}) = (X'\Omega^{-1}X)^{-1} \quad (9)$$

Anselin (2006) suggested that spatial regression should be employed particularly for studies which use spatially aggregated socio-demographic and economic data.

Given the detected spatial autocorrelation in dependent and independent variables (See Table 2.1), we take spatial effects into account by extending traditional SUR to the SUR with a spatial lag process. We assume a spatial lag effect in the SUR model based on theoretical considerations. Geographical space increasingly plays a core role in crime analysis (Messner and Anselin, 2004). Spatial diffusion processes, for instance, have been consistently found in violent crimes (Morenoff and Sampson, 1997; Cohen and Tita, 1999; Messner et al., 1999; Smith et al., 2000; Morenoff et al., 2001). In addition, beneficial effects of crime prevention strategies in a neighborhood also diffuse to adjacent neighborhoods (Bowers and Johnson, 2003; Weisburd et al., 2006). This suggests that crime rate and socio-economic characteristics in one spatial unit may be systematically related to those in adjacent units, thus resulting in spatial dependence (Smith et al., 2000). Also, the advantage of embracing a spatial lag process in the SUR model is twofold. First, the correlations among the errors in different equations (also different years) will be utilized to improve the regression estimates. Second, spatial correlation among geographic units will be taken care by spatial lag terms and utilized to improve coefficient estimation. The SUR with spatial lag process can be in a following form:

$$y = \lambda W y + X\beta + \varepsilon, \quad \varepsilon \sim N(0, \Omega) \tag{10}$$

where W is a R by R matrix of weights with elements w_{ij} represent the proximity of spatial unit i and j. Based on the Queen's 1st order contiguity protocol as suggested by Andresen (2011), a spatial weight matrix is created using GeoDa 1.4.0 (Anselin et al., 2006). The coefficient λ quantifies the relationship between crime rate y in one BG and crime rates in contiguous BGs. A statistically significant λ can be interpreted as that a spatial diffusion process of crime rate is detected (Doreian, 1980). We calculate the spatial lag of dependent variable $y^L = Wy$ and spatial lag of all independent variables, and apply them in the estimation of the SUR model. The variable y^L can be understood as the "violent crime potential" similar to the "street robbery potential" studied in Smith et al. (2000). As suggested by Tita and Radil (2010), it is feasible to combine the response lag and predictor lag terms in a single model. We use the SUR model with spatial lag process to simultaneously estimate equations using crime rates of different years as dependent variables and all independent variables selected above, in order to obtain the model parameters of each group (i.e. each equation in the model system).

2.3.3 Wald statistic test

A key feature of SUR is that we can test differences between predictors across equations. Based on the estimated parameters of the SUR models, we use the Wald statistic test to investigate whether the correlates of crime vary in their effects on violent crime over time (Clogg et al., 1995). The null hypothesis (H_0) is that the parameter (β_i) for one independent variable in one equation is statistically equal to that of another equation (β_j), that is:

H_0: $\beta_i = \beta_j$

H_A: $\beta_i \neq \beta_j$

This is the same as testing the following hypothesis:

H_0: $\beta_i - \beta_j = 0$

H_A: $\beta_i - \beta_j \neq 0$

For testing the null hypothesis, a general formula for the computation of the Wald statistic is as follows:

$$W = \left(\hat{\beta}_i - \hat{\beta}_j\right)' \left[var\left(\hat{\beta}_i\right) + var\left(\hat{\beta}_j\right)\right]^{-1} \left(\hat{\beta}_i - \hat{\beta}_j\right) \quad (11)$$

where β_i and β_j are the parameter vectors including all parameter estimates for two equations, $var(\cdot)$ is the estimated variance-covariance matrix for the parameters. The Wald statistic is asymptotically distributed as χ^2 distribution under the null hypothesis.

If the test fails to reject the null hypothesis at an acceptable level of confidence, it can be interpreted to mean that the parameter of one independent variable in one equation is not significantly different from that corresponding to the same independent variable in other equations. Therefore, the relationship between crime rates and socio-economic characteristics are stable across years, and if the average group's performance is worse than other single year groups, there will be no point in using multi-year average crime rate. Whereas, if the null hypothesis can be rejected at a significant level, model parameters of one equation are significantly different from those of others, it can be interpreted to mean that the relationship is significantly heterogeneous over years. In this case, if the average group performs better than other groups, using multi-year average crime rate may help to minimize the impact of annual fluctuations and reduce the potential bias of a single high or low crime year.

2.4 Analysis and results

2.4.1 Dependent variables

The present research investigates the appropriateness of dependent variable selection by examining the temporal stability of correlations between socio-economic indicators and violent crime rates. The intent is to test the parameter stability of well-proved socio-economic variables in each single year in order to examine if the criminal processes and characteristics are stable across years. Six dependent variables are used: violent crime rate of the year of 1998, 1999, 2000, 2001, 2002, and five-year average. The reason we investigate violent crime in this study is simply because of the data availability. The reason we use crime rate rather than count is that crime count which measures the quantity of criminal activity is a quite restrictive measure. Andresen (2006a) pointed out that as an absolute term it does not give an indication of why crime happens more in certain places but less in others. We employ the measure of crime rate to evaluate the risk of crime by controlling for the population. The rate is measured per 1,000 residential population living in the BG. Violent crime rate in this study is composed of all four violent offenses: murder and non-negligent manslaughter, forcible rape (including intent to rape), robbery, and aggravated assault. The distribution of crime rate is long tailed, and as a matter of practice it is transformed using the natural logarithm (Smith et al., 2000). The transformation also ensures that back transformed estimates of crime rate are strictly positive. Table 2.1 gives the acronyms (VR98, VR99, VR00, VR01, VR02, AVG), definitions and descriptive statistics for these dependent variables.

2.4.2 Independent variables

The independent variables in this research are selected based on the social disorganization theory (Shaw and McKay, 1942; Sampson and Groves, 1989), one of the most important theories in criminology. Social disorganization refers to "the inability of local communities to realize the common values of their residents" (Bursik, 1988, p. 521). Rooted in social ecology (Park et al., 1967), social disorganization theory looks for explanations of criminal activity in general characteristics of the social structure and social control on which offenders and victims are embedded. Over the last several decades, a large body of studies has emerged and sought to test the empirical validity of the theory by using multivariate statistical modeling approaches and various crime data sources. They have consistently shown that socio-economic deprivation, residential mobility (or population turnover), family disruption, and ethnic heterogeneity lead to community social disorganization, which in turn, increases violent crime rates (Andresen, 2006a, 2006b; Cahill and Mulligan, 2003; Charron, 2009; Porter and Purser, 2010; Sampson, 1985).

We use a series of variables to capture the four structural elements of the theory: socio-economic deprivation, residential mobility, family disruption, and ethnic heterogeneity. All elements have been recognized to characterize socially disorganized and distressed areas, and have been identified as main sources of social disorganization (Cahill, 2004; Law and Quick, 2013). All variables below are proxy measures of the theory and have been informed by an extensive review of the relevant literature (e.g., Cahill and Mulligan, 2003; Sun et al., 2004; Andresen, 2006a, 2006b; McCord and Ratcliffe, 2007; Ye and Wu, 2011; Law and Quick, 2013). Socio-economic deprivation is measured as percentage of population below the poverty level and percentage of unemployed population. Residential mobility is represented by percentage of renter occupied housing units and percentage of population living in households and are nonrelatives. Family disruption is measured through number of single-parent families. Ethnic heterogeneity is captured by percentage of foreign born population and racial heterogeneity measured by means of the Heterogeneity Index (see equation 12) which is designed to capture the degree of racial heterogeneity in each BG (Blau, 1977; Sampson & Groves, 1989; Cahill and Mulligan, 2003).

$$\text{Heterogeneity Index } = 1 - \sum p_i{}^2 \qquad (12)$$

where i indicates the number of racial group, p_i is the proportion of the population in a given group (Sampson and Groves, 1989). The index is defined as a function of the different racial groups within a given area, and considers both number of racial groups and population size of each group. The range of the index is (0, 1), within which 0 means maximum homogeneity, and 1 means maximum heterogeneity. We use all 5 different racial groups to generate the index, i.e. White, Black or African American, American Indian and Alaska Native, Asian, and Native Hawaiian and Other Pacific Islander.

In spatial analysis of crime, it is important to consider the spatial dependency in the data. BGs are relatively small geographic units, thus socio-economic characteristics of any BG may not only affect its own crime rate but also crime rates of its neighboring BGs. For this purpose, we also consider the spatial lag of explanatory variables for spatial modeling. Spatial weight matrix is calculated based on the Queen's 1st order contiguity by using GeoDa 1.4.0. Table 2.1 provides descriptive statistics of all these explanatory variables.

Table 2.1 Descriptive statistics for variables used in the analysis (N = 567 observations)

Acronym	**Variable description**	**Mean**	**S.D.**	**Min.**	**Max.**	**Moran's I[a]**
Dependent variables						
VR98	Violent crime rate per 1,000 inhabitants in 1998 (log transformed)	0.05	4.56	-11.51	6.10	0.29 (11.59)

VR99	Violent crime rate per 1,000 inhabitants in 1999 (log transformed)	0.13	4.53	-11.51	5.80	0.20 (7.88)
VR00	Violent crime rate per 1,000 inhabitants in 2000 (log transformed)	0.31	4.39	-11.51	6.66	0.23 (9.18)
VR01	Violent crime rate per 1,000 inhabitants in 2001 (log transformed)	0.47	4.30	-11.51	6.91	0.20 (7.90)
VR02	Violent crime rate per 1,000 inhabitants in 2002 (log transformed)	0.55	4.13	-11.51	6.10	0.23 (9.02)
AVG	Five-year average violent crime rate per 1,000 inhabitants (log transformed)	0.91	3.49	-11.51	6.32	0.24 (9.08)
Independent variables						
POVRTY	Percent of population below the poverty level in 1999	18.52	15.91	0.00	76.81	0.53 (19.45)
UEMPLOY	Percent of population age 16 and over in civilian labor force that is unemployed	3.99	3.42	0.00	30.77	0.25 (10.07)
RENTER	Percent of housing units that are renter occupied	49.77	28.81	0.00	100.00	0.43 (16.18)
NOFAM	Percent of population living in households and are nonrelatives	9.34	8.75	0.00	65.18	0.59 (23.68)
SIGFAM	Number of single-parent families (log transformed)	3.35	2.14	-11.51	5.98	0.29 (10.96)
FOREIGN	Percent of total population that is foreign born	6.09	8.40	0.00	95.86	0.44 (16.47)
HI	Heterogeneity Index	0.33	0.16	0.00	0.76	0.50 (18.95)

POVRTY^L^	Spatially lagged POVRTY	18.30	11.97	1.18	59.35	0.83 (29.66)
UEMPLOY^L^	Spatially lagged UEMPLOY	3.97	2.04	0.34	12.36	0.69 (26.51)
RENTER^L^	Spatially lagged RENTER	49.70	20.14	10.56	98.68	0.78 (27.51)
NOFAM^L^	Spatially lagged NOFAM	9.15	6.49	3.46	43.43	0.91 (35.59)
SIGFAM^L^	Spatially lagged SIGFAM	3.38	1.54	-11.51	5.24	0.71 (26.78)
FOREIGN^L^	Spatially lagged FOREIGN	6.08	5.71	0.00	49.77	0.76 (27.49)
VR98^L^	Spatially lagged dependent variable VR98	0.19	2.80	-11.51	3.44	0.72 (27.09)
VR99^L^	Spatially lagged dependent variable VR99	0.38	2.57	-11.51	3.56	0.64 (23.89)
VR00^L^	Spatially lagged dependent variable VR00	0.54	2.55	-11.51	3.69	0.72 (25.06)
VR01^L^	Spatially lagged dependent variable VR01	0.66	2.38	-9.18	3.73	0.67 (25.05)
VR02^L^	Spatially lagged dependent variable VR02	0.79	2.26	-11.51	3.86	0.69 (25.87)
AVG^L^	Spatially lagged dependent variable AVG	1.09	2.00	-8.96	3.61	0.73 (27.83)

Note: The numbers in parentheses are z-scores.

[a] We calculate the global Moran's I Index values and z-scores using GeoDa 1.4.0, this is also done by using Queen's 1st contiguity weight which is also created using GeoDa 1.4.0.

The correlations test (see Table 2.6 in the Appendix) for all independent variables indicates that none of the pair-wise correlations between explanatory variables is larger than 0.8, which is a common threshold for the issue of collinearity (Andresen, 2006a). The variable of poverty population has a strong positive relationship (0.702) with the spatial lag of itself, the variable of population living in households and are nonrelatives has a strong positive relationship (0.797) with the spatial lag of itself. Both are standard relationships. In addition, the correlations among all six spatially lagged dependent variables are strong. Particularly, the correlations between the spatially lagged 5-year average crime rate and other 5 single years are all strong. However, considering the stable patterns of the geographic distribution of crime rates in 5 years and 5-year average (see Figure 2.1), these high correlations can be seen as standard relationships.

2.4.3 Empirical results

Maps showing the spatial distribution of violent crime rates in all five years appear in Figure 2.1. All maps are created using ArcGIS 10.1 (ESRI, 2012). The violent crime rate in 1998 shows a pattern of lower crime rate in the northern and southern areas and higher crime rate in the middle areas of the city, especially downtown areas and those found in eastern, western and southern immediate surrounding areas. In other 4 years and 5-year average, the geographic distribution of crime rate shows a similar pattern as in 1998, the BGs with the highest crime rates are those consistently found in downtown areas and in the immediate surrounding areas. The extreme northern, northwest, northeast, southwest, and southeast edges of the city experience very low crime rates. Also, it should be noted that the extreme western BG always belongs to the highest crime rates category in all five years.

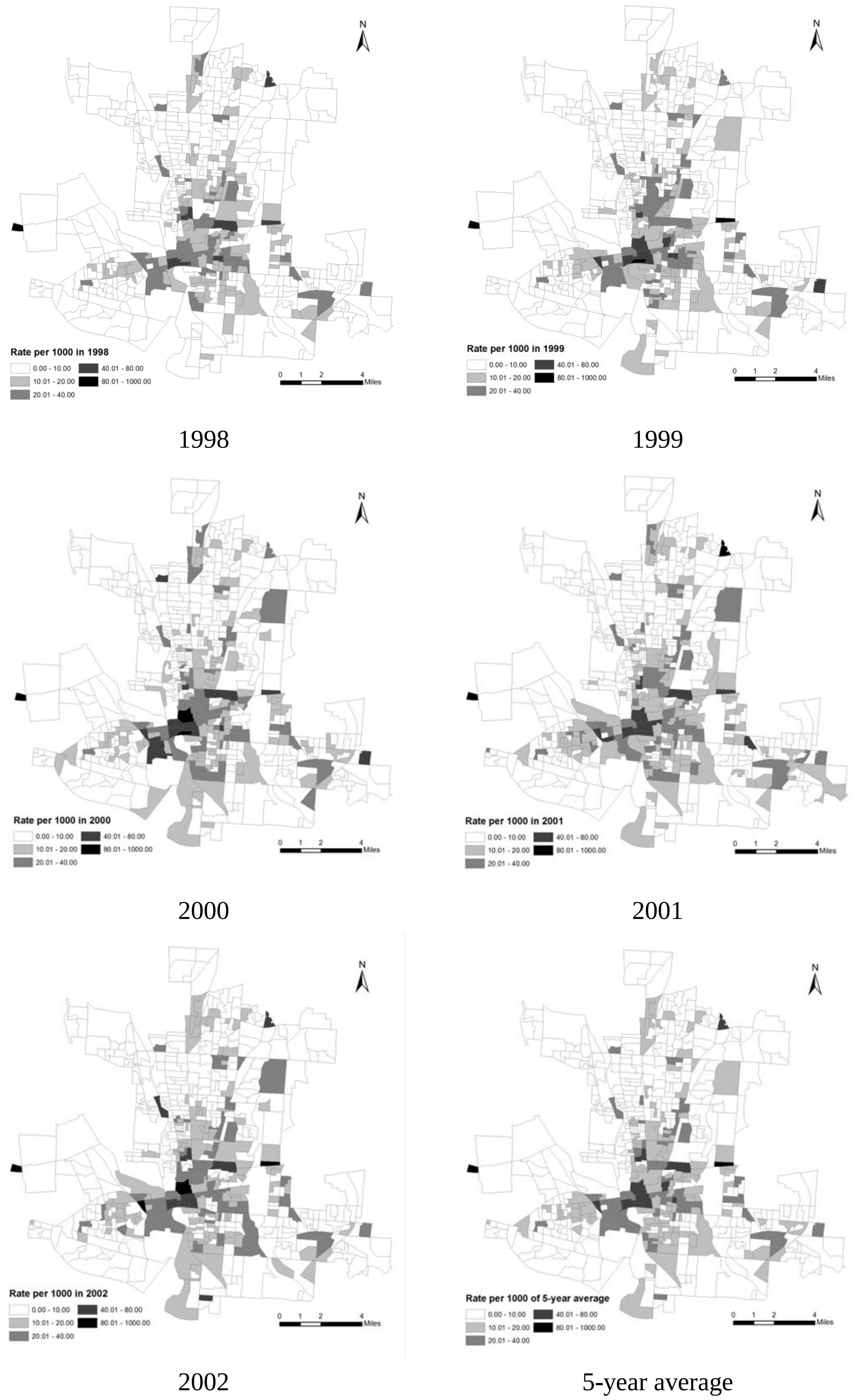

Figure 2.1 Violent crime rates per 1,000 in Columbus in five single years and five-year average

The results of global Moran's I indicate significant spatial autocorrelation in violent crime rates (See Table 2.1). The patterns observed in Figure 2.1 are consistent with the test and show that high crime rates tend to be clustered and low crime rates also tend to be clustered. The result also suggests a decreasing spatial autocorrelation of violent crime rate from 1998 to 1999, but a stable spatial autocorrelation from 1999 to 2002. Although it has been suggested that spatial autocorrelation at this level "is not a major issue" (Cahill and Mulligan, 2003, p. 591), for the purpose of unbiased spatial modeling, we believe it is appropriate to adjust for spatial autocorrelation in the analysis.

Following common practice (e.g., Novak et al., 1999; Cahill and Mulligan, 2003), the difference of means test is performed using Stata 11.2 (StataCorp., 2011) to compare crime rates between five single years and five-year average. Table 2.2 provides the results of paired-samples t-tests. There are three observed yearly fluctuations in crime data, violent crime rates of two single years (1998 and 1999) are significantly lower than the average crime rate, and one year (2001) is significantly higher than the average crime rate. In light of the temporal stability, the common practice followed by the literature as reviewed above would suggest the use of the five-year average measure in this case. However, we propose that the criteria to support the averaging of crime data across multiple years are not only the presence of year-by-year volatility, but also the temporal stability of relationships between socio-economic characteristics and crime.

Table 2.2 Results of t-tests of crime rates between five single years and five-year average

	Group							
	Single year			5-year average				
Comparison	M	S.D.	n	M	S.D.	n	t	*p*-value
1998 vs. Avg[a]	9.04	20.73	567	10.20	25.17	567	-4.60	< 0.001
1999 vs. Avg	9.44	17.59	567	10.20	25.17	567	-1.80	0.037
2000 vs. Avg	10.78	34.19	567	10.20	25.17	567	1.35	0.178
2001 vs. Avg	11.68	43.04	567	10.20	25.17	567	1.84	0.033
2002 vs. Avg	10.51	21.96	567	10.20	25.17	567	1.12	0.265

Note: [a] Avg stands for the 5-year average group.

To verify this, we test the stability of model parameters for the SUR system with respect to socio-economic predictors in the five years under consideration. The SUR and Wald test are performed using Stata 11.2. Table 2.3 provides the estimation results for the SUR without controlling for spatial effect, i.e. excluding the spatial-lagged dependent variables, and the Moran's I coefficients for residuals in all groups. The results show that Moran's I of residuals in all groups are significant, which indicates the existence of spatial autocorrelation in residuals before any spatial effects were controlled for in the model. This finding suggests that it is necessary to take spatial effects into consideration

Table 2.3 Parameter estimates from the SUR without spatial lag and Moran's I of residuals

Variables	1998	1999	2000	2001	2002	AVG
Constant	-3.212***	-3.462***	-2.854***	-1.667**	-2.637***	-1.084*
POVRTY	-	-	-	-	-	-
UEMPLOY	0.275***	0.318***	0.313***	0.274***	0.307***	0.226***
RENTER	-	-	-	-	-	-
NOFAM	-	-	-	-	-	-
SIGFAM	0.542***	0.472***	0.351**	0.335**	0.433***	0.273**
FOREIGN	-	-	-	-	-	-
HI	-	-	-	-	-	-
$POVRTY^L$	-	-	-	-	-	-
$UEMPLOY^L$	-	-	-	-	-	-
$RENTER^L$	-	-	-	-	-	-
$NOFAM^L$	0.093*	0.126**	0.124**	0.091**	0.118***	0.091**
$SIGFAM^L$	-	-	-	-	-	-
$FOREIGN^L$	-0.132**	-0.114**	-0.109**	-0.190***	-0.132***	-0.134***
Moran's I[a]	0.173 (6.435)	0.069 (2.928)	0.135 (5.678)	0.085 (3.411)	0.111 (4.516)	0.106 (4.264)
R^2	0.111***	0.109***	0.097***	0.120***	0.128***	0.109***
N^b	567	567	567	567	567	567

Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

[a] We calculate the global Moran's I Index values and z-scores (the numbers in parentheses) for residuals using GeoDa 1.4.0, this is also done by using Queen's 1st contiguity weight.

[b] *N* is the number of observation.

Table 2.4 provides the estimation results for the SUR system with spatial lag process. Compared to the Table 2.3, the goodness of fit (as measured by R^2) for all groups are improved. Most importantly, the Moran's I Index values and z-scores of residuals for all groups suggest that the spatial autocorrelation in residuals is not significant anymore. This supports the use of spatially lagged dependent variables in controlling spatial effects in the model. In Table 2.4, the R^2 is stronger in the case of the 1998 (0.224) and 2002 (0.210) groups. The 1999 group has the lowest R^2 (0.141), lower than the 2000 (0.167), 2001 (0.150) and five-year average (0.152) groups. This is because social-economic conditions have both contemporaneous and lagged effects on crime rates (Rosenfeld & Fornango, 2007). Performance of the five-year average group is lower than the three best performing single year groups. The Breusch-Pagan test of independence suggests that the residuals from the six equations are not independent (p<0.001), which indicates that the SUR estimation is more appropriate in this case.

Table 2.4 Parameter estimates from the SUR with spatial lag and Moran's I of residuals

Variables	1998	1999	2000	2001	2002	AVG
Constant	-2.356***	-3.095***	-2.473***	-1.575*	-2.409***	-1.034*
POVRTY	-	-	-	-	-	-
UEMPLOY	0.148*	0.274***	0.244***	0.236***	0.241***	0.187***
RENTER	-	-	-	-	-	-
NOFAM	-	-	-	-	-	-
SIGFAM	0.411***	0.413***	0.292**	0.302**	0.375***	0.242**
FOREIGN	-	-	-	-	-	-
HI	-	-	-	-	-	-
$POVRTY^L$	-	-	-	-	-	-
$UEMPLOY^L$	-	-	-	-	-	-
$RENTER^L$	-	-	-	-	-	-
$NOFAM^L$	0.074*	0.111**	0.100**	0.084*	0.106**	0.080**
$SIGFAM^L$	-	-	-	-	-	-
$FOREIGN^L$	-0.103**	-0.103**	-0.088*	-0.173***	-0.116**	-0.119***
Spatially lagged	0.473***	0.204**	0.339***	0.195**	0.302***	0.206**

dependent variable						
Moran's I[a]	-0.032 (-1.080)	-0.015 (-0.512)	-0.014 (-0.515)	0.008 (0.350)	-0.001 (0.051)	0.020 (0.872)
R^2	0.224***	0.141***	0.167***	0.150***	0.185***	0.152***
N[b]	567	567	567	567	567	567

Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

[a] We calculate the global Moran's I Index values and z-scores (the numbers in parentheses) using GeoDa 1.4.0, this is also done by using Queen's 1st contiguity weight.

[b] N is the number of observation.

Results of the Wald statistic test for exploring the model parameter stability of the SUR equation system are shown in Table 2.5. Only the test for constants of the 1998 group and five-year average group can reject the null hypothesis at a significant level, indicating that the constant differs in its relationship with violent crime rates across the two groups. All other four variables of the two groups fail to reject the null hypothesis at a significant level, suggesting that there are not any significant differences in relationship with violent crime rates across the 1998 and five-year average group for all four independent variables. When comparing other four groups (1999, 2000, 2001, 2002) with the average group respectively, we do not find any significant parameter differences across groups for all independent variables and constants. That is, the association of all independent variables with violent crime rates is not significantly different across these groups. The Wald test in the final column of Table 2.5 which jointly tests all six groups reveals the same fact as before. Therefore, we can conclude that the crime characteristics are stable across five years.

Table 2.5 Wald test for parameter stability of the SUR equations

	98 vs. Avg[a]	99 vs. Avg	00 vs. Avg	01 vs. Avg	02 vs. Avg	All 5 groups
Constant	4.61*	3.76	0.54	1.04	0.98	6.65
UEMPLOY	0.47	2.72	1.24	0.95	1.48	5.99
NOFAM^L	0.03	1.07	0.48	0.02	1.15	5.94
SIGFAM	3.09	3.57	0.32	0.48	3.17	2.44
FOREIGN^L	0.24	0.27	1.09	3.07	0.02	8.43

Note: * $p < 0.05$, ** $p < 0.01$. The null hypothesis of Wald test for the first column is H_0: $\beta_{98} - \beta_{Avg} = 0$, the second column is H_0: $\beta_{99} - \beta_{Avg} = 0$, the third is H_0: $\beta_{00} - \beta_{Avg} = 0$, the fourth is H_0: $\beta_{01} - \beta_{Avg} = 0$, the fifth is H_0: $\beta_{02} - \beta_{Avg} = 0$, and the last is H_0: $\beta_t - \beta_{Avg} = 0$ for $t = 1998, \dots, 2002$.

[a] Avg stands for the 5-year average group.

2.5 Discussion

A synthesis of the results of SUR model and Wald test leads to several relevant conclusions. Of particular importance is that, although the difference of means test suggests significant yearly volatility in crime data, both the SUR model and Wald test show different results. Table 2.4 and 2.5 suggest that there are not any significant differences in relationship with violent crime rates across all single year groups and five-year average group for all four significant explanatory variables. The results also indicate that there is a considerable stability in the signs of the explanatory variables across all groups. In other words, the crime characteristics are stable across five years. This finding yields twofold implications. The first is in regard to the appropriateness of averaging crime data across multiple years. The second is in regard to the selection of dependent variables in spatial analysis of crime.

In previous studies, the main criterion for averaging crime data across multiple years is the presence of annual fluctuation, that is, if it is observed by using the difference of means test, crime data of multiple years should be averaged. However, our findings reveal that the criteria to support the averaging of crime data should be not only the presence of year-by-year volatility, but also the temporal stability of relationships between socio-economic characteristics and crime.

Our findings also reveal that averaging crime data across multiple years is not necessarily superior to single year crime rate. Previous studies suggest that in case of the annual fluctuation, using average crime rate is able to reduce the influence of fluctuation in the year-to-year occurrence of crime. But this study answers an important question of this practice, that is, how useful will the analysis of these averaged data be for revealing the relationship between crimes and socio-demographic and economic characteristics of every single year. Our analysis suggests that, although the difference of means test indicates that crime rates in 1998 (lower), 1999 (lower) and 2001 (higher) are different from the average crime rate, the results of SUR model and Wald test suggest that all significant independent variables provide statistically identical effects on crime rates of all single years and five-year average. In other words, single-year crime rates can exhibit similar performance in crime modeling analysis as average crime rate. What’s more, considering the goodness of fit, the 1998, 2000 and 2002 groups perform better than the average group.

Other key findings of this analysis are as follows. First, in agreement with previous studies, all significant social disorganization variables have the expected signs. To be specific, net of other explanatory variables, (1) socio-economic deprivation which is captured by percentage of unemployed population (UEMPLOY) has a statistically significant, positive association with violent crime rates; (2) residential mobility captured by lag of percentage of population living in households and are nonrelatives ($NOFAM^L$) is significantly and positively related to violent crime rates; (3) family disruption captured by log transformed number of single-parent families (SIGFAM) is also significantly and positively related to violent crime rates; (4) ethnic heterogeneity characterized by lag of percentage of foreign-born population ($FOREIGN^L$) has a statistically significant, negative association with violent crime rates. Among them, socio-economic deprivation is the strongest predictor in the model. In contrast, family disruption's effect on violent crime rates is relatively modest in comparison to other predictors in the model. In addition, the spatially lagged dependent variables are also significantly and positively related to violent crime rates, indicating that proximity to BGs with high rates of violent crime corresponds to higher violent crime rates. This also proves that the consideration of spatial lag effect is necessary for this model.

Second, the spatial lag form of ethnic heterogeneity shows significant impact is because ethnic groups tend to show a cluster pattern. The variable foreign-born population has a high Moran's I value (0.44) with a high z-score (16.47), suggesting that BGs with high percent of foreign-born population tend to significantly cluster across the city. What's more, the Census 2000 shows that population of each ethnic group (White, Black or African American, American Indian and Alaska Native, and Asian) tend to cluster in Columbus, either intentionally or unintentionally.

Third, the percentage of unemployed population, percentage of population living in households and are nonrelatives, and number of single-parent families are modestly stronger for the 1999 group than it is for other five groups, indicating that the socio-economic deprivation, residential mobility and family disruption elements impact violent crime rate in 1999 to a greater extent than crime rates in other years and five-year average.

Fourth, the spatial autocorrelation in the residuals is carefully investigated in this analysis. Comparing the Moran's I Index values and z-scores of residuals in the SUR without controlling spatial effects and the SUR with spatial effects were controlled for, we found that introducing spatially lagged dependent variables into the model are effective in removing the spatial autocorrelation from residuals. This finding indicates the importance of taking spatial effects into consideration.

Whilst this study addresses one specific issue of the selection of dependent variable, in recent years, a number of other important statistical issues in crime analysis and their impacts on crime analysis have been investigated. For example, the appropriateness of aggregating crime across types (Andresen and Linning, 2012); measure of population at risk (Andresen, 2011); the choice of spatial unit of analysis (Ouimet, 2000; Wooldridge, 2002; Jacob, 2006; Andresen and Malleson, 2013); and the choice of spatial autocorrelation statistic (Townsley, 2009; Bernasco and Elffers, 2010; Tita and Radil, 2010). Keeping these studies in mind we concentrate only on the selection of dependent variable in this study. Admittedly, the conclusions we obtained may be influenced by the statistical issues above. For example, first, violent crime rate in this study is composed of all four violent offenses: murder and nonnegligent manslaughter, forcible rape, robbery, and aggravated assault. The temporal stability of model parameters that we found using violent crime rate in this study needs to be substantiated in the case of property crime or other disaggregated crime types. Second, the census residential population is used as the population at risk when calculating crime rates. Although this is a common practice, recent studies also suggest alternative measures of the population at risk such as the ambient population (Andresen, 2006a). It was found that the use of ambient population has an impact on the results of crime analysis (Andresen, 2011), although Harries (1991) suggested that this improvement is most often not worth the costs in both time and money. Third, in terms of the choice of spatial unit, although some found that it has no impact on the result of crime analysis, see for example, Land et al. (1990), Ouimet (2000), Wooldredge (2002), and no impact on “the power of the statistical tests performed” (Ouimet, 2000, p. 150), others suggested the use of smaller spatial units (Andresen, 2011). The use of small spatial unit provides stronger coefficients than the larger spatial unit (Ouimet, 2000). The BG is the finest scale for which the U.S. Census Bureau tabulates and publishes sample data. The use of other larger units such as census tract in this study may results in weaker coefficients.

2.6 Conclusions

This study yields several relevant findings. We first present the importance of dependent variable selection, and question the appropriateness of averaging crime rates over multiple years. Although it is a common practice to use either single-year or multi-year average of crime, to our knowledge, evidence to support either practice seems to be absent, as per our comprehensive literature review. A common approach for evaluating the annual fluctuation in crime data is conducting difference of means test for crime data, and averaging crime data across years if fluctuation is observed. However, socio-demographic and economic variables may be able to account for certain degree of fluctuation. By applying a unified modeling framework to assess the stability of model parameters in crime rate analysis, this study adds to existing literature by empirically studying the appropriateness of using single-year and multi-year average measure. First, a difference of means test reveals three significant fluctuation in crime data, that is, crime rates in 1998 (lower), 1999 (lower) and 2001 (higher) are significantly different from the average rate. However, keeping in mind that social indicators may explain and moderate this variation, we present that the criteria to support the averaging of crime data across years are not only the presence of year-by-year volatility, but also the temporal stability of relationships between socio-economic characteristics and crime. To verify this, seemingly unrelated regression is considered a superior approach.

Consequently, by employing comprehensive social disorganization variables, the significant determinants of violent crime rate in Columbus are provided. Based on the estimated parameters of significant variables, the Wald statistic test suggests strong parameter stability between five single years (1998, 1999, 2000, 2001 and 2002) and the average of five years. Although the difference of means test reveals several fluctuations, a more comprehensive regression and model parameters test indicates an inverse result. This implies the importance of taking social indicators into consideration when selecting the form of the dependent variable (i.e. average or single year).

Finally, we conclude that temporal stability refers to not only the year-by-year volatility in crime data, but more importantly, the temporal stability of relationships between socio-economic characteristics and crime. If crime processes are stable over a period of time, and the effects of socio-demographic and economic determinants on crime of different years are significantly stable, it is unnecessary to average crime over multiple years.

Although not the main focus for this research, we also find that the social disorganization theory performs well in explaining violent crime in Columbus. All four elements of the theory are significant, and all variables show expected signs for their estimated parameters. This finding can also add evidence to current empirical research of the theory.

References

Andresen, M. A. (2006a). Crime measures and the spatial analysis of criminal activity. British Journal of Criminology, 46(2), 258-285.

Andresen, M. A. (2006b). A spatial analysis of crime in Vancouver, British Columbia: a synthesis of social disorganization and routine activity theory. The Canadian Geographer/Le Géographe canadien, 50(4), 487-502.

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

Andresen, M. A., & Linning, S. J. (2012). The (in) appropriateness of aggregating across crime types. Applied Geography, 35(1), 275-282.

Andresen, M. A., & Malleson, N. (2011). Testing the stability of crime patterns: implications for theory and policy. Journal of Research in Crime and Delinquency, 48(1), 58-82.

Andresen, M. A., & Malleson, N. (2013). Spatial Heterogeneity in Crime Analysis. In Crime Modeling and Mapping Using Geospatial Technologies (pp. 3-23).

Netherlands: Springer.

Anselin, L. (2006). Spatial econometrics. In T. C. Mills, & K. Patterson (Eds.), Econometrics theory: Vol. 1. Palgrave handbook of econometrics (pp. 901-941). Basingstoke: Palgrave Macmillan.

Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: an introduction to spatial data analysis. Geographical Analysis, 38(1), 5-22.

Bernasco, W., Block, R., & Ruiter, S. (2013). Go where the money is: modeling street robbers' location choices. Journal of Economic Geography, 13(1), 119-143.

Bernasco, W., Elffers, H. (2010). Statistical analysis of spatial crime data (chapter 33). In: A. Piquero, & D. Weisburd (Eds.). Handbook of quantitative criminology (pp. 699-724). Springer.

Bjerk, D. (2010). Thieves, thugs, and neighborhood poverty. Journal of Urban Economics, 68(3), 231-246

Blau, P. M. (1977). Inequality and heterogeneity: a primitive theory of social structure. New York: Free Press.

Bowers, K. J., & Johnson, S. D. (2003). Measuring the geographical displacement and diffusion of benefit effects of crime prevention activity. Journal of Quantitative Criminology, 19(3), 275-301.

Browning, C. R., Byron, R. A., Calder, C. A., Krivo, L. J., Kwan, M. P., Lee, J. Y., & Peterson, R. D. (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.

Bursik, R. J. (1988). Social disorganization and theories of crime and delinquency: problems and prospects. Criminology, 26(4), 519-552.

Cahill, M. E. (2004). Geographies of urban crime: an intraurban study of crime in Nashville, TN, Portland, OR and Tucson, AZ (Doctoral dissertation). University of Arizona.

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

Cahill, M. E., & Mulligan, G. F. (2007). Using geographically weighted regression to explore local crime patterns. Social Science Computer Review, 25(2), 174-193.

Chamlin, M. B., & Cochran, J. K. (2004). Excursus on the population size-crime relationship, Western Criminology Review, 5, 119.

Charron, M. (2009). Neighbourhood characteristics and the distribution of police-reported crime in the city of Toronto. Statistics Canada, Canadian Centre for Justice Statistics.

Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. American Journal of Sociology, 1261-1293.

Cohen, J., & Tita, G. (1999). Diffusion in homicide: exploring a general method for detecting spatial diffusion processes. Journal of Quantitative Criminology, 15(4), 451-493.

Davidson, R., & MacKinnon, J. G. (1993). Estimation and inference in econometrics. Cambridge, Massachusetts: Harvard University Press.

Doreian, P. (1980). Linear models with spatially distributed data spatial disturbances or spatial effects? Sociological Methods & Research, 9(1), 29-60.

Ellis, L., & Walsh, A. (2000). Criminology: a global perspective. Boston, MA: Allyn and Bacon.

Erdogan, S., Yalçin, M., & Dereli, M. A. (2013). Exploratory spatial analysis of crimes against property in Turkey. Crime, Law and Social Change, 59(1), 63-78.

ESRI (Environmental Science Research Institute). (2012). ArcGIS 10.1. Redlands, CA, USA.

Griffith, D. A. (2003). Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Berlin: Springer-Verlag.

Griffith, D. A., & Amrhein, C. G. (1991). Statistical analysis for geographers. Englewood Cliffs, NJ: Prentice Hall.

Harries, K. D. (1991). Alternative denominators in conventional crime rates. In: P. Brantingham & P. Brantingham (Eds.), Environmental criminology (2nd ed.). (pp. 147– 165). London: Sage.

Jacob, J. C. (2006). Male and female youth crime in Canadian communities: assessing the applicability of social disorganization theory. Canadian Journal of Criminology and Criminal Justice, 48(1), 31-60.

Krivo, L. J., & Peterson, R. D. (1996). Extremely disadvantaged neighborhoods and urban crime. Social Forces, 75(2), 619-648.

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

Langford, M., & Unwin, D. J. (1994). Generating and mapping population density surfaces within a geographical information system. The Cartographic Journal, 31(1), 21-26.

Law, J., & Quick, M. (2013). Exploring links between juvenile offenders and social disorganization at a large map scale: a Bayesian spatial modeling approach. Journal of Geographical Systems, 15(1), 89-113.

Light, M. T., & Harris, C. T. (2012). Race, space, and violence: exploring spatial dependence in structural covariates of white and black violent crime in US counties. Journal of Quantitative Criminology, 28(4), 559-586.

McCord, E. S., & Ratcliffe, J. H. (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.

Messner, S. F., & Anselin, L. (2004). Spatial analyses of homicide with areal data. In Spatially integrated social science (pp. 127-144).

Messner, S. F., Anselin, L., Baller, R. D., Hawkins, D. F., Deane, G., & Tolnay, S. E. (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., & Golden, R. M. (1992). Racial inequality and racially disaggregated homicide rates: an assessment of alternative theoretical explanations. Criminology, 30(3), 421-448.

Minnesota Population Center (2011). National Historical Geographic Information System: Version 2.0. Minneapolis, MN: University of Minnesota. https://www.nhgis.org/

Moran, P. A. (1950). Notes on continuous stochastic phenomena. Biometrika, 17-23.

Morenoff, J. D., & Sampson, R. J. (1997). Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970–1990. Social Forces, 76(1), 31-64.

Morenoff, J. D., Sampson, R. J., & Raudenbush, S. W. (2001). Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. Criminology, 39(3), 517-558.

Novak, K. J., Hartman, J. L., Holsinger, A. M., & Turner, M. G. (1999). The effects of aggressive policing of disorder on serious crime. Policing: An International Journal of Police Strategies & Management, 22(2), 171-194.

Oberwittler, D., & Wikström, P. O. H. (2009). Why small is better: Advancing the study of the role of behavioral contexts in crime causation. In Putting crime in its place (pp. 35-59). New York: Springer.

Ouimet, M. (2000). Aggregation bias in ecological research: how social disorganization and criminal opportunities shape the spatial distribution of juvenile delinquency in Montreal. Canadian Review of Criminology, 42, 135-156.

Páez, A., & Scott, D. M. (2004). Spatial statistics for urban analysis: a review of techniques with examples. GeoJournal, 61(1), 53-67.

Park, R. E., Burgess, E. W., & McKenzie, R. D. (1967). The city. Chicago: University of Chicago Press.

Peterson, R. D., Krivo, L. J., & Harris, M. A. (2000). Disadvantage and neighborhood violent crime: Do local institutions matter?. Journal of Research in Crime and Delinquency, 37(1), 31-63.

Porter, J. R., & Purser, C. W. (2010). Social disorganization, marriage, and reported crime: a spatial econometrics examination of family formation and criminal offending. Journal of Criminal Justice, 38(5), 942-950.

Rangel, T. F., Diniz-Filho, J. A. F., & Bini, L. M. (2010). SAM: a comprehensive application for spatial analysis in macroecology. Ecography, 33(1), 46-50.

Ratcliffe, J. H. (2004). Geocoding crime and a first estimate of a minimum acceptable hit rate. International Journal of Geographical Information Science, 18(1), 61-72.

Ratcliffe, J. (2010). Crime mapping: spatial and temporal challenges. In: Piquero A.R., & D. Weisburd (Eds.), Handbook of quantitative criminology (pp. 5-24). New York: Springer.

Rosenfeld, R., & Fornango, R. (2007). The impact of economic conditions on robbery and property crime: the role of consumer sentiment. Criminology, 45(4), 735-769. Chicago.

Sampson, R. J. (1985). Race and criminal violence: a demographically disaggregated analysis of urban homicide. Crime & Delinquency, 31(1), 47-82.

Sampson, R. J. (1987). Urban black violence: the effect of male joblessness and family disruption. American Journal of Sociology, 348-382.

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

Schweitzer, J. H., Kim, J. W., & Mackin, J. R. (1999). The impact of the built environment on crime and fear of crime in urban neighborhoods. Journal of Urban Technology, 6(3), 59-73.

Shaw, C. R., and McKay, H. D. (1942). Juvenile delinquency and urban areas. Chicago: University of Chicago Press.

Smith, W. R., Frazee, S. G., & Davison, E. L. (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.

StataCorp. (2011). Stata statistical software: release 12. College Station, TX: Stata Corp LP.

Sun, I. Y., Triplett, R., & Gainey, R. R. (2004). Neighborhood characteristics and crime: a test of Sampson and Groves' model of social disorganization. Western Criminology Review, 5, 1.

Tita, G. E., & Radil, S. M. (2010). Spatial regression models in criminology: Modeling social processes in the spatial weights matrix (chapter 6). In: A. R.

Piquero, & D. Weisburd (Eds.) Handbook of quantitative criminology (pp. 101-121). New York: Springer.

Townsley, M. (2009). Spatial autocorrelation and impacts on criminology. Geographical Analysis, 41(4), 452-461.

U.S. Census Bureau (2003). Census of population and housing, summary file 4: technical documentation (p. 2000).

Weisburd, D. L., Groff, E. R., & Yang, S. M. (2012). The criminology of place: street segments and our understanding of the crime problem. Oxford University Press.

Wooldredge, J. (2002). Examining the (ir)relevance of aggregation bias for multi-level studies of neighborhoods and crime with an example comparing census tracts to official neighborhoods in Cincinnati. Criminology, 40(3), 681-710.

Ye, X., & Wu, L. (2011). Analyzing the dynamics of homicide patterns in Chicago: ESDA and spatial panel approaches. Applied Geography, 31(2), 800-807.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association, 57(298), 348-368.

Appendix

Table 2.6 Correlations for independent variables

X_1	1																		
X_2	0.731	1																	
X_3	0.671	0.689	1																
X_4	0.723	0.678	0.731	1															
X_5	0.730	0.686	0.762	0.769	1														
X_6	0.777	0.798	0.828	0.796	0.845	1													
X_7	0.318	0.344	0.353	0.350	0.343	0.370	1												
X_8	0.373	0.338	0.312	0.328	0.372	0.366	0.379	1											
X_9	0.163	0.205	0.212	0.139	0.192	0.185	0.595	0.269	1										
X_{10}	-0.035	0.015	0.061	-0.012	0.017	0.035	0.484	-0.043	0.531	1									
X_{11}	0.237	0.222	0.132	0.185	0.200	0.161	-0.117	0.130	-0.108	-0.371	1								
X_{12}	-0.105	-0.077	-0.059	-0.170	-0.122	-0.120	0.152	-0.059	0.370	0.171	-0.062	1							
X_{13}	0.198	0.214	0.280	0.159	0.204	0.211	0.183	0.177	0.356	0.022	0.162	0.376	1						
X_{14}	0.443	0.447	0.412	0.449	0.449	0.472	0.702	0.335	0.467	0.469	-0.223	0.046	0.035	1					
X_{15}	0.545	0.548	0.512	0.571	0.582	0.576	0.434	0.422	0.242	-0.011	0.103	-0.169	0.147	0.590	1				
X_{16}	0.245	0.290	0.288	0.216	0.256	0.270	0.501	0.210	0.619	0.543	-0.302	0.233	0.147	0.686	0.344	1			
X_{17}	-0.036	0.044	0.075	-0.024	-0.006	0.053	0.442	-0.010	0.473	0.797	-0.442	0.189	0.000	0.534	-0.037	0.677	1		
X_{18}	0.322	0.350	0.322	0.332	0.356	0.331	-0.036	0.188	-0.114	-0.351	0.359	-0.102	0.219	0.020	0.395	-0.065	-0.416	1	
X_{19}	-0.190	-0.165	-0.168	-0.262	-0.192	-0.229	0.039	-0.153	0.246	0.234	-0.197	0.654	0.212	0.078	-0.213	0.371	0.264	-0.129	1
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}

Note: $X_{1:}$ spatially lagged violent crime rate per 1,000 in 1998; $X_{2:}$ spatially lagged violent crime rate per 1,000 in 1999; $X_{3:}$ spatially lagged violent crime rate per 1,000 in 2000; $X_{4:}$ spatially lagged violent crime rate per 1,000 in 2001; $X_{5:}$ spatially lagged violent crime rate per 1,000 in 2002; $X_{6:}$ spatially lagged five-year average crime rate per 1,000; $X_{7:}$ poverty population (%); $X_{8:}$ unemployment rate (%); $X_{9:}$ renter occupied housing units (%); $X_{10:}$ population living in households and are nonrelatives (%); $X_{11:}$ single parent families; $X_{12:}$ foreign born population (%); $X_{13:}$ Heterogeneity Index; $X_{14:}$ spatially lagged X_7; $X_{15:}$ spatially lagged X_8; $X_{16:}$ spatially lagged X_9; $X_{17:}$ spatially lagged X_{10}; $X_{18:}$ spatially lagged X_{11}; $X_{19:}$ spatially lagged X_{12}.

Chapter 3 Persistence of violent crime hot spots

3.1 Introduction

Crime events are not evenly distributed across space and time (Eck and Weisburd, 1995). Spatial concentration of crime events forms so called crime “clusters” or “hot spots”. These are arguably the most fundamental indicators that summarize the unevenness in the pattern of crime occurrences and thus help to identify high risk places that require intervention (Shiode, 2011). Hot spots are deemed to have a certain degree of temporal persistence. According to Sherman (1995, p. 36), for instance, “hot spots are small places in which the occurrence of crime is so frequent that it is highly predictable, at least over a one year period”. However, the literature regarding the persistence of hot spots has focused on relatively short time periods (e.g., days, weeks, or seasons). Studies on short time periods potentially suffer from high variance in risk estimates (Mohler, 2014) and fail to adequately address the mechanisms feeding and breaking longer-term persistence. To date, little is known about the socio-economic and environmental correlates of hot spot persistence. The objective of this study therefore is to illustrate a modeling approach useful to examine the persistence of crime hot spots over longer time periods. This study addresses a gap in the literature by identifying the neighborhood structural correlates that perpetuate or interrupt the presence of crime hot spots. Specifically, the study is motivated by the following considerations.

First, scholars have acknowledged the significance of characterizing the persistence of crime hot spots for crime prevention and reduction (e.g., Ratcliffe, 2004; Johnson, Lab, and Bowers, 2008; Law, Quick, and Chan, 2015), however, there is a lack of literature in the study of long-term persistence (e.g., Weisburd et al., 2004). Contemporary research efforts have been devoted to short-term persistence, such as days (Ratcliffe, 2002), weeks (Johnson, Lab, and Bowers, 2008), months (Ratcliffe, 2004), and seasons (Andresen and Malleson, 2013). Compared to short-term hot spots which usually experience high level of volatility and suffer from high variance in risk estimates (Weisburd et al., 2004; Mohler, 2014), the existence of long-term hot spots can be an important indicator of consistent problem areas and socio-economic and environmental deprived neighborhoods (Ratcliffe and McCullagh, 1999). There is thus a need for a clear understanding of the long-term persistence of hot spots.

Second, the mechanisms feeding and breaking the hot spot persistence remain under-investigated. A systematic understanding of the mechanisms is valuable for two reasons. Firstly, numerous studies, especially those on crime prevention through environment design (CPTED), have suggested that focused crime prevention strategies on persistent hot spots can effectively reduce crime (Cozens, Saville, and Hillier 2005). Secondly, researchers have observed that some beneficial effects of crime reduction strategies diffuse from focused hot spots into adjacent neighborhoods immediately surrounding the focused ones (Braga, Papachristos, and Hureau, 2014). All in all, it is ultimately an important premise to identify the socio-economic and environmental correlates underlying the persistence of hot spots to provide insights to inform these efforts and strategies. Furthermore, it is important to examine whether our conventional understanding regarding the neighborhood structural determinants of cross-sectional crime can also account for various degrees of hot spot persistence.

Our case study is violent crime in Columbus, Ohio. Using nine years' worth of crime information and two decennial census datasets, we explore the question of why crime hot spots are persistent in certain areas over multiple years. This study contributes to the literature in the following ways. First, the measure of cluster membership (being part of a hot spot) can conveniently capture the magnitude of the hot spot persistence over several years. Second, use of an ordered probit regression provides a methodology to explore the structural determinants of cluster membership. Third, the identified neighborhood characteristics provide insights that can inform crime reduction initiatives with the aim of providing "long-term intervention to reduce the risk factors associated with offending" (Craglia, Haining, and Wiles, 2000, p. 728) – including modifications to the built environment, allocation of police and crime prevention resources, and introduction of community problem-solving techniques (Johnson, Lab, and Bowers, 2008).

3.2 Context and data sources

3.2.1 Crime data

The City of Columbus is located in central Ohio, and had approximately 0.6 million residents in 1990 and 0.7 million residents in 2000. Crime data was collected by the Columbus Division of Police, and covers a wide range of attribute information, such as location of crime incident, date and time when crime happened, date when crime was cleared, Uniform Crime Reporting (UCR) code, and type of crime, among other attributes. According to the FBI's UCR Program, violent crime consists of four criminal offenses: murder and non-negligent manslaughter, forcible rape, robbery, and aggravated assault. We aggregate the four types of crime into yearly cross-sectional violent crime dataset. To be specific, there are 41,827 violent crimes in total over nine years: 4,962 in 1994, 4,978 in 1995, 4,715 in 1996, 4,550 in 1997, 4,049 in 1998, 4,333 in 1999, 4,578 in 2000, 4,871 in 2001, and 4,791 in 2002. Apart from the 5 per cent randomly and erroneously coded addresses by police personnel, the geocoding accuracy of our dataset reaches to 95 per cent success rate which is high enough to minimize the concern of bias in the analysis.

3.2.2 Census data

Considering the long-term nature of crime dataset, from 1994 to 2002, and the US Census is administered every 10 years, the present study uses two decennial Census data to reduce possible bias result from the gap: Census 1990 and Census 2000. The Census dataset is drawn from two sources: the US Census Bureau and the National Historical Geographic Information System (Minnesota Population Center, 2011). The former provides Summary File 3 (SF3) which consists of Census 1990 and 2000 general demographic, social, economic and housing characteristics. Most of our independent variables aggregated at the zonal level by block group are drawn from the SF3. The latter is used to generate supplementary socio-demographic and economic independent variables.

3.2.3 Spatial unit of analysis

The spatial unit used in the study is Census block group (BG), the finest scale for which the U.S. Census Bureau tabulates and publishes sample data. BGs generally contain between 600 and 3000 people and are made up of on average 40 Census blocks. We use two sets of BG level boundaries: BG boundaries of 1990 and 2000. This corresponds to the two decennial Census data mentioned above. The study area in both 1990 and 2000 boundaries does not cover the whole city for three reasons. First, crime data in some suburbs in BGs of Columbus are not maintained by Columbus Division of Police which is our data source. We have to exclude them from the study area. Second, given that two sets of BG boundaries are used, and the boundaries of the city and BGs changed from 1990 to 2000, to maintain the comparability of the two, 7 BGs in 1990 and 11 in 2000 have to be excluded from the study area. Third, Census data is not available for Census units having less than 100 people (U.S. Census Bureau, 2003), therefore, 7 BGs in 1990 and 4 in 2000 have to be excluded. The resulting dataset consists of n=652 BGs.

3.3 Methods

3.3.1 Spatial clustering analysis

The first step in our analysis is to investigate the spatial characteristics of violent crime throughout the study area from 1994 to 2002 by means of Kulldorff's spatial scan statistics (Kulldorff, 2011). Scan statistics have been increasingly used by criminologists for detecting statistically significant spatial, temporal or space-time hot spots in crime data (e.g., Ceccato and Haining, 2004; Nakaya and Yano, 2010). These statistics operate by gradually scanning a circle or an ellipse window across the study area, recording the number of observed and expected observations inside the window at each BG. In this process, different size of the scanning window is used. The window with the maximum likelihood will be the most likely violent crime cluster. Simulation based pseudo-*p*-values are assigned to clusters, with which significant clusters can be identified and spatial patterns of yearly violent crime can be intuitively compared.

For this task we use SaTScan 9.1.1 (Kulldorff, 2011), and spatial scan analysis is performed to detect yearly spatial clusters of violent crime. Three input data are used: case data (count of violent crime in each 1990 BG), population file (population of each BG from Census 1990), and coordinate file (coordinates of the centroids of 1990 BGs). The case data are generated by aggregating violent crime points to the 1990 block group level. Based on these aggregated case data the discrete Poisson model is used as probability model. The maximum spatial cluster size is set as 10% of population at risk implying the reported clusters can include at most 10% of the total population at risk. The value of 10% is selected through an iterative process that generates significant secondary clusters with no common geographic areas with the most likely clusters. The value of 10% is supported by other similar studies such as Uittenbogaard and Ceccato (2012). The Monte Carlo replication is set to 999 times in the model to assess the significance of detected clusters. We then combine the most likely and secondary clusters to generate the cluster membership variables (the frequency that each BG belongs to detected clusters over nine years) for further analysis.

3.3.2 Pycnophylactic interpolation

We take 1990 as the base year for the analysis. Decennial change variables (Δ) are generated by subtracting the Census 1990 variables from Census 2000 variables to capture the amount of variation (increase or decrease) between the two. Given changes in BG boundaries between 1990 and 2000, we use Tobler's (1979) pycnophylactic interpolation method to generate two compatible zoning systems. Pycnophylactic interpolation is the most appropriate approach in this case because of its property of mass preservation (Páez et al., 2013a, 2013b). An ArcGIS extension tool developed by Qiu, Zhang, and Zhou (2012) is used to perform the interpolation for all Census 2000 variables in ArcGIS 10.1. To avoid interpolation bias, we exclude the non-residential land use from the source zone to ensure that the variables can only be redistributed to populated units. An example of the interpolation result is shown in Figure 3.1, in which a BG level variable (poverty rate) from Census 2000 is interpolated into 1990 BG boundaries.

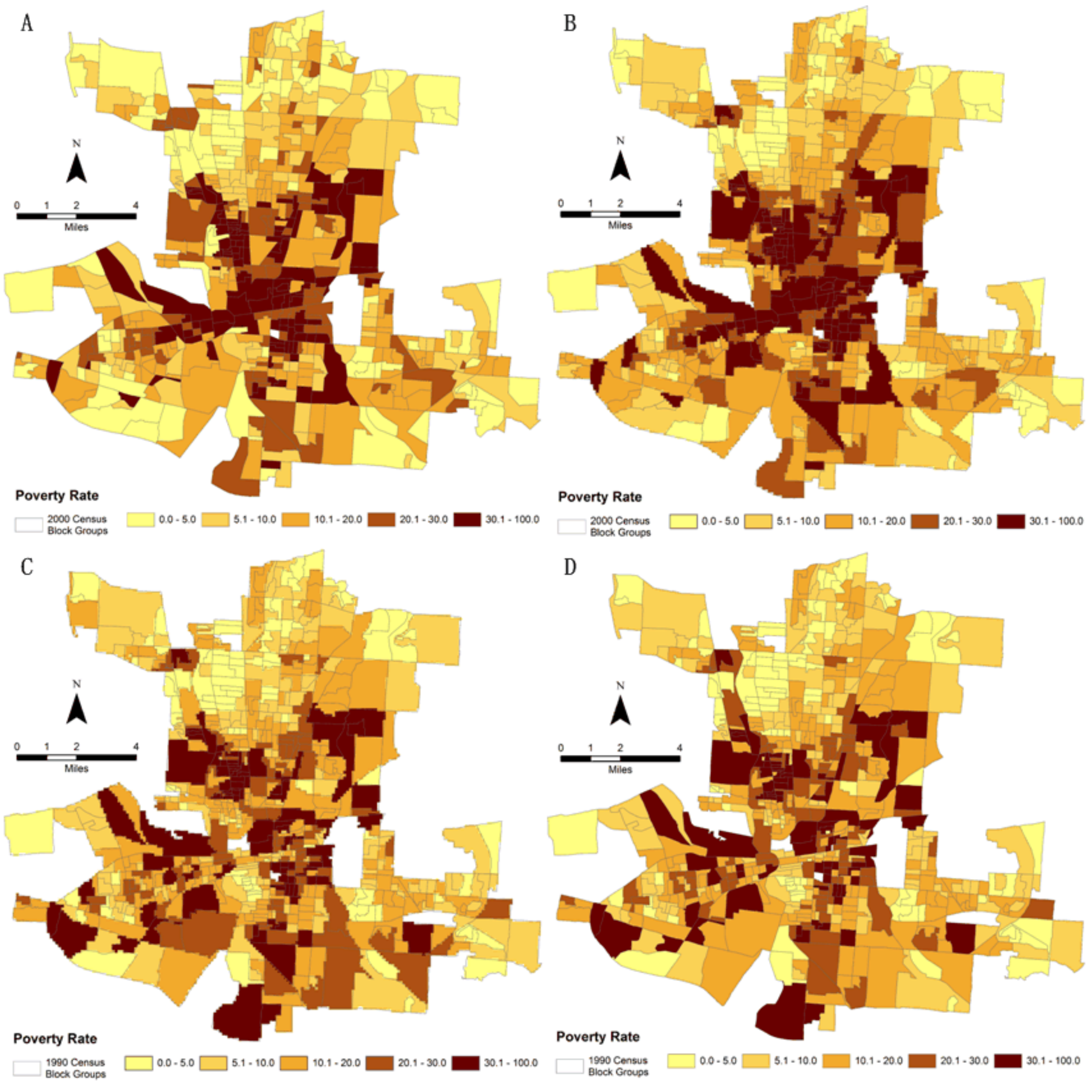

Figure 3.1 Poverty rate from Census 2000.
Map A indicates the spatial pattern of poverty rate in source zone (2000 BG boundaries); map B indicates the interpolated poverty rate (raster surface) in source zone; map C indicates the estimated poverty rate (raster) in target zone (1990 BG boundaries); map D shows the estimated poverty rate (vector) in target zone.

3.3.3 Ordered probit regression

Based on the detected clusters of nine years, the frequency that each BG belongs to clusters over nine years can be counted, and the cluster membership variable can be obtained. The cluster membership variable (y) can be defined as an ordered variable with 10 possible categories, that is:

$$y_{(R\times 1)} = (y_1, \cdots, y_i, \cdots, y_R)', \ \forall i = 1, \cdots, n \tag{13}$$

$$\text{with} \quad y_i = \begin{cases} 0, \ \text{if BG } i \text{ does not belong to any cluster in 9 years} \\ 1, \ if \text{ BG } i \text{ belongs to cluster in 1 year} \\ \vdots \\ j, \ if \text{ BG } i \text{ belongs to clusters in } j \text{ years} \\ \vdots \\ 9, \ \text{if BG } i \text{ belongs to clusters in 9 years} \end{cases}$$

where i represents the BG number, $\forall j = 0, \cdots, m$ ($m = 9$ in this case), R is the total amount of BGs in the study area. Ordinal outcome y is a function of a latent variable y^* that describes the unobserved probability to be included in certain category. The higher the value of y^* in one BG, the more likely the BG is to report a higher category of cluster membership (i.e. the BG belongs to clusters in more years). The latent variable y^* can be defined as follows (see López, Angulo, and Artal, 2012):

$$y_{(R\times 1)} = (y_1^*, \cdots, y_i^*, \cdots, y_R^*)', \ \forall i = 1, \cdots, n \tag{14}$$

$$\text{with} \quad y_i = \begin{cases} 0, \text{ if } & \mu_{-1} < y_i^* \leq \mu_0 \\ 1, \ if & \mu_0 < y_i^* \leq \mu_1 \\ & \vdots \\ j, \text{ if } & \mu_{j-1} < y_i^* \leq \mu_j \\ & \vdots \\ 9, \ if & \mu_8 < y_i^* < \mu_9 \end{cases}$$

where $\mu_0, \cdots, \mu_8$ are unobserved (but estimable) thresholds defined to satisfy the conditions that $\mu_0 < \mu_1 < \cdots < \mu_8$, and $\mu_{-1} = -\infty$, $\mu_9 = +\infty$. In this study there are 10 categories, thus the value range of y^* can be divided into 10 intervals. The threshold values (μ) correspond to the cut-offs where a BG moves from reporting one category of cluster membership to another.

Given the general existence of spatial dependence in crime analysis, the spatially aggregated socio-economic variables of one BG determine not only the probability that the BG itself to be included in a higher level of the category, but also the probabilities of its neighbors. In this context, we take spatial lag of socio-economic variables into consideration, and the model can be defined in a following form (see López, Angulo, and Artal, 2012):

$$y^* = X\beta + \varepsilon, \ \varepsilon \sim N(0, 1) \tag{15}$$

$$\text{with} \quad X_{(R\times(2k+1))} = [\iota \quad Z \quad WZ]; \quad \beta_{((2k+1)\times 1)} = [\beta_0 \quad \beta_Z \quad \beta_{WZ}]'$$

where X denotes an $R \times (2k+1)$ matrix, ι is an $R \times 1$ vector of ones, Z is an $R \times k$ matrix of independent variables, W denotes an $R \times R$ matrix of spatial weights with elements W_{ij} represent the spatial proximity of unit i and j, WZ is an $R \times k$ matrix of spatial lag of independent variables, β is a $(2k+1) \times 1$ vector of the regression coefficients, ε denotes an $R \times 1$ vector of disturbances. The spatial weight matrix is created using GeoDa 1.4.0 (Anselin, Syabri, and Kho, 2006), based on the Queen's 1st order contiguity protocol as suggested by Andresen (2011).

Spatial association is typically a concern with linear regression and a vast literature exists that addresses this topic (Anselin, 1988; Griffith, 1988; Haining, 1993). In contrast, little is known about the issue of spatial association in models with categorical variables, and existing work has been based mostly on modifications of the well-known Moran's coefficient (Amaral, Anselin, and Arribas-Bel, 2013). In the present case, we use the spatial fit indicator developed by Páez et al. (2013) to assess the degree of spatial fit of models with categorical variables. This indicator is used to assess whether the model can replicate the degree of spatial association of the original variable on cluster membership. The spatial fit indicator is based on the $Q(m)$ statistic for the spatial association of categorical variables (for a detailed illustration of the statistic see Ruiz, López, and Páez, 2010, Páez et al., 2012, and Ruiz, López, and Páez, 2012). The spatial fit indicator was developed to compare the similarity level of spatial association between the predicted variable and the observed variable. It is defined in the following form (see Páez et al., 2013):

$$\Phi = \frac{\hat{Q}(m) - Q(m)}{\hat{Q}(m) + Q(m)} \tag{16}$$

where $Q(m)$ denotes the value of the statistic for the original cluster membership variable $y_{(R\times1)}$, $\hat{Q}(m)$ denotes the value of the statistic for the predicted cluster membership variable ($\hat{y}_{(R\times1)}$) obtained from a candidate model. The indicator of spatial fit (Φ) is bounded between -1 and 1, where positive Φ indicates over-fitting of the predictions (more spatial structure than observed), whereas negative Φ indicates under-fitting (less spatial structure than observed). A value of zero, finally, indicates an identical degree of spatial association between the predicted and observed variables. Confidence intervals have been developed for Φ, which can be used to evaluate whether two variables have significantly different levels of spatial association at a predetermined level of confidence.

3.4 Analysis and results

3.4.1 Dependent variables

Figure 3.2 shows the procedure of generating the dependent variable. The top two map layers represent crime clusters in 1994 and 1995 as detected by spatial scan. By overlaying nine cluster maps, the cluster membership can be obtained by counting the number of times a BG belongs to clusters, as shown in the third map layer in Figure 3.2. The summary statistics of the 10 categories of cluster membership is shown in the frequency tables (see Table 3.1). The ordered probit model assumes the same relationship exists between each pair of outcome categories of the dependent variable. That is, the coefficients obtained from the model represent the relationship between the first category versus all higher categories are the same as that represent the relationship between the second category and all other higher categories.

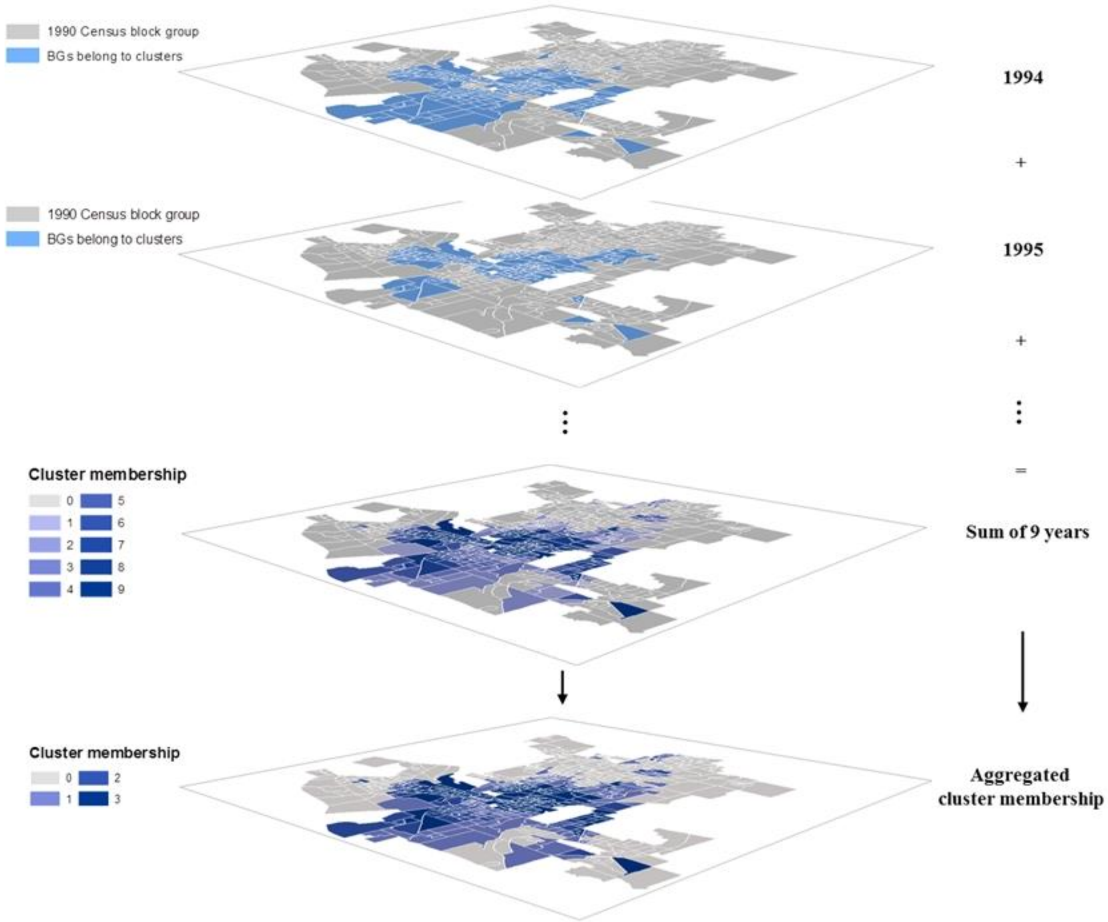

Figure 3.2 Generating the dependent variable

Preliminary model estimates reveal that the intercept parameters are not significantly different from each other, which indicates that the 10 categories need to be combined into fewer categories. To reduce potential bias, we reduce the number of categories by merging the sparsely populated categories (e.g., category 4, 5, 6) to others. 10 categories are re-aggregated in this way into 4, as shown in the fourth map layer in Figure 3.2: category 0 (the zero category) stays the same and denotes that the BG does not belongs to any clusters in 9 years; category 1 (low persistence) indicates that the BG belongs to a cluster between 1 and 3 years; category 2 (medium persistence) means that the BG belongs to a cluster between 4 and 6 years; and category 3 (high persistence) indicates that the BG is part of a cluster for 7 to 9 years. Table 3.1 shows the frequency table of the aggregated cluster membership.

Table 3.1 Aggregation of the frequency of cluster membership

Original categories	**Freq.**	**Percent**	**New categories**	**Freq.**	**Percent**
0	287	44.02	0	287	44.02
1	63	9.66	1	141	21.63
2	42	6.44			
3	36	5.52			
4	27	4.14	2	69	10.58
5	25	3.83			
6	17	2.61			
7	36	5.52	3	155	23.77
8	47	7.21			
9	72	11.04			
Total	652	100.00	Total	652	100.00

Note: When performing the Kulldorff's spatial scan statistics, we use the discrete Poisson model as probability model, we set the maximum spatial cluster size as 10%, we combine the most likely and secondary clusters to generate the cluster membership.

3.4.2 Independent variables

Independent variables are selected based on two spatial theories of crime: social disorganization theory (SDT) (Sampson and Groves, 1989) and routine activity theory (RAT) (Cohen and Felson, 1979). SDT generally explains crimes from the perspective of neighborhood characteristics, while the RAT suggests that crimes occur at the intersection of motivated offenders, suitable targets, and the absence of capable guardians. Although the two theories suggest different emphasis of analysis (the former considers neighborhood explanations, while the latter focuses on the individual criminal victimization), researchers argue that elements of the RAT can also be measured by neighborhood level variables (Andresen, 2006b), and an integration of the two can provide good indicators of various crimes (Andresen, 2006a, 2006b) and "the most robust theoretical explanation for geographical studies of crime" (Cahill, 2004, p. 17).

Informed by an extensive review of the relevant literature (e.g., Kelly, 2000; Cahill and Mulligan, 2003; Andresen, 2006a, b; Law and Quick, 2013; Chun, 2014; He et al., 2015), 10 variables are selected to capture the four elements of SDT: socio-economic deprivation, family disruption, residential mobility, and ethnic heterogeneity. Socio-economic deprivation is captured by percentage of population below poverty line, median household income, and percentage of population with bachelor's degree. The median age of housing units is used to capture the neighborhood deprivation from a built environment perspective (Cahill and Mulligan, 2003; Law and Quick, 2013). Family disruption is captured by number of single-parent families. Residential mobility is captured by percentage of renter-occupied housing units and percentage of population live in the same house as 5 years ago. Ethnic heterogeneity is captured by the Heterogeneity Index (Equation (17)) and immigrant concentration. The Heterogeneity Index represents the degree of racial diversity within a BG (Sampson and Groves, 1989). Immigrant concentration is measured by the percentage of foreign-born population and the percentage of Hispanics and Latinos. Current studies provide controversial relationships between immigrant concentration and violent crime, thus the relationship still needs to be examined.

$$\text{Heterogeneity Index } = 1 - \sum {p_i}^2 \qquad (17)$$

where i denotes the number of racial group, and p_i is the percentage of population within a given group. The index ranges from 0 to 1, where 0 indicates maximum homogeneity, and 1 indicates maximum heterogeneity. All five racial groups are used: White; Black; American Indian, Eskimo, or Aleut; Asian or Pacific Islander; and other races.

Five variables at BG level are selected to capture the three elements of RAT: motivated offender, suitable target, and guardianship, three of which are overlapped with SDT variables. The concentration of motivated offender is measured by percentage of population below the poverty level and median household income. Suitable target is captured by percentage of renter-occupied housing units. Guardianship is captured by percentage of vacant housing units and population density per square mile. Furthermore, six geographical variables are used to capture the overall spatial tendency: the distance from the centroid of BGs to Central Business

District (CBD), X Coordinates (easting) and Y Coordinates (northing) of BGs, squared X and squared Y, and the interaction of X times Y. In addition, the spatial lag of each SDT variables and RAT variables are also included by assuming that the spatially aggregated socio-economic variables of one BG determine not only the probability that the BG itself to be included in a higher level of the category, but also the probabilities of its neighboring BGs. Table 3.2 gives descriptive statistics of all independent variables and the dependent variable.

Poverty rate, median age of housing units, single-parent families, rental residences, Heterogeneity Index and vacant housing units have the expectation of positive relationships with the persistence of hot spots. While educated population, median household income, percentage of population live in the same house as 5 years ago, immigrant concentration, population density and distance to CBD have the expectation of negative relationships.

The correlation test for all variables shows that three pair-wise correlations exceed 0.8 or –0.8, a common threshold for collinearity test as suggested by Andresen (2006a): X Coordinates have a strong positive correlation (0.8635) with squared X, Y Coordinates have a strong positive correlation (0.8761) with squared Y, and poverty rate from Census 1990 has a strong negative relationship (-0.828) with decennial change of poverty rate (Δ_{00-90}). This is not surprising, since these three variables (Δ_{00-90}, squared X, and squared Y) are calculated by using the former three variables (poverty rate, X Coordinates, and Y Coordinates), and warrants for careful treatment in model estimation. In addition, population with bachelor's degree has a relatively strong correlation (0.794) with the spatial lag of itself, and population density has a relatively strong relationship (0.770) with the spatial lag of itself. Both are standard relationships and carefully treated in later model estimation.

Table 3.2 Descriptive statistics for variables used in the analysis

Variables	**Mean**	**S.D.**	**Min**	**Max**
Dependent Variable				
Cluster membership	1.14	1.22	0.00	3.00
Base variables from Census 1990				
% of population below poverty level	0.21	0.19	0.00	1.00
Median household income (in thousands)	24.53	11.32	0.00	76.45
% of educated population	0.13	0.12	0.00	0.67
Heterogeneity Index	0.22	0.16	0.00	0.61
% of foreign-born population	0.03	0.05	0.00	0.77
% of Hispanics and Latinos	0.01	0.02	0.00	0.16
# of single-parent families	36.90	45.71	0.00	501.00
% of renter-occupied housing units	0.51	0.30	0.00	1.00
% of population living in the same house as 5 years ago	0.47	0.19	0.00	1.00
% of vacant housing units	0.08	0.07	0.00	0.38
population density (in thousands)	2.93	2.16	0.00	16.68
Interpolated variables from Census 2000				
% of population below poverty level	0.16	0.14	0.01	1.00
Median household income (in thousands)	30.29	20.44	1.80	193.33
% of educated population	0.14	0.13	0.00	1.00
Heterogeneity Index	0.40	0.64	0.00	0.99
% of foreign-born population	0.05	0.07	0.00	0.82
% of Hispanics and Latinos	0.02	0.02	0.00	0.37
# of single-parent families	46.37	48.46	0.14	511.02
% of renter-occupied housing units	0.42	0.25	0.04	0.99

% of population living in the same house as 5 years ago	0.40	0.25	0.01	0.99
% of vacant housing units	0.07	0.06	0.00	0.45
population density (in thousands)	2.37	1.78	0.07	17.19
Spatially lagged base variables				
W * % of population below poverty level	0.21	0.16	0.00	0.84
W * Median household income (in thousands)	24.36	9.21	0.00	60.92
W * % of educated population	0.13	0.10	0.00	0.46
W * Heterogeneity Index	0.22	0.12	0.00	0.55
W * % of foreign-born population	0.03	0.03	0.00	0.21
W * % of Hispanics and Latinos	0.01	0.01	0.00	0.05
W * % of single-parent families	37.63	30.97	0.00	191.33
W * % of renter-occupied housing units	0.51	0.22	0.00	1.00
W * % of population living in the same house as 5 years ago	0.47	0.15	0.04	0.85
W * % of vacant housing units	0.07	0.05	0.00	0.35
W * population density (in thousands)	2.77	1.65	0.11	11.15
Decennial change variables				
Δ_{00-90}† of population below poverty level	-0.04	0.17	-0.87	0.92
Δ_{00-90} of median household income (in thousands)	5.76	16.21	-37.49	151.51
Δ_{00-90} of educated population (log)	-3.37	1.27	-10.30	0.01
Δ_{00-90} of Heterogeneity Index (log)	-1.32	1.45	-11.51	-0.17
Δ_{00-90} of foreign-born population	0.02	0.07	-0.26	0.60
Δ_{00-90} of Hispanics and Latinos	0.01	0.03	-0.15	0.33
Δ_{00-90} of single-parent families	9.47	34.46	-321.06	208.75
Δ_{00-90} of renter-occupied housing units	-0.09	0.27	-0.94	1.00
Δ_{00-90} of vacant housing units	-0.01	0.07	-0.30	0.39
Δ_{00-90} of population living in the same house as 5 years ago	-0.07	0.21	-0.96	0.72
Δ_{00-90} of population density (in thousands)	-0.57	1.95	-9.54	14.60
Other independent variables				
Median age of housing units	34.71	13.70	0.00	51.00
Distance to CBD (in thousands)	7.06	3.84	0.26	16.86
X Coordinates‡	3.67	1.45	0.00	8.23
Y Coordinates	3.52	1.72	0.00	7.54

Squared X	15.57	11.88	0.00	67.77
Squared Y	15.33	13.68	0.00	56.89
X times Y	12.61	7.14	0.00	39.99

Note: Spatial units are Columbus Census block groups (N = 652), mapped in Figure 1-3.

† Δ_{00-90} indicates the decennial change of the corresponding variable calculated by subtracting its Census 1990 value from Census 2000 value.

‡ X Coordinates are calculated by subtracting the minimum value of X and then divided by 1,000. Y Coordinates are calculated in the same way.

3.4.3 Results

The results of applying the spatial scan show that violent crimes generally show significant concentration patterns in Columbus. Given that detected secondary clusters generally have no common geographic area with the most likely clusters, according to Chen et al. (2008), they are of interest and able to reject the null hypothesis on their own strength. Figure 3.3 provides the detected significant clusters by combining the most likely and secondary clusters. There are 246 BGs (out of 652) in total belong to clusters in 1994, 190 in 1995, 239 in 1996, 195 in 1997, 173 in 1998, 203 in 1999, 207 in 2000, 208 in 2001, and 214 in 2002. An apparent yearly fluctuation can be observed from 1994 to 1998 in terms of the amount of clustering BGs. The amount hits a low in 1998, then starts level off in 1999 and gradually increases until 2002.

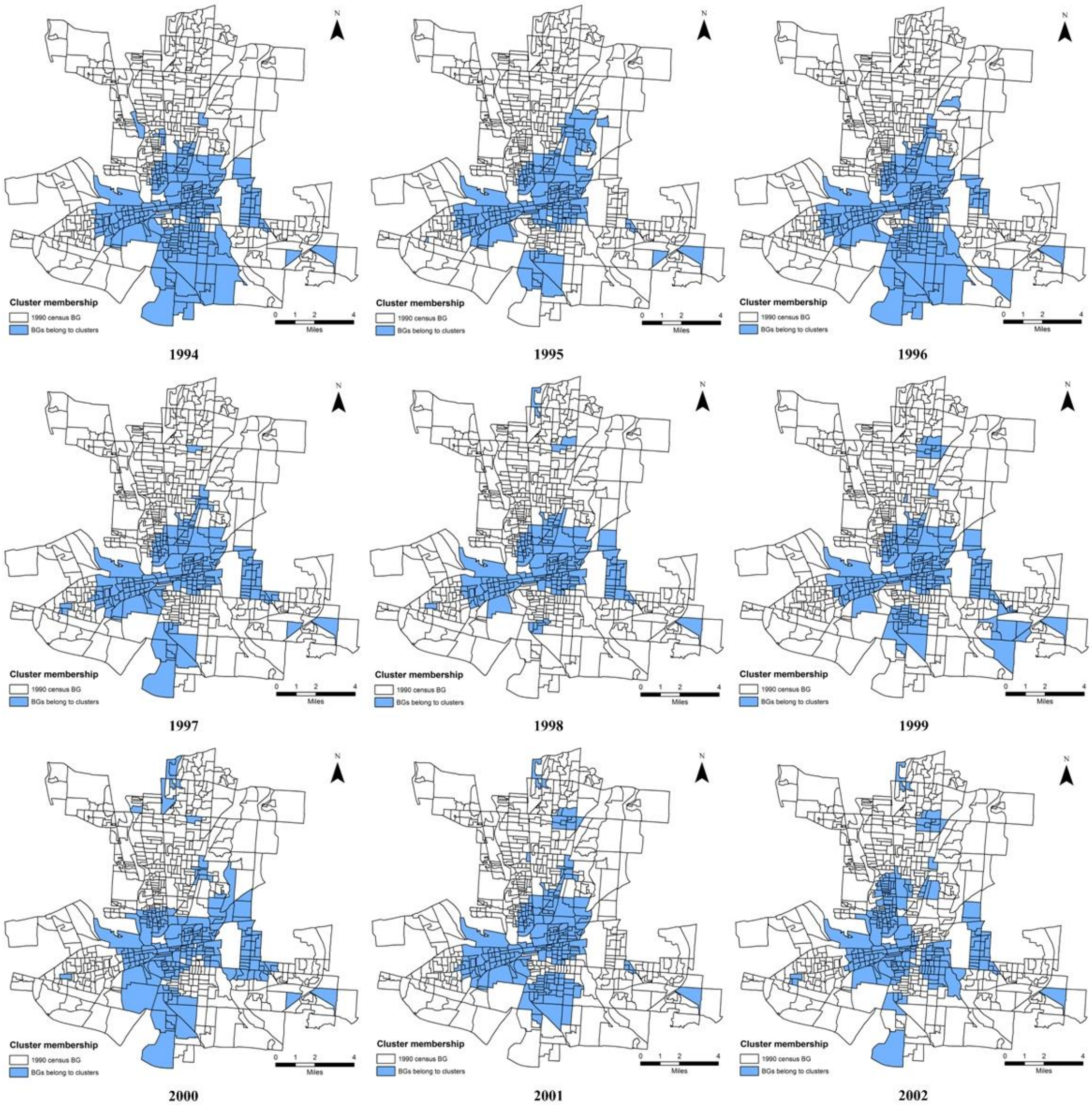

Figure 3.3 Significant violent crime clusters from 1994 to 2002 (a combination of the most likely and secondary clusters)

Regarding the spatial patterns, BGs that belong to clusters are those found in downtown areas as well as northern, eastern and western immediate surrounding areas, among which the sizes and patterns of eastern immediate surrounding BGs are constantly stable in nine years while the northern BGs are stable in the first eight years. Southwards, we also observe constant clusters, but their size substantially varies over nine years, with the largest in 1994 and the smallest in 1998. Northwards, a small number of clusters emerge in 1997, the size and location of which varies slightly year by year. The pattern of clusters in 2002 is notably different from other eight years in a way. Stable clusters immediately north of the CBD break down, and lead to an isolated cluster in western surrounding areas of the downtown, while new clusters emerge in more northern areas. Meanwhile, southern clusters break down into two separate clusters. All these finally result in a discrete pattern of crime clusters in 2002. Moreover, except for a few consistent clusters in the southwestern edges of the city, the northern, western, and eastern edges of the city rarely experience crime clusters.

Of particular interest are the correlates of hot spot persistence during the period. To examine this, the cluster membership variable is regressed against the full set of SDT and RAT variables by using an ordered probit model. The results reported in Table 3.3 show a goodness of fit of McFadden's Pseudo R^2 of 0.449. The likelihood ratio χ^2 of 743.96 with a *p*-value of 0.000 indicates that the model as a whole is statistically significant. The thresholds do not have substantive interpretations (Whalen et al., 2012).

A backward stepwise approach is used to select variables. The final model consists of 15 variables that are statistically significant at least at the 0.05 level of significance and have expected signs. Of these, five are base variables, three are spatially lagged base variables, two capture decennial change, four are geographical variables capturing overall spatial tendency, and one is the built environment variable capturing the neighborhood deprivation. To be specific, the base variable of median household income from Census 1990 is significant and negative; however, the decennial change variable is not. This indicates that BGs with higher median household income in 1990 are less likely to be part of persistent hot spots. However, decennial changes in median household income do not seem to affect this. Both the base and decennial change variable of educated population are significant and negative, which implies that higher starting levels of educated population reduce the probability of being part of a persistent cluster. Moreover, an increase in educated population between the two Census years further reduces the probability of a BG being in a persistent cluster. The base variable of population living in the same house as five years ago is not significant, but the decennial change variable is significantly negative. This indicates that a decennial increase in the population living in the same house as five years ago reduces the probability of a BG being in a persistent cluster. The base variable of single-parent families is significantly positive, which indicates that BGs with single-parent families in 1990 are more likely to be part of persistent hot spots. The base variable of foreign-born population is not significant, but the spatially lagged foreign-born population is significantly negative. This implies that BGs with neighbors with more immigrant concentration in 1990 are less likely to be part of persistent clusters. In contrast, the spatially lagged poverty rate is significant and positive, which implies that BGs with neighbors that experience high poverty rates are more likely to be part of persistent hot spots. Both the base variable and spatially lagged renter-occupied housing units are significant and positive, the base variable of population density is negative. This indicates that higher starting levels of renter-occupied housing units increase the probability of being part of a persistent cluster, but higher levels of population density in 1990 decrease the probability. The variable of median age of housing unit is significant and positive, which suggests that BGs with more aged housing units are more likely to be part of persistent clusters. Regarding the geographical variables, distance to CBD is significantly negative, X Coordinates is significantly positive, Y Coordinates is significantly negative. This implies that violent crime hot spots are generally more persistent in BGs close and to the east and south of the CBD areas.

Table 3.3 Parameter estimates from the ordered probit regression (N = 652 observations)

Variables	**Coefficient**	***p*-value**
Base independent variables		
% of population below poverty level	-	-
Median household income (in thousands)	-0.027	0.000
% of educated population (log)	-0.088	0.001
Heterogeneity Index	-	-
% of foreign-born population	-	-
% of Hispanics and Latinos	-	-
# of single-parent families	0.003	0.009
% of renter-occupied housing units (log)	0.103	0.018
% of population living in the same house as 5 years ago	-	-
% of vacant housing units	-	-
Population density (in thousands)	-0.159	0.000
Spatially lagged base variables		

W * % of population below poverty level	2.459	0.000
W * Median household income (in thousands)	-	-
W * % of educated population	-	-
W * Heterogeneity Index	-	-
W * % of foreign-born population	-6.809	0.004
W * % of Hispanics and Latinos	-	-
W * % of single-parent families	-	-
W * % of renter-occupied housing units	1.390	0.000
W * % of population living in the same house as 5 years ago	-	-
W * % of vacant housing units	-	-
W * Population density (in thousands)	-	-
Decennial change variables		
Δ_{00-90} of population below poverty level	-	-
Δ_{00-90} of median household income (in thousands)	-	-
Δ_{00-90} of educated population (log)	-0.152	0.041
Δ_{00-90} of Heterogeneity Index (log)	-	-
Δ_{00-90} of foreign-born population	-	-
Δ_{00-90} of Hispanics and Latinos	-	-
Δ_{00-90} of single-parent families	-	-
Δ_{00-90} of renter-occupied housing units	-	-
Δ_{00-90} of population living in the same house as 5 years ago	-0.008	0.005
Δ_{00-90} of vacant housing units	-	-
Δ_{00-90} of population density (in thousands)	-	-
Other independent variables		
Median age of housing units	0.023	0.000
Distance to CBD (in thousands)	-0.269	0.000
X Coordinates	0.468	0.000
Y Coordinates	-0.915	0.000
Squared X	-	-
Squared Y	0.133	0.000
X times Y	-	-
μ_1	-1.442	
μ_2	-0.134	
μ_3	0.564	

Log likelihood	-457.078	
McFadden’s Pseudo R^2	0.449	
LR $\chi 2$ (15)	743.960	0.000
Spatial fit indicator Φ	-0.0356	
Interval of Φ	[-0.1146, 0.0158]	

Marginal effects from the ordered probit model are reported in Table 3.4. For all categories, the 15 variables are statistically significant at least at the 0.05 level of significance and have expected signs (except for the Δ_{00-90} of educated population in category 1 which has a *p*-value = 0.068). Interpretation of the marginal effects is as the rate of change in probability given a unit change in a variable. For instance, consider the effect of educated population and renter-occupied housing unit: a one-unit increase in educated population is associated with being 3.3% more likely to be in the category 0 (zero persistence), 1.0% less likely to be in the category 1 (low persistence), 1.3% less likely to be in the category 2 (medium persistence), and 0.9% less likely to be in the category 3 (high persistence); one unit increase in renter-occupied housing units is associated with being 3.8% less likely to be in the zero persistence category, 1.1% more likely to be in the low persistence category, 1.6% more likely to be in the medium persistence category, and 1.1% more likely to be in the high persistence category.

Table 3.4 Ordered probit marginal effects for 4 categories of cluster membership (N = 652 observations)

Variables	Category 0	Category 1	Category 2	Category 3
	dy/dx	dy/dx	dy/dx	dy/dx
Base independent variables				
% of population below poverty level	-	-	-	-
Median household income (in thousands)	0.010	-0.003	-0.004	-0.003
	(0.001)	(0.011)	(0.001)	(0.001)
% of educated population (log)	0.033	-0.010	-0.013	-0.009
	(0.001)	(0.007)	(0.002)	(0.003)
Heterogeneity Index	-	-	-	-
% of foreign-born population	-	-	-	-
% of Hispanics and Latinos	-	-	-	-
# of single-parent families	-0.0013	0.0004	0.0005	0.0004
	(0.009)	(0.025)	(0.012)	(0.014)
% of renter-occupied housing units (log)	-0.038	0.011	0.016	0.011
	(0.018)	(0.037)	(0.021)	(0.024)
% of population living in the same house as 5 years ago	-	-	-	-
% of vacant housing units	-	-	-	-
population density (in thousands)	0.059	-0.017	-0.024	-0.017
	(0.000)	(0.000)	(0.000)	(0.000)
Spatially lagged base variables				
W * % of population below poverty level	-0.910	0.267	0.378	0.265
	(0.000)	(0.001)	(0.000)	(0.001)
W * Median household income (in thousands)	-	-	-	-
W * % of educated population	-	-	-	-
W * Heterogeneity Index	-	-	-	-
W * % of foreign-born population	2.520	-0.740	-1.046	-0.734
	(0.004)	(0.016)	(0.006)	(0.007)
W * % of Hispanics and Latinos	-	-	-	-
W * % of single-parent families	-	-	-	-
W * % of renter-occupied housing units	-0.514	0.151	0.214	0.150
	(0.000)	(0.006)	(0.001)	(0.001)
W * % of population living in the same house as 5 years ago	-	-	-	-
W * % of vacant housing units	-	-	-	-

W * population density (in thousands)	-	-	-	-
Decennial change variables				
Δ_{00-90} of population below poverty level	-	-	-	-
Δ_{00-90} of median household income (in thousands)	-	-	-	-
Δ_{00-90} of educated population (log)	0.056	-0.017	-0.023	-0.016
	(0.042)	(0.068)	(0.047)	(0.047)
Δ_{00-90} of Heterogeneity Index (log)	-	-	-	-
Δ_{00-90} of foreign-born population	-	-	-	-
Δ_{00-90} of Hispanics and Latinos	-	-	-	-
Δ_{00-90} of single-parent families	-	-	-	-
Δ_{00-90} of renter-occupied housing units	-	-	-	-
Δ_{00-90} of population living in the same	0.003	-0.001	-0.001	-0.001
house as 5 years ago	(0.005)	(0.017)	(0.008)	(0.009)
Δ_{00-90} of vacant housing units	-	-	-	-
Δ_{00-90} of population density (in thousands)	-	-	-	-
Other independent variables				
Median age of housing units	-0.009	0.003	0.004	0.002
	(0.000)	(0.004)	(0.000)	(0.000)
Distance to CBD (in thousands)	0.100	-0.029	-0.041	-0.029
	(0.000)	(0.000)	(0.000)	(0.000)
X Coordinates	-0.173	0.051	0.072	0.050
	(0.000)	(0.000)	(0.000)	(0.000)
Y Coordinates	0.338	-0.099	-0.141	-0.099
	(0.000)	(0.002)	(0.000)	(0.000)
Squared X	-	-	-	-
Squared Y	-0.049	0.014	0.020	0.014
	(0.000)	(0.002)	(0.000)	(0.000)
X times Y	-	-	-	-

Note: *p*-values are reported in parentheses.

As discussed above, spatial association might be a concern. We thus proceed to assess the spatial fit of the model by using the spatial fit indicator previously described. A vector of predicted categories is obtained by using the estimated probabilities from the ordered probit model and a multinomial random number generator. The parameters for this analysis are as follows. Spatial contiguity is defined by *m*-surrounding of size 3 with an overlap degree of 1 (i.e. at most one common neighbor between proximate *m*-surroundings). Since there are 4 outcomes (i.e. the number of categories), this yields a ratio of symbolized locations to symbols equals of 5.08, which exceeds the recommended minimum value of 5. As seen in Table 3.3, the spatial fit indicator Φ has a value of -0.0356, and the 95% confidence interval of Φ is [-0.1146, 0.0158] (note the change of sign). This provides evidence that the degree of spatial association between the dependent variable and the projected variable is statistically the same. In other words, we can conclude that the model is satisfactory in terms of replicating the observed spatial association in cluster membership.

3.5 Discussion

A synthesis of the results of spatial scan statistics and ordered probit regression leads to several relevant conclusions. First, the spatial patterns of crime clusters in each year are consistent with the concentric zone model proposed by Park, Burgess, and McKenzie (1967) in SDT which suggests that crime is highest in zone of transition (near the CBD) and lowest in commuter zone (near the city's edge). This implies that more crime prevention resources should be allocated to transition zone of the city to break the persistence of clusters. What's more, compared to the western and northern areas, more attention shall be paid on the eastern and southern persistent clusters during the period. Second, the results from ordered probit model reveal several possible mechanisms driving the persistence of violent crime hot spot. This finding has practical and theoretical implications.

First, traditional SDT treats immigration as a disorganizing force leading to increased crime rates in neighborhoods. Controlling for poverty rates, however, our results indicate that a spatially lagged immigrant population variable has a negative effect on the persistence of crime hot spots. Our finding thus adds credibility to the "immigration revitalization" hypothesis (see Lee and Martinez 2002, p. 365), which argues that "immigration revitalizes poor areas and strengthens social control due to strong familial and neighborhood institutions and enhanced job opportunities associated with enclave economies, and thereby decreases crime" (Lee and Martinez 2002, p. 366). Furthermore, since this is a spatially lagged variable in the present analysis, this suggests that the effect may lead to "diffusion of benefits" beyond the neighborhood. More importantly, this finding extends our understanding concerning this relationship by confirming that the negative effect is not only cross-sectional, but also lasting over longer periods of time.

The model also reveals that concentrated poverty, concentrated rental residences, and decennial change of educated population are among the strongest variables, in addition to the immigrant concentration. Although the decennial change in poverty rates is not significant, reducing the base level of concentrated poverty tends to reduce the probability that a zone will be part of a persistent hot spot. Several policy tools exist that address poverty rates, including job creation programs in socio-economic deprived areas, changing tax collection policies specified for low income families, increasing social security/disability benefits, and so on. Other programs that increase the income of poor residents could prove effective as well, such as supplemental security insurance, unemployment insurance, temporary assistance for underprivileged families, and veteran's benefits. Furthermore, more police resources and crime reduction efforts could be allocated to neighborhoods that have a large number of rental residences and aged housing units. From a long run perspective, growth of education level can also steadily lower the probability that a BG being part of a persistent cluster.

3.6 Conclusions

This study investigated the spatial and temporal nature of the persistence of violent crime hot spots in Columbus from 1994 to 2002. Our findings provide insights regarding the neighborhood structural factors that influence the persistence of crime hot spots. Our results contribute to the literature in a number of ways. First, our findings lend support to Cahill's (2004, p. 17) argument that a synthesis of the two theories provides "the most robust theoretical explanation for geographical studies of crime". Second, the negative effect of immigrant concentration adds credibility to the immigration revitalization perspective. Third, the results extend our knowledge regarding the effects of socio-economic deprivation and residential mobility by confirming that not only cross-sectional violent crime patterns can be accounted for by cross-sectional socio-economic deprivation and residential mobility, but also the decennial variation of socio-economic deprivation and residential mobility tend to impact the longitudinal persistence of violent crime hot spots. This issue may not be that important for cross-sectional analysis of crime, but when it comes to the longitudinal study of crime pattern persistence, noticing the unequal performance of social disorganization elements may benefits the rational selection of explanatory variables. What's more, these findings provide strategic insights to enable the government and law enforcement agencies to design focused crime reduction policies and efficient crime prevention strategies.

It bears noting that the findings in this study are limited to violent crime. A thorough consideration of spatial analytical issues was given in this study, which increases our confidence on the reliability of results. Several avenues for future research can be discerned, including exploring hot spot persistence at different time scales (particularly longer time periods), examining the persistence of other disaggregated types of violent crime, identifying neighborhood structural predictors that influence the persistence of individual types of crime, and comparing the consistency and difference between aggregated and individual types of violent crime.

References

Amaral, P. V., Anselin, L., & Arribas-Bel. D. (2013). Testing for spatial error dependence in probit models. Letters in Spatial and Resource Sciences, 6(2), 91-101.

Andresen, M. A. (2006a). Crime measures and the spatial analysis of criminal activity. British Journal of Criminology, 46(2), 258-85.

Andresen, M. A. (2006b). A spatial analysis of crime in Vancouver, British Columbia: A synthesis of social disorganization and routine activity theory. The Canadian Geographer/Le Géographe canadien, 50(4), 487-502.

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

Andresen, M. A., & Malleson, N. (2013). Crime seasonality and its variations across space. Applied Geography, 43, 25-35.

Anselin, L. (1988). Spatial econometrics: Methods and models. Dordrecht: Kluwer.

Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: an introduction to spatial data analysis. Geographical Analysis, 38(1), 5-22.

Braga, A. A., Papachristos, A. V., & Hureau, D. M. (2014). The effects of hot spots policing on crime: An updated systematic review and meta-analysis. Justice Quarterly, 31(4), 633-663.

Cahill, M. E., & Mulligan, G. F. (2003). The determinants of crime in Tucson, Arizona1. Urban Geography, 24(7), 582-610.

Cahill, M. E. (2004). Geographies of Urban Crime: An Intraurban Study of Crime in Nashville, TN, Portland, OR and Tucson, AZ. Doctoral dissertation. University of Arizona.

Ceccato, V., & Haining, R. (2004). Crime in border regions: The Scandinavian case of Öresund, 1998–2001. Annals of the Association of American Geographers, 94(4), 807-826.

Chen, J., Roth, R. E., Naito, A. T., Lengerich, E. J., & MacEachren, A. M. (2008). Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of US cervical cancer mortality. International Journal of Health Geographics, 7(1), 57.

Chun, Y. (2014). Analyzing space–time crime incidents using eigenvector spatial filtering: an application to vehicle burglary. Geographical Analysis, 46(2), 165-184.

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

Cozens, P. M., Saville, G., & Hillier, D. (2005). Crime prevention through environmental design (CPTED): a review and modern bibliography. Journal of Property Management, 23(5), 328-356.

Craglia, M., Haining, R., & Wiles, P. (2000). A comparative evaluation of approaches to urban crime pattern analysis. Urban Studies, 37(4), 711-729.

Eck, J. E., & Weisburd, D. L. (2015). Crime places in crime theory. Crime and place:

Crime prevention studies, 4, 1-33.

Griffith, D. A. (1988). Advanced spatial statistics: Special topics in the exploration of quantitative spatial data series. Dordrecht: Kluwer.

Haining, R. (1993). Spatial data analysis in the social and environmental sciences. Cambridge University Press.

He, L., Páez, A., Liu, D., & Jiang, S. (2015). Temporal stability of model parameters in crime rate analysis: An empirical examination. Applied Geography, 58, 141-152.

Johnson, S. D., & Bowers, K. J. (2008). Stable and fluid hotspots of crime: differentiation and identification. Built Environment, 34(1), 32-45.

Kelly, M. (2000). Inequality and crime. Review of Economics and Statistics, 82(4), 530-39.

Kulldorff, M. (2011). SaTScan user guide for version 9.0.” Accessed 1 March 2015 from http://www.satscan.org.

Law, J., & Quick, M. (2013). Exploring links between juvenile offenders and social disorganization at a large map scale: a Bayesian spatial modeling approach. Journal of Geographical Systems, 15(1), 89-113.

Law, J., Quick, M., & Chan, P. W. (2015). Analyzing hotspots of crime using a Bayesian spatiotemporal modeling approach: a case study of violent crime in the Greater Toronto Area. Geographical Analysis, 47(1), 1-19.

Lee, M. T., & Martinez Jr, R. (2002). Social disorganization revisited: Mapping the recent immigration and black homicide relationship in northern Miami. Sociological Focus, 35(4), 363-380.

López, F. A., Angulo, A., & Artal. A. (2012). Capturing spatial clusters of activity in the Spanish Mediterranean Axis." In Defining the spatial scale in modern regional analysis: New challenges from data at local level, 255-80, edited by E. F. Vázquez and F. R. Morollón. Berlin Heidelberg: Springer.

Minnesota Population Center. (2011). "National historical geographic information system: Version 2.0." Minneapolis, MN: University of Minnesota. https://www.nhgis.org/. Accessed 18 Feb, 2016.

Mohler, G. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. International Journal of Forecasting, 30(3), 491-497.

Nakaya, T., & Yano, K. (2010). Visualising crime clusters in a space-time cube: An exploratory data analysis approach using space-time kernel density estimation and scan statistics. Transactions in GIS, 14(3), 223-239.

Páez, A., Ruiz, M., López, F., & Logan, J. (2012). Measuring ethnic clustering and exposure with the Q statistic: an exploratory analysis of Irish, Germans, and Yankees in 1880 Newark. Annals of the Association of American Geographers, 102(1), 84-102.

Páez, A., López, F. A., Ruiz, M., & Morency, C. (2013). Development of an indicator to assess the spatial fit of discrete choice models. Transportation Research Part B: Methodological, 56, 217-233.

Páez, A., Farber, S., Mercado, R., Roorda, M., & Morency, C. (2013). Jobs and the Single Parent: An Analysis of Accessibility to Employment in Toronto. Urban Geography, 34(6), 815-842.

Park, R. E., Burgess, E. W., & McKenzie. R. D. (1967). The city. Chicago: University of Chicago Press.

Qiu, F., Zhang, C., & Zhou, Y. (2012). The development of an areal interpolation ArcGIS extension and a comparative study. GIScience & Remote Sensing, 49(5), 644-663.

Ratcliffe, J. H. (2002). Aoristic signatures and the spatio-temporal analysis of high volume crime patterns. Journal of Quantitative Criminology, 18(1), 23-43.

Ratcliffe, J. H. (2004). The hotspot matrix: A framework for the spatio-temporal targeting of crime reduction. Police Practice and Research, 5(1), 5-23.

Ratcliffe, J. H., & McCullagh, M. J. (1999). Hotbeds of crime and the search for spatial accuracy. Journal of Geographical Systems, 1(4), 385-398.

Ruiz, M., López, F., & Páez, A. (2010). Testing for spatial association of qualitative data using symbolic dynamics. Journal of Geographical Systems, 12(3), 281-309.

Ruiz, M., López, F., & Páez, A. (2012). Comparison of thematic maps using symbolic entropy. International Journal of Geographical Information Science, 26(3), 413-439.

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

Sherman, L. W. (1995). Hot spots of crime and criminal careers of places. Crime and Place, 4, 35-52.

Shiode, S. (2011). Street-level Spatial Scan Statistic and STAC for Analysing Street Crime Concentrations. Transactions in GIS, 15(3), 365-383.

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. Journal of the American Statistical Association, 74(367), 519-530.

Uittenbogaard, A., & Ceccato, V. (2012). Space-time clusters of crime in Stockholm, Sweden. Review of European Studies, 4(5), 148.

U.S. Census Bureau. (2003). Census of population and housing, summary file 4: Technical documentation." https://www.census.gov/prod/cen2000/doc/sf4.pdf. Accessed 3 March, 2015.

Weisburd, D., Bushway, S., Lum, C., & Yang, S. M. (2004). Trajectories of crime at places: A longitudinal study of street segments in the city of Seattle. Criminology, 42(2), 283-322.

Whalen, K. E., Páez, A., Bhat, C., Moniruzzaman, M., & Paleti, R. (2012). T-communities and Sense of Community in a University Town: Evidence from a Student Sample using a Spatial Ordered-response Model. Urban Studies, 49(6), 1357-1376.

Chapter 4 Built environment and violent crime: An environmental audit approach using Google Street View

4.1 Introduction

Social disorganization theory has consistently demonstrated the relationship between social disorder and violent crime. Accordingly, numerous studies argue that physical signs of neighborhood disorder can impact violent crime (Anderson et al., 2012). The important role of micro-level physical characteristics of built environment in inducing and deterring violent crime has been confirmed in a number of studies (Brantingham and Brantingham, 1993; Sampson and Raudenbush, 2004; Groff et al., 2009). This is intuitive: physical features of neighborhood environment are an integral part of neighborhood ecology (Greensberg and Schneider, 1994), and in this sense, obvious signs of physical decay tend to indicate an absence of informal social control and collective efficacy in the neighborhood (Sampson et al., 1997). This, in turn, can lead to disproportionate concentration of violent crimes. While researchers have become increasingly interested in these complex mechanisms of social-environmental interaction (Malleson and Birkin, 2012), their understanding have been limited by the lack of fine-scale quantitative data of the built environment.

Environmental audits are a direct way of developing inventories of the features of the built environment (Moniruzzaman and Páez, 2012a). Environmental audit tools for physical activities (e.g., walking, cycling) and neighborhood safety have undergone extensive development and testing over the past several decades (e.g., Perkins et al., 1992; Day et al., 2006; Clifton et al., 2007; Hoehner et al., 2007; Moniruzzaman and Páez, 2016). These inventories have been more successful in helping to decide "what" to audit. However, questions of "where" and "how" to audit remain under-investigated. In other words, current literature fails to address appropriate selection of target sites and effective data collection method. First, sites are conventionally selected through sampling (e.g., Perkins et al., 1993; McMillan et al., 2010). While convenient, this approach fails to account and control for underlying effects of socio-economic correlates – which may result in misleading conclusions (Moniruzzaman and Páez, 2012a). Environmental audits, therefore, should be conducted at target sites most likely to yield novel information. Second, most conventional environmental audits involve on-site observations (Edwards et al., 2013). Auditors have to visit sites and directly collect environmental data. This approach is costly, time-consuming, and burdensome, especially for large-scale studies and dispersed distribution of study sites. Effective tools to enable quick and systematic collection of environmental characteristics associated with violent crime remain under-investigated.

Technology advancements provide new tools to explore geographical landscapes (Clarke et al., 2010; Chen et al., 2013, He et al., 2013). Google Street View (GSV), for instance, has recently been used as a tool to investigate urban environment (Salesses et al., 2013). It has been successful in helping to virtually and remotely audit environmental features associated to public health issues and physical activities (Clarke et al., 2010; Edwards et al., 2013). The results of existing research confirm the reliability of this approach for all but the most ephemeral attributes of the environment even over a period of years (Clarke et al., 2010). While much research has been devoted to walkability, for instance, to date little attention has been given to the use of these technologies to identify environmental features associated with violent crime.

In this research, we illustrate how GSV can be used to study the relationship between violent crime and physical characteristics of urban residential blocks. Specifically, we explore two questions: “where” the audit needs to be conducted, and “how” to effectively audit by considering the purpose and cost. This study contributes to the literature in the following ways. First, we address the question of “where” to audit by using Poisson regression model with spatial filtering (Griffith, 2002). After controlling for relevant socio-economic correlates, target sites can be determined by analyzing residual spatial pattern as retrieved by a spatial filter. Second, we use GSV to systematically identify physical features associated to violent crime, after socio-economic correlates have been controlled for in the model. This procedure is fairly general, and can be applied to other cases where on-site audits are challenging or unfeasible. Third, we develop a desktop data collection tool to facilitate virtual audit data collection. With some modification, the tool can be retooled for multi-purpose environmental audits. Fourth, this study provides an alternative way to re-evaluate influential theories of crime. Results obtained from the Poisson estimation and GSV-based environmental audit give credence to social disorganization theory, routine activity theory, and broken window theory. This can help to provide practical insights for Crime Prevention Through Environmental Design (CPTED) initiatives and inform focused crime prevention efforts.

The remainder of the chapter is organized as follows. In section 2, we summarize previous studies on the relationship between physical features of the built environment and violent crime. We also discuss current methods used for environmental audit. In section 3, we discuss the modeling techniques, development of desktop data collection tool, and GSV-based audit procedure. In section 4, we provide the results of Poisson model estimation, audit site selection, and analysis of contingency tables. In section 5, we discuss the theoretical and practical implications of the study. The chapter then concludes with a discussion of contributions, future directions, and limitations of the present study.

4.2 Background

The literature on built environment and violent crime provides evidences that the environmental context can induce or inhibit violent crime (Sampson and Raudenbush, 2004; Spicer et al., 2012). Of particular interests are various types of fine-scale physical characteristics of built environments. To better assess the physical features, various audit instruments have been developed, that can also be enhanced with emerging geospatial technologies. In this section, a brief review of primary criminological theories and empirical evidence is presented, followed by a summary of current environmental audit approaches.

4.2.1 Physical features of built environment and violent crime

The primary criminological theories underlying current empirical studies of environmental correlates of crime are broken windows theory (BWT) (Wilson and Kelling, 1982), crime prevention through environmental design (CPTED) (Jeffery, 1971), and situational crime prevention (SCP) (Clarke, 1995).

The BWT argues that visible evidence of physical disorder, such as broken windows, vacant/abandoned properties and abandoned cars on the street, are not only evidence of possible criminal activity, but can also facilitate further crime and other deviant behaviors (Wilson and Kelling, 1982). This link has previously been discussed in the work of Jacobs (1961), who emphasized the important role of pedestrian traffic, visual signs of environmental disorder, and social interaction between residents. A decayed physical environment conveys at least two messages: 1) the physical environment of neighborhood is neglected; and/or 2) informal and formal social control for deviant behavior has broken down (Cohen et al., 2000). The BWT literature emphasizes the importance of well-kept property to keep it from falling into disarray.

Consistent with BWT, literature on the CPTED argues that crime and fear of crime can be inhibited though proper design and modification of physical features of residential environment (Crowe, 2000). Seven key elements are suggested to create “defensible spaces” and reduce crime opportunities (see inter alia Marzbali et al., 2012; Cozens and Love, 2015). These are 1) territorial reinforcement (e.g., fences, property signs, etc.); 2) surveillance; 3) access control; 4) activity support (e.g., legitimate activities in parks); 5) image/maintenance (e.g., graffiti and litter); 6) target hardening (e.g., locks, alarms); and 7) geographical juxtaposition (i.e., surrounding land uses, such as schools or bars). Similar to CPTED, SCP also suggests that crime opportunities can be reduced through environmental modification, but with a focus on certain types of crime and involving the use of technologies and procedures (Cornish and Clarke, 2003). A growing body of research provides evidence that various types of violent crime can be controlled through CPTED and SCP strategies. An extended review of these concepts can be found in Cozens and Love (2015).

Empirical studies give credence to the above theories by evaluating the effects of various aspects of physical environment on violent crime, such as vegetation (Wolfe and Mennis, 2012), street lighting (Steinbach et al., 2015), and graffiti and litter (Loukaitou-Sideris et al., 2001), among many other features related to the design of residential and non-residential property (Marzbali et al., 2012).

Specifically, the effects of vegetation are mixed and depend on the type of vegetation. Vegetation can provide cover to potential criminals (Donovan and

Prestemon, 2012), and can provide surveillance cover to potential victims (Wolfe and Mennis, 2012). Small and view-obstructing trees are associated with increased crime (Kuo et al., 1998). However, larger trees are associated with decreased crime, as they signal the presence of informal/formal social control in neighborhood and increase the risk of committing crime (Donovan and Prestemon, 2012).

Street lighting has been associated to fear of crime and perception of safety, for a review see Farrington and Welsh (2002). Although violent crime can happen both at daytime and nighttime, studies found that improved lighting can reduce fear of crime (e.g., Welsh and Farrington, 2008). Although some studies have failed to find significant correlations between increase in crime and reduction in lighting (e.g., Pain et al., 2006), others suggest that decayed lighting leads to reduced visibility and deters residents from making journeys. In turn, this may result in an absence of sufficient surveillance on the street, decrease of community pride, loss of informal social control, and further opportunities for criminal activity (Steinbach et al., 2015).

In terms of graffiti, many studies understand it as vandalism (Shobe and Banis, 2014). The presence of graffiti can produce disorder, chaos, and fear in neighborhoods, and indicates that crime is left unchecked (Doran and Lees, 2005). Austin and Sanders (2007) also confirmed the strong link between the presence of graffiti and concentration of violent crime, and the effects vary by the types of graffiti. Based on those findings, literature suggests that a comprehensive elimination of urban graffiti is needed (Keizer et al., 2008).

In addition to the above physical features, studies have also associated violent crime with vacant and abandoned properties (Hannon, 2005), abandoned cars (Braga and Bond, 2008), and block watch signs (Donovan and Prestemon, 2012). These findings find their roots in both theoretical and practical development of criminological theories such as BWT, CPTED and SCP. Like all theories, these are enriched by empirical studies and continue to evolve as controversial evidence emerges. For example, some studies on BWT found little evidence for the effectiveness of policing practices in reducing crime (e.g., Harcourt and Ludwig, 2006). Some studies on CPTED have found contradictions between territoriality and surveillance (e.g., Reynald, 2009). Territoriality is enhanced by high fences due to their ability to keep people out of the property, whereas surveillance may be hampered by the presence of fences. However, as Cozen and Love (2015) suggested, future practices need to absorb emerging evidence and theories from environmental criminology to further assess the effectiveness of CPTED.

Our review of the literature on the relationship between physical environment and violent crime provides some insights.

Firstly, many argue that both physical disorder and violent crime can be affected by concentrated socio-economic disadvantage, therefore, the link may be blurred in the absence of controls for poverty and collective efficacy (Sampson and Raudenbush, 2004; Gau and Pratt, 2010). Furthermore, the effects of physical features may be conditional on socio-economic conditions, such as variation of population mobility (Wilcox, 2004) and disadvantage index (Stucky and Ottensmann, 2009). Therefore, to more effectively measure the effect of the physical environment, social structural conditions need to be controlled by means of a wide array of social disorganization variables and routine activity variables. This is something we achieve in our approach by using a statistical approach to controlling for a wide array of covariates before selecting sites for environmental audits. In this way the effects of physical disorder can be explored in a more systematic way.

And secondly, the above theories are influential but as noted they remain the subject of debate. Studies of the rigorous influence of BWT and CPTED have remained speculative, due in part to the difficulty in data collection of fine-scale physical features. Conventional data collection practices provide valuable methods, but are limited by the involvement of burdensome field work and cost. The approach illustrated in this study allows for cost-effective systematic collection of environmental data associated with violent crime.

4.2.2 Environmental audit approach

Environmental audits represent a direct and practical way to assess physical features of environment associated with physical activity and safety. Environmental audit programs for the purpose of crime prevention have been launched in many jurisdictions, including Toronto's Safety Audit Initiative (Thompson and Gartner, 2007), City of Saskatoon's Local Area Planning Program (Kellett et al., 2008), and Edmonton's Safer Cities Initiatives (City of Edmonton, 2000). Numerous audit instruments have been developed, such as the Pedestrian Environment Data Scan (PEDS) (Clifton et al., 2007), the Revised Block Environmental Inventory (RBEI) (Perkins et al., 1992), the Irvine-Minnesota Inventory (IMI) (Day et al., 2006), the University of Maryland Urban Design Tool (Ewing et al., 2006), the Analytic Audit Tool (Hoehner et al., 2007), the Systematic Pedestrian and Cycling Environmental Scan (SPACES) (Pikora et al., 2002), the Public Open Space Desktop Auditing Tool (POSDAT) (Edwards et al., 2013), and several manuals for CPTED (ODPM, 2004).

Despite the widespread availability of these instruments, fine-scale environmental data are not typically collected in a systematic way (Parmenter et al., 2008). Conventional audit practices require environmental features to be collected through on-site observation. Trained auditors have to walk or drive through the audit sites, record environmental data through filling out paper checklist (Clifton et al., 2007), filling digital checklist on handheld device (Moniruzzaman and Páez, 2012a), videotaping (Cohen et al., 2000), on-site photographing (Austin and Sanders, 2007), or using GPS-enabled geospatial video (Mills et al., 2010) and/or GPS-enabled passive photographing (Oliver et al., 2013). The need to deploy research personnel on site makes audits time-consuming, costly, burdensome, and potentially risky.

Fortunately, the growth of geospatial technologies now provides increased opportunities to remotely and virtually observe and characterize geographical landscapes. Google Street View (GSV), for instance, has recently been utilized to investigate urban built environments (Clarke et al., 2010; Edwards et al., 2013). A growing body of studies identifies GSV as a valuable and reliable data source for the assessment of environmental correlates of physical activities such as walking (Griew et al., 2013) and cycling (Vanwolleghem et al., 2014), and health issues such as children's healthy behaviour (Odgers et al., 2012) and elder's wellbeing (Burton et al., 2011). It has been suggested that GSV should be used with caution, due to the temporal instability of date stamps (month and year) and absence of time information (day) (Curtis et al., 2013; Salesses et al., 2013). However, previous studies generally report accurate and consistent agreements between field audit and GSV-based virtual audit, and dramatic savings of up to 90 per cent of audit time (e.g., Clarke et al., 2010; Edwards et al., 2013).

Compared to conventional on-site audit, GSV has several advantages. First, when the study sites are dispersed across a large city, on-site audit can be challenging and costly. However, in this case, GSV-based virtual audit can reduce transportation time and cost and increase efficiency. Second, compared to on-site observation which suffers from temporal fluctuation of environmental features, GSV provides a unified data source for studies of various purposes. The present and historical GSV imageries are provided with date stamps (month and year). Results obtained from GSV-based audit are comparable and can be cross-validated. Third, in-person field audit in high crime neighborhoods may put auditors at risk. However, GSV-based audit allows auditors remotely and virtually observe the study sites without taking any risk.

Although the novel tool has been increasingly employed, it has rarely been utilized to identify physical features associated with violent crime. Salesses et al. (2013) studied the link between urban environment and perception of safety using GSV imagery. Their study provides empirical test of GSV methods, but does not follow a systematic site selection procedure. This is a point that our approach addresses.

4.3 Data and methods

A systematic method to measure the effect of physical environment on violent crime is used. At the foundation of our approach are statistics of violent crime and a wide array of social disorganization variables and routine activity variables. These variables are examined using a Poisson regression model estimated at block group (BG) level. After controlling for social structural conditions, we hypothesize that the residual spatial pattern is attributable to the absence of physical environmental variables (one source of spatial autocorrelation being omitted relevant variables; McMillen, 2003). To retrieve the spatial residual pattern, we use eigenvector-based spatial filtering when estimating the Poisson model (Griffith, 2002). Selection of sites for environmental audits is based on analysis of the residual pattern, if any. Finally, these candidate sites are audited using GSV. In this section, we describe the technical aspects of this approach.

4.3.1 Crime data and census data

Violent crime data was obtained from the Columbus Division of Police. The dataset provides a series of spatial and temporal attributes, such as time, date, location and type of crime incidents. Four types of criminal offenses are classified as violent crime by the FBI: murder and non-negligent manslaughter, forcible rape, robbery, and aggravated assault. The period covered by the dataset is from 1998 to 2002.

Previous studies have suggested that crime data does not necessarily need to be averaged if the relationships between socio-economic indicators and multi-year crimes are significantly stable (He et al., 2015). We use violent crime of 2002 in the present study for two reasons. First, there is a stable temporal relationship in Columbus between violent crime and social structural correlates of crime (He et al., 2015). Second, the use of 2002 data minimizes the gap between crime data and street view imagery (i.e., 2007).

The dataset records 4,791 violent crimes in 2002. The geocoding success rate of crime points is 95.0 per cent. The distribution of violent crime of 2002 in Columbus is shown is Table 4.1. The variable is skewed to the right, with a range from 0 to 67 and mean value of 8.61. According to MacDonald and Lattimore (2010), this suggests that a Poisson distribution may appropriately fit these data.

Data from the US Census 2000 was used to generate socio-economic and demographic independent variables. The Census data is mainly collected from the Summary File 3 (SF3) from the US Census Bureau. Furthermore, supplementary data were obtained from the National Historical Geographic Information System (Minnesota Population Center, 2011). Independent variables were selected based on our review of social disorganization theory (SDT) (Sampson and Groves, 1989) and routine activity theory (RAT) (Cohen and Felson, 1979) as described next.

SDT uses four elements to explain crime: socio-economic deprivation (SED), residential mobility (RM), family disruption (FD), and ethnic heterogeneity (EH). Informed by extensive empirical literature (e.g., Andresen, 2006; Chun, 2014; He et al., 2015; He et al., 2016), we use poverty rate, percentage of educated population, and unemployment rate to capture SED. RM is captured by percentage of rental residences and locational stability (represented by population living in the same house as 5 years ago). FD is captured by number of single-parent families. EH is captured by immigrant concentration (represented by percentage of foreign-born population) and Heterogeneity Index. The index is defined as $1 - \sum {p_i}^2$ (Sampson and Groves, 1989), where i denotes the number of racial group, and p_i is the percentage of population within a given group. Racial groups in this study include: White; Black; American Indian, Eskimo, or Aleut; Asian or Pacific Islander; and other races. The index ranges from zero to one. A value of zero indicates maximum homogeneity, and one indicates maximum heterogeneity.

RAT uses three elements to understand crime: motivated offender (MO), suitable target (ST), and guardianship. We use percentage of young male population, poverty rate, and unemployment rate to capture MO. ST is captured by percentage of rental residences. Guardianship is captured by population density and percentage of vacant properties. Table 4.1 provides descriptive statistics for all variables used in the regression analysis. The collinearity test for all variables does not identify pair-wise correlation in excess of 0.8 or -0.8, a common threshold as suggested by Andresen (2006).

A positive relationship with crime is expected for poverty rate, unemployment rate, rental residences, single-parent families, Heterogeneity Index, vacant properties, and young male population. A negative relationship, on the other hand, is expected for educated population, locational stability, immigrant concentration, and population density.

Table 4.1 Descriptive statistics for variables used in the regression analysis (N = 567 observations)

Variable	**Mean**	**S.D.**	**Min**	**Max**
Dependent Variable				
Count of violent crime in 2002	8.61	9.74	0.00	67.00
Socio-economic Deprivation				
% of population below the poverty level (***Motivated***	18.52	15.91	0.00	76.81

Offenders)				
% of population with bachelor's degree	15.91	12.77	0.00	52.04
% of unemployed population (***Motivated Offenders***)	3.99	3.42	0.00	30.77
Ethnic Heterogeneity				
Heterogeneity Index	0.33	0.16	0.00	0.76
% of foreign-born population	6.09	8.40	0.00	95.86
Residential Mobility				
% of renter-occupied housing units ***(Suitable Targets)***	49.77	28.81	0.00	100.00
% of population living in the same house as 5 years ago	46.01	19.21	0.00	100.00
Family Disruption				
# of single-parent families	53.51	48.20	0.00	397.00
Guardianship				
population density (in thousands)	2.69	1.99	0.02	14.90
% of vacant housing units	8.22	6.65	0.00	46.00
Motivated Offenders				
# of young male population age 15-24 (log transformed)	4.13	1.21	-11.51	7.70

4.3.2 Poisson regression with eigenvector-based spatial filtering

The decision of "where" to audit is based on the use of Poisson regression with eigenvector-based spatial filtering (Griffith, 2002). The model is used to estimate the relationship between count of violent crime and a series of social disorganization and routine activity variables at BG level.

The use of Poisson regression in violent crime analysis is reviewed by MacDonald and Lattimore (2010). The unimodal and skewed nature of Poisson distribution makes Poisson regression more appropriate to model (typically skewed) crime counts. In order to better capture spatial autocorrelation and produce model estimates that are not affected by spatial autocorrelation, recent studies incorporate eigenvector-based spatial filtering (Griffith, 2002; Moniruzzaman and Páez, 2012a; Chun, 2014). As a nonparametric approach, it is developed to transform spatially autocorrelated residuals into spatially independent residuals by transferring the embedded residual pattern to the mean of the model. The model thus incorporates a filtered variable that yields clean residuals.

Spatial filtering uses a judiciously selected set of eigenvectors obtained from a transformed spatial weight matrix, which takes the following form:

$$(\mathrm{I} - 11^{\mathrm{T}}/\mathrm{n})\mathrm{C}(\mathrm{I} - 11^{\mathrm{T}}/\mathrm{n}) \tag{18}$$

where I denotes an identity matrix, 1 denotes a column vector of ones, T denotes the matrix transpose operator, n is the number of spatial units, and C denotes an n-by-n spatial weight matrix.

Eigenvectors are orthogonal and uncorrelated, and represent n distinct map patterns. A combination of eigenvectors can describe systematic spatial patterns and act as proxies for omitted variables in the specification of Poisson regression model. Therefore, they can help to identify spatial autocorrelation from the unexplained parts of dependent variable (Chun, 2014). A Poisson regression with eigenvector-based spatial filtering can be written in a following form (see Chun, 2014):

$$\mathrm{Exp}[\mathrm{Y}] = g^{-1}(\mathrm{X}\beta + \mathrm{E}\gamma) \tag{19}$$

where Y denotes the dependent variable which follows a Poisson distribution, Exp[Y] denotes the calculus of expectation operator, g denotes a link function (logarithm in most cases), X denotes a vector of linear independent variables, β is a vector of regression coefficients, E denotes the spatial filter (which is a linear combination of selected eigenvectors), and γ denotes a parameter associated with the spatial filter. An assumption of Poisson regression is that the conditional mean equals to variance. To minimize the bias of over-dispersion, the intervals of confidence can be corrected by including an estimation of dispersion parameter for the Poisson distribution (Moniruzzaman and Páez, 2012b).

The filtered spatially auto-correlated parts from model residuals provide information of systematic over- and under-estimation of the dependent variable. In other words, the filter identifies the locations where the behavior (i.e., violent crime) is more common than expected after controlling for the covariates (i.e., those where the model under-estimates crime), and locations where crime is less common than expected (i.e., those where the model over-estimates crime). After controlling for socio-economic correlates (and possibly meso-scale environmental features), we hypothesize that systematic over- and under-estimation are caused by the omission of finer-scale physical features of built environment that may encourage or hinder violent crime in a similar systematic way.

4.3.3 Audit site selection

Audit sites are selected based on the spatial filter in the following manner. The spatial filter is divided into Quantiles. In the present study we use three quantiles, Quantile 1 (Q1) for the bottom third of observations (negative values), Quantile 2 (Q2) for the middle third (around zero), and Quantile 3 (Q3) for the top third (positive values). Next, we select sites from BGs where crime counts are over-estimated (i.e., from Q1); further, we select sites from BGs where crime counts are under-estimated (i.e., from Q3).

For selection of audit sites we are interested in BGs in Q1 (over estimation) and Q3 (under estimation). We hypothesize that the systematic over- and under-estimation is related to the omission of fine-scale built environment variables that associate with violent crime but were not part of the model. To be specific, we hypothesize that: (1) in blocks where crimes are over-predicted (Q1) more built environment factors will be found that hinder crime; and (2) in blocks where crimes are under-predicted (Q3) more built environment factors that facilitate crime will be found.

During the process of selecting BGs we heed the suggestion that, in the manner of control-case studies, paired observations at similar levels of crime count from two groups should be selected (Moniruzzaman and Páez, 2012a). In other words, when one BG is selected from over-estimation group, the other BG having similar value of crime count shall be selected from under-estimation group. Those paired observations should include wide-ranging values from both groups. The size of selection should meet the demand of sample size for further statistical analysis (in the present case, contingency table analysis). And the distribution of selected sites should be as random as possible across the study area.

4.3.4 Environmental audit using Google Street View

Google Street View (GSV) is a technology featured in Google Maps and Google Earth (Google Inc., 2016). Launched on May 25, 2007, it provides 360° panoramic street-level imageries which covering 39 countries and about 3,000 cities. The GSV images were taken from positions along streets, and from a height of approximately 2.5 meter and at about 10 or 20 meter intervals. It allows users to virtually walk down streetscapes, and to navigate forward and backward, pan 360 degrees, vertically rotate 290 degrees, and zoom in/out.

To facilitate the virtual audit and data collection, a desktop audit tool was developed using ArcGIS Engine 10.0 (ESRI Inc., 2016) and Visual Studio 2008 (Microsoft Inc., 2016). The tool was developed based on the revised block environmental inventory (RBEI) designed by Perkins et al. (1992) to measure the physical environment of urban residential blocks. Shapefile data can be imported into the tool; environmental factors can be input using popup input windows, and can be saved as attribute data of corresponding polygon. Figure 4.1 shows the interfaces of the data collection tool.

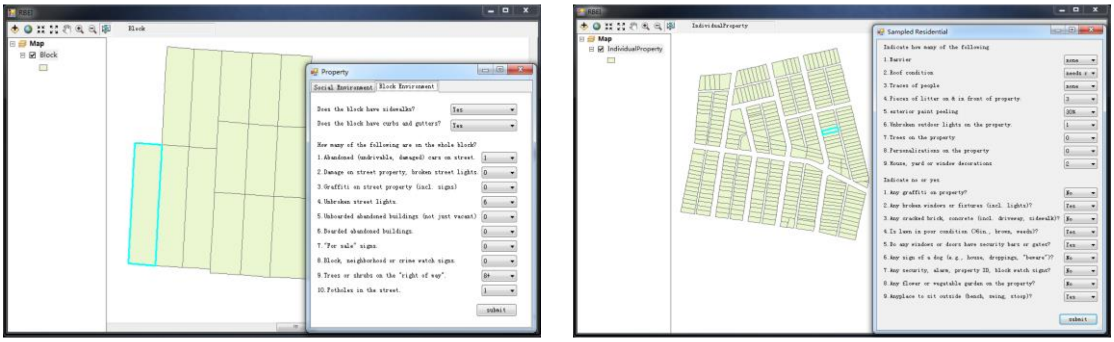

Figure 4.1 The interface of RBEI tool for block level (left) and individual property level (right) data collection

The RBEI measures three types of environmental features that have been associated with violent crime: physical signs of social disorder and incivility, territorial functioning, and defensible space (Perkins et a., 1992). First, it has been suggested that physical incivility (e.g., property damage and abandoned buildings) can influence residents' fear of crime or perception of safety. Second, territorial functioning features (e.g., yard decoration) reflect pride of ownership and territoriality. They have been linked to increased social interaction and sense of community. Third, the defensible space features can not only encourage territorial pride, but also physically hinder the commission of crime. For example, the presence of a fence can reflect both pride of territoriality as well as provide a physical barrier. These three types of features are measured at both residential block level and individual residential property level. Tables 4.2 and 4.3 provide the list of factors used to capture the features at two levels, along with the corresponding hypotheses to be examined based on the literature and the RBEI checklist.

Table 4.2 Block level environmental factors and corresponding hypotheses

Factors	Hypotheses
Physical Incivility	
Graffiti	Q1[a]: More blocks with none graffiti on street property, fewer blocks with presence of graffiti. Q3[b]: More blocks with presence of graffiti on street property, fewer blocks with none graffiti. See Doran and Lees (2005).
Damage on street property	Q1: More blocks with none damage on street property, fewer blocks with presence of damage. Q3: More blocks with presence of damage on street property, fewer blocks with none damage. See Perkins et al. (1992).
Unboarded abandoned buildings	Q1: More blocks with none unboarded abandoned buildings, fewer blocks with presence of unboarded abandoned buildings. Q3: More blocks with presence of unboarded abandoned buildings, fewer blocks with none unboarded abandoned buildings. See Hannon (2005).
Boarded abandoned buildings	Q1: More blocks with none boarded abandoned buildings, fewer blocks with presence of boarded abandoned buildings. Q3: More blocks with presence of boarded abandoned buildings, fewer blocks with none boarded abandoned buildings. See Hannon (2005).
Abandoned cars	Q1: More blocks with none abandoned/damaged cars on street, fewer blocks with presence of abandoned/damaged cars. Q3: More blocks with presence of abandoned/damaged cars, fewer blocks with none abandoned/damaged cars on street. See Braga and Bond (2008).
Potholes in the street	Q1: More blocks with none/few potholes in the street, fewer blocks with some/dense potholes in the street. Q3: More blocks with some/dense potholes in the street, fewer potholes with none/few potholes in the street.
Territorial Functioning	
Block or crime watch signs	Q1: More blocks with presence of crime watch signs, fewer blocks with none crime watch signs. Q3: More blocks with none crime watch signs, fewer blocks with presence of crime watch signs. See Donovan and Prestemon (2012).
Trees or shrubs	Q1: More blocks with some/dense trees or shrubs on the ‘right of way’, fewer blocks with none/few trees or shrubs. Q3: More blocks with none/few trees or shrubs on the ‘right of way’, fewer blocks with some/dense trees or shrubs. See Donovan and Prestemon (2012).
Defensible Space	
Unbroken street lights	Q1: More blocks with more than seven unbroken street lights, fewer blocks with less than seven unbroken street lights. Q3: More blocks with less than seven unbroken street lights, fewer blocks with more than seven unbroken street lights. See Farrington and Welsh (2002).

Note: [a] Q1 denotes Quantile 1 where crimes are over-predicted. [b] Q3 denotes Quantile

3 where crimes are under-predicted.

Table 4.3 Environmental factors at individual property level and corresponding hypotheses

Factors	**Hypotheses**
Physical Incivility	
Litter	Q1: More properties with none/few pieces of litter on & in front of property, fewer properties with some/dense pieces of litter. Q3: More properties with some/dense pieces of litter on & in front of property, fewer properties with none/few pieces of litter. See Loukaitou-Sideris et al. (2001).
Graffiti	Q1: More properties with none graffiti, fewer properties with presence of graffiti. Q3: More properties with presence of graffiti, fewer properties with none graffiti. See Loukaitou-Sideris et al. (2001).
Exterior paint peeling	Q1: More properties with none/few dilapidated exterior, fewer properties with some/dense dilapidated exterior. Q3: More properties with some/dense dilapidated exterior, fewer properties with none/few dilapidated exterior. See Perkins et al. (1993).
Roof condition	Q1: More properties with new roof, fewer properties with roof needs repair. Q3: More properties with roof needs repair, fewer properties with new roof. See Perkins et al. (1993).
Broken windows or fixtures	Q1: More properties with none broken windows or fixtures, fewer properties with presence of broken windows or fixtures.Q3: More properties with presence of broken windows or fixtures, fewer properties with none broken windows or fixtures. See Wilson and Kelling (1982).
Traces of people	Q1: More properties with animate traces of people, such as toys left outdoors, fewer properties with inanimate or none traces of people. Q3: More properties with inanimate or none traces of people, fewer properties with animate traces of people. See Perkins et al. (1992).
Cracked brick, concrete	Q1: More properties with none cracked brick and concrete, fewer properties with presence of cracked brick and concrete. Q3: More properties with presence of cracked brick and concrete, fewer properties with none cracked brick and concrete. See Perkins et al. (1993).
Territorial Functioning	

Trees on the property	Q1: More properties with more than two trees on the property, fewer properties with less than two trees. Q3: More properties with less than two trees, fewer properties with more than two trees on the property. See Perkins et al. (1993).
Personalizations on the property	Q1: More properties with more than one personalization on the property, fewer properties with less than one personalization. Q3: More properties with less than one personalization, fewer properties with more than one personalization. See Perkins et al. (1993).
House or yard decorations	Q1: More properties with more than three house/yard decorations, fewer properties with less than three house/yard decorations. Q3: More properties with less than three house/yard decorations, fewer properties with more than three house/yard decorations. See Perkins et al. (1993).
Signs of dog	Q1: More properties with presence of sign of dog, fewer properties with none sign of dog. Q3: More properties with none sign of dog, fewer properties with presence of sign of dog. See Perkins et al. (1993).
Security/watch signs	Q1: More properties with presence of security/watch signs, fewer properties with none security/watch signs. Q3: More properties with none security/watch signs, fewer properties with presence of security/watch signs. See Donovan and Prestemon (2012).
Garden	Q1: More properties with presence of flower/vegetable garden on the property, fewer properties with none flower/vegetable garden. Q3: More properties with none flower/vegetable garden, fewer properties with presence of flower/vegetable garden. See Perkins et al. (1993).
Lawn in poor condition	Q1: More properties with lawn in good condition, fewer properties with lawn in poor condition. Q3: More properties with lawn in poor condition, fewer properties with lawn in good condition. See Perkins et al. (1993).
Place to sit outside	Q1: More properties with place to sit outside, fewer properties with none place to sit outside. Q3: More properties with none place to sit outside, fewer properties with place to sit outside. See Perkins et al. (1993).
Defensible Space	
Barrier	Q1: More properties with barrier on property or perimeter, fewer properties with none barrier on property or perimeter. Q3: More

	properties with none barrier on property or perimeter, fewer properties with barrier on property or perimeter. See Cozens and Love (2015).
Unbroken outdoor lights	Q1: More properties with more than two unbroken outdoor lights on the property, fewer properties with less than two unbroken outdoor lights. Q3: More properties with less than two unbroken outdoor lights on the property, fewer properties with more than two unbroken outdoor lights. See Cozens and Love (2015).
Security bars on windows	Q1: More properties with presence of security bars or gates on windows or doors, fewer properties with none security bars or gates. Q3: More properties with none security bars or gates on windows or doors, fewer properties with presence of security bars or gates. See Cozens and Love (2015).

4.4 Experiments and results

4.4.1 Poisson regression analysis

The results of estimating the Poisson model for violent crime are shown in Table 4.4. A backward stepwise approach was used to specify the model, retaining all variables that were statistically significant at least at the conventional 0.05 level of significance. The goodness of fit of the model is summarized by a Pseudo R^2 of 0.530. An extreme low value of Moran's I in residuals (0.003) with an insignificant Z score (0.366) indicates that spatial autocorrelation has been removed from model residuals by the spatial filter. An over-dispersion parameter (2.066) was calculated to correct the intervals of confidence (for more details, see Moniruzzaman and Páez, 2012b).

The final model consists of eight significant variables all with signs as expected. Concretely, percentage of poverty population is significant and positive, while percentage of population with bachelor's degree is significantly negative. This indicates that BGs with higher poverty rate tend to have more violent crime, while BGs with higher educated population tend to have less violent crime. Percentage of foreign-born population has a significantly negative relationship with violent crime, which indicates that BGs with higher immigrant concentration experience lower violent crime. This is in agreement with the "immigrant revitalization" argument (Lee and Martinez, 2002; p. 365). Percentage of renter-occupied housing units is significantly positive, which suggests that BGs with more rental residences experience more violent crime. Single-parent family has a statistically significant and positive relationship with violent crime, which indicates that BGs with more disrupted families tend to experience more violent crime. Population density is significant and negative, thus indicating that higher population density in neighborhoods can contribute to significantly reduce violent crime. Percentage of vacant housing units is significantly positive, which indicates that BGs with more vacant properties experience increased violent crime. Young male population is significant and positive, which as expected suggests that BGs with higher young male population (15 to 24) tend to experience more violent crime. Among these variables, educated population is the strongest predictor in the model, while percentage of vacant properties is the second strongest one. Population density and young male population have the least effect on violent crime, compared to other significant predictors.

The spatial filter is synthesized in the usual way from 16 eigenvectors to yield a coefficient of 1. Selection of the eigenvectors was based on their ability to reduce the degree of residual spatial autocorrelation below significance. As seen in Figure 4.2, the distribution of the spatial filter is fairly symmetrical. The quantiles used for selection of BGs below are also shown there.

Table 4.4 Parameter estimates from the Poisson regression with spatial filter (the dependent variable is count of violent crime in 2002)

Variable	**Coefficient**	***p*-value**
Constant	1.013	0.000
Socio-economic Deprivation		
% of population below the poverty level (***Motivated Offenders***)	0.006	0.012
% of population with bachelor's degree	-0.023	0.000
% of unemployed population (***Motivated Offenders***)	-	-
Ethnic Heterogeneity		
Heterogeneity Index	-	-
% of foreign-born population	-0.010	0.028
Residential Mobility		
% of renter-occupied housing units ***(Suitable Targets)***	0.010	0.000
% of population living in the same house as 5 years ago	-	-
Family Disruption		
# of single-parent families	0.003	0.000
Guardianship		
population density (in thousands)	-0.081	0.000
% of vacant housing units	0.017	0.000
Motivated Offenders		
# of young male population age 15-24 (log transformed)	0.146	0.002
Spatial Filter (16 eigenvectors)	1.000	0.000
Pseudo R^2	0.530	
N	567	
Moran's I of residuals	0.003	
Z score for Moran's I	0.366	
Over-dispersion parameter (estimated)	2.066	

4.4.2 Selection of audit sites

For selection of sites for audits we classify the spatial filter into three quantiles, each containing 189 BGs (see Figure 4.3). The RBEI requires environmental audit is conducted at two levels: blocks and sampled residential properties, with at least eight residential properties sampled from each block. We selected 169 blocks from BGs in Q1 where we expect to find more crime-hindering factors; 162 blocks from BGs in Q3 where we expected to find more crime-facilitating factors. This generates a total of 331 blocks, which is a sufficient sample size for contingency table analysis. From those, we then selected 459 individual residential properties: 230 from over-predicted blocks and 229 from under-predicted blocks. Figure 4.3 shows that the selected block groups are randomly distributed across two quantiles.

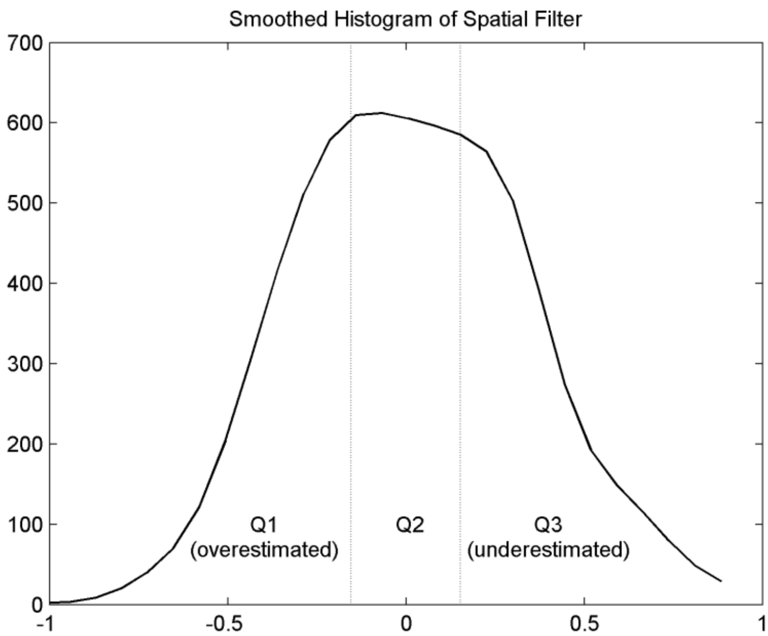

Figure 4.2 Smoothed histogram of spatial filter

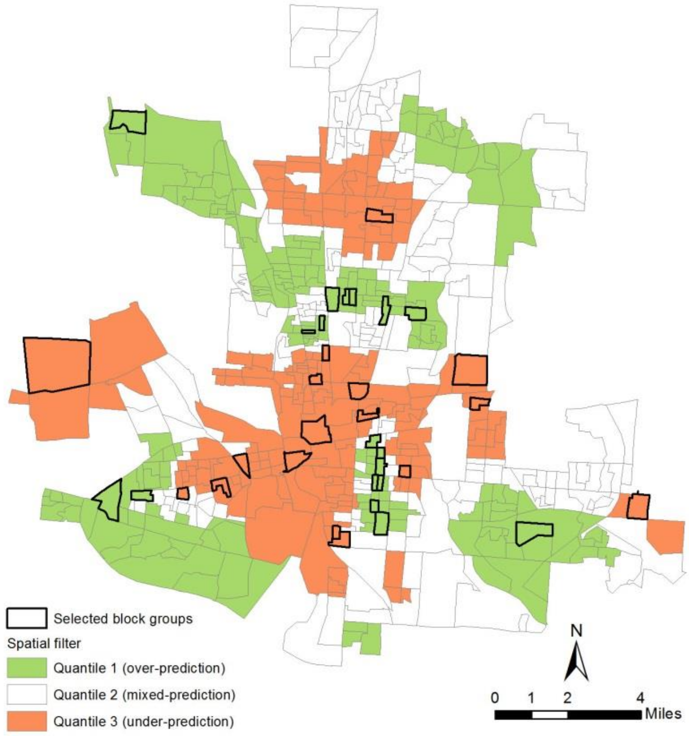

Figure 4.3 Selected BGs from Q1 and Q3, target sites are selected from those BGs

4.4.3 Environmental audit using Google Street View

The first author conducted the virtual audit in September, 2015. Using the desktop data collection tool with the RBEI embedded, environmental features as listed in Table 4.2 and 4.3 were identified from GSV. To avoid biasing the audits, blocks (and individual properties) from over- and under- predicted BGs were merged into a single shapefile so that at the time of the audits the auditor was not aware of the block's status as Q1 or Q3.

For most audit sites in Columbus, Google provides historical street view imageries available for several dates, for example, August 2007, May 2009, June 2011, June 2012, June 2014, August 2015, and November 2015. To minimize the interval between our crime data and GSV imageries, we collect physical features mainly from images drawn from August 2007. Images from 2009 were used less than 3% of the time, and only when 2007 imageries were absent for some blocks, the resolution of 2007 imageries was too coarse to identify certain environmental features, or environmental features (such as outdoor lights) were blocked by trees. The study by Clarke et al. (2010) indicates that a five year interval can provide reliable environmental indicators.

4.4.4 Contingency table analysis

The dBASE table of the shapefile that contains observed environmental factors of blocks and individual properties was analyzed using contingency tables. In the dBASE table, a number of environmental factors were recorded as categorical responses (e.g., absent or present). Other continuous responses (e.g., percentage of exterior paint peeling) were categorized for analysis. Contingency tables display the frequency distribution of observed cases from both Q1 and Q3. A chi-square (χ^2) test can be implemented to compare the observed frequency of each cell to the expected frequency under the null hypothesis of independence. Tables 4.5 and 4.6 present the result of this analysis of environmental factors at the level of blocks and individual properties.

At block level, the null hypothesis of independence is rejected for nine environmental factors at the 0.05 level of significance. Specifically, six physical features that capture physical incivility are significant. Two factors significantly capture territorial functioning. One factor significantly describes defensible space (see Table 4.5).

At the level of individual residential property, 17 factors are significant at least at the 0.05 level, and one factor that captures territorial functioning was found to be not significant. Of these factors, seven capture physical incivilities, seven more capture territorial functioning, and three capture defensible spaces (see Table 4.6).

Table 4.5 Summary of environmental factors at block level using $\chi 2$ independence tests (N = 331 residential blocks)

Environmental factors	**Chi-square statistic**	***p*-value**
Physical Incivility		
Graffiti	48.524	0.000
Unboarded abandoned buildings	29.478	0.000
Boarded abandoned buildings	92.985	0.000
Abandoned cars	17.791	0.000
Damage on street property	55.610	0.000
Potholes in the street	59.261	0.000
Territorial Functioning		
Block or crime watch signs	20.065	0.000
Trees or shrubs	8.336	0.004
Defensible Space		
Unbroken street lights	12.576	0.002

Table 4.6 Summary of environmental factors at individual property level using $\chi 2$ independence tests (N = 459 individual properties)

Environmental factors	**Chi-square statistic**	***p*-value**
Physical Incivility		
Litter	156.311	0.000

Graffiti	27.005	0.000
Exterior paint peeling	30.476	0.000
Roof condition	81.260	0.000
Broken windows or fixtures	27.005	0.000
Traces of people	78.095	0.000
Cracked brick, concrete	104.454	0.000
Territorial Functioning		
Trees on the property	33.937	0.000
Personalizations on the property	4.518	0.034
House or yard decorations	35.195	0.000
Signs of dog	5.649	0.017
Security/watch signs	65.049	0.000
Garden	66.465	0.000
Lawn in poor condition	33.885	0.000
Place to sit outside	2.465	0.116
Defensible Space		
Barrier	62.902	0.000
Unbroken outdoor lights	15.730	0.000
Security bars on windows	65.050	0.000

Tables 4.7 to 4.10 shows block-level factors analyzed using contingency tables. Physical signs include graffiti, damage on street property, potholes in street, unboarded and boarded abandoned buildings, and abandoned cars on street. Responses for potholes in street were categorized as none/few (observed potholes <= 2), and some/dense (observed potholes > 2). Observed frequencies of the presence of graffiti, property damage, abandoned buildings, abandoned cars, and potholes in street are significantly lower than expected in blocks of Q1 (where there is less crime than predicted by the model), but significantly higher than expected in blocks of Q3 (where there is more crime than predicted by the model).

Table 4.7 Physical incivility factors analyzed using contingency table (1)

		Graffiti		Damage on street property		Potholes in street		
		None	Present	None	Present	None/few	Some/dense	Total
Q1	Count	144	25	152	17	155	14	169*3
	Expected	114.4	54.6	121.5	47.5	124.1	44.9	
Q3	Count	80	82	86	76	88	74	162*3
	Expected	109.6	52.4	116.5	45.5	118.9	43.1	
	Total	224	107	238	93	243	88	331*3

Table 4.8 Physical incivility factors analyzed using contingency table (2)

		Unboarded abandoned buildings		Boarded abandoned buildings		Abandoned cars on street		
		None	Present	None	Present	None	Present	Total
Q1	Count	157	12	138	31	158	11	169*3
	Expected	137.9	31.1	94.5	74.5	144.5	24.5	
Q3	Count	113	49	47	115	125	37	162*3
	Expected	132.1	29.9	90.5	71.5	138.5	23.5	
	Total	270	61	185	146	283	48	331*3

We use unbroken street lights to capture defensible space at block level. These were audited as continuous responses and reclassified as three categorical responses for analysis: less than 4, 4 to 7, and more than 7. Table 4.9 shows that blocks in Q1 (actual crime less than predicted by model) exceed their expected frequencies of more than 7 unbroken street lights. However, for blocks in Q3 (actual crime exceeds expected value from model), the frequency of more than 7 unbroken lights is significantly lower than expected, and the frequencies of 4 to 7 and less than 7 unbroken lights are significantly higher than expected.

Table 4.9 Defensible space factors analyzed using contingency table

		Unbroken street lights			
		Less than 4	4 to 7	More than 7	Total
Q1	Count	36	83	50	169
	Expected	44.4	87.3	37.3	
Q3	Count	51	88	23	162
	Expected	42.6	83.7	35.7	
	Total	87	171	73	331

Territorial functioning is characterized by block or crime watch signs and trees/shrubs at block level. As shown in Table 4.10, for blocks of Q1, the presence of crime watch signs is significantly prevalent. For blocks of Q3, the observed presence of crime watch signs is significantly lower than expected. Trees/shrubs were audited as continuous responses and reclassified as two categories: none/few (trees/shrubs <= 3), some/dense (trees/shrubs > 3). Observations of some or dense trees are more frequent than expected in blocks of Q1, but none or few trees are more frequently observed than expected in blocks of Q3.

Table 4.10 Territorial functioning factors analyzed using contingency table

		Crime watch signs		Trees or shrubs		
		None	Present	None/few	Some/dense	Total
Q1	Count	127	42	39	130	169*2
	Expected	141.9	27.1	51.1	117.9	
Q3	Count	151	11	61	101	162*2
	Expected	136.1	25.9	48.9	113.1	
	Total	278	53	100	231	331*2

Tables 4.11 to 4.16 show the results contingency table analysis for individual property level variables. Physical incivilities at individual property level are captured by factors listed in Tables 4.11, 4.12 and 4.13. Continuous responses for litter and exterior paint peeling were reclassified as none/few (number of litter <= 2; percentage of peeling <= 10%), some/dense (number of litter > 2; percentage of peeling > 10%). Responses for graffiti, broken windows, and cracked brick are originally categorical: none and present. Responses for roof condition are categorical either: new roof, average roof, and roof needs repair. Categorical responses for traces of people are: animate, inanimate, and none. If can be seen that for individual residential properties in Q1, the presences of some/dense litter, graffiti, some/dense dilapidated exterior, roof needs repair, broken windows, cracked brick, and invisible traces of people are significantly lower than expected. However, these features significantly exceed the expectation for individual properties in Q3. Frequencies of properties with new roof and animate traces of people are significantly prevalent in Q1.

Table 4.11 Physical incivility factors (1)

		Litter		Graffiti		Dilapidated exterior		
		None/few	Some/dense	None	Present	None/few	Some/dense	Total
Q1	Count	212	18	213	17	198	32	230*3
	Expected	147.8	82.2	201.4	28.6	172.4	57.6	
Q3	Count	83	146	189	40	146	83	229*3
	Expected	147.2	81.8	200.6	28.4	171.6	57.4	
	Total	295	164	402	57	344	115	459*3

Table 4.12 Physical incivility factors (2)

		Roof condition			Broken windows or fixtures		
		New	Average	Needs repair	None	Present	Total
Q1	Count	91	127	12	213	17	230*2
	Expected	59.1	126.8	44.1	192.4	37.6	
Q3	Count	27	126	76	171	58	229*2
	Expected	58.9	126.2	43.9	191.6	37.4	
	Total	118	253	88	384	75	459*2

Table 4.13 Physical incivility factors (3)

		Traces of people			Cracked brick, concrete		
		Animate	Inanimate	None	None	Present	Total
Q1	Count	89	126	15	191	39	230*2
	Expected	55.6	130.8	43.6	137.3	92.7	
Q3	Count	22	135	72	83	146	229*2
	Expected	55.4	130.2	43.4	136.7	92.3	
	Total	111	261	87	274	185	459*2

Defensible space factors at individual property level include barriers, unbroken outdoor lights, and security bars or gates on windows or doors. Barriers can be walls, fences, or hedges of any sort or height. These features were audited as categorical responses: none, on property (barrier is part of the property), and on perimeter (barrier is around the perimeter of the entire property). As shown in Table 4.14, for properties in Q1, the frequency of properties with barrier around the perimeter, properties with more than two unbroken outdoor lights, and properties with security bars on windows is significantly higher than expected. The presence of properties without barrier is significantly less than expected. For properties in Q3, in contrast, the presence of properties with no barrier, no security bars, and less than or equal to two unbroken outdoor lights is more prevalent.

Table 4.14 Defensible space factors

		Barrier			Unbroken outdoor lights		Security bars on windows & doors		
		None	On property	Perimeter	<=2	>2	No	Yes	Total
Q1	Count	14	89	127	149	81	52	178	230*3
	Expected	47.1	83.7	99.2	167.9	62.1	66.1	163.9	
Q3	Count	80	78	71	186	43	80	149	229*3
	Expected	46.9	83.3	98.8	167.1	61.9	65.9	163.1	
	Total	94	167	198	335	124	132	327	459*3

Factors listed in Tables 4.15 and 4.16 capture territorial functioning. Trees on property were audited as continuous responses and reclassified into two categories for analysis: more than two trees, less than or equal to two trees. Continuous responses for personalization signs on property were categorized into two: more than one personalization, less than or equal to one personalization. Continuous responses for house or yard decoration were categorized as more than three, less than or equal to three. Other factors (sign of dog, security signs, flower or vegetable garden, condition of lawn, and place to sit outside) were originally audited as categorical responses: yes and no. As shown in the tables, for properties in Q1, the frequency of properties with more than two trees, more than one personalization, and more than three decorations is significantly higher than expected. The presence of properties with sign of dog, security signs and garden is significantly prevalent. The factor of place to sit outside on the other hand is not significant.

Table 4.15 Territorial functioning factors (1)

		Trees		Personalization		Decorations		Signs of dog		
		<=2	>2	<=1	>1	<=3	>3	No	Yes	Total
Q1	Count	130	100	207	23	87	143	171	59	230*4
	Expected	158.8	71.2	213	17	118.8	111.2	181.4	48.6	
Q3	Count	187	42	218	11	150	79	191	38	229*4
	Expected	158.2	70.8	212	17	118.2	110.8	180.6	48.4	
	Total	317	142	425	34	237	222	362	97	459*4

Table 4.16 Territorial functioning factors (2)

		Security/watch signs		Garden		Lawn in poor condition		Place to sit outside		
		No	Yes	No	Yes	No	Yes	No	Yes	Total
Q1	Count	138	92	39	191	189	41	18	212	230*4
	Expected	174.9	55.1	80.7	149.3	160.3	69.7	23.1	206.9	
Q3	Count	211	18	122	107	131	98	28	201	229*4
	Expected	174.1	54.9	80.3	148.7	159.7	69.3	22.9	206.1	
	Total	349	110	161	298	320	139	46	413	459*4

All of the significant environmental attributes lend support to our hypotheses presented in Tables 4.2 and 4.3.

First (see Figure 4.4), we find more built environment features that hinder crime in blocks where observed crimes are less than predicted (Q1, Green Zones). Figure 4.5 shows examples of these physical features: barriers are more frequently observed, along with security bars/gates, yard decoration, personalization signs of house number, good roof condition, crime watch sign and dog sign; where as a less frequent presence of abandoned building, car, graffiti, dense litter, poor condition of lawn, dilapidated exterior and broken windows.

Figure 4.4 Summary of significant environmental correlates of violent crime identified in GSV.

Correlates linked to green zones are crime-hindering factors which are more frequently found in BGs where observed crimes are less than predicted (Q1). Correlates linked to red zones are crime-facilitating factors which are more frequently found in BGs where observed crimes are more than predicted (Q3).

Figure 4.5 Examples of crime-hindering environment features found in GSV images. The left image shows barriers, security bars on windows, and yard decoration. The middle image shows personalization sign of house number, and good lawn and roof condition. The right image shows neighborhood crime watch sign.

Second, we generally find more built environment factors that facilitate crime in blocks where observed crimes exceed the expectation (Q3, Red Zones). Figure 4.6

shows examples of physical signs of environmental decay found through GSV: abandoned buildings and cars are more frequent, as are graffiti, dense litter, poor condition of lawn, dilapidated exterior and broken windows. On the other hand, there is a dearth of barriers, security bars/gates, house decoration and personalization, crime watch signs, and signs of dog.

Figure 4.6 Examples of physical signs of environmental decay found in GSV images. The left image shows abandoned building, graffiti on the door, and dense litter. The middle image shows abandoned cars, and poor condition of lawn. The right image shows abandoned building, dilapidated exterior, broken windows, dense litter, and poor condition of lawn.

Interpretation of several physical attributes must be conducted with a measure of caution. First, some aspects of properties cannot be observed from GSV images, which may lead to some bias in certain areas. For example, the number of unbroken outdoor light, percentage of exterior paint peeling, and house decorations. In our study, however, GSV provided images of front and back yards for most blocks, which gives us confidence about the accuracy of quantity of physical features in our audits.

Second, the RBEI suggests that on-site audit should be conducted between 5 pm an 8 pm on a weeknight, or between noon and 8 pm on weekend. Although all imageries were taken at daytime, we cannot know if said time requirement was satisfied because GSV did not provide the exact time when images were taken. Third, street lights and outdoor lights at daytime imageries may have not been turned on, which makes it sometimes difficult to identify broken lights from well-functioning lights. Therefore, unbroken street lights are collected through an approximate identification.

Third, caution also needs to be exerted when interpreting more ephemeral physical signs such as litter and graffiti. However, many studies suggested that there's a time-lag effect of socio-economic fluctuation on crime (Rosenfeld and Fornango, 2007), and crime itself can in turn impact socio-economic outcomes (Malby and Davis, 2012). In light of this, we conjecture the presence of a similar time-lag effect of environmental decay on crime, and a similar time-lag effect of crime on physical decay of residential environment. This is an issue in need of further research.

4.5 Discussion

Results obtained from Poisson regression with spatial filtering and GSV-based environmental audit inform several theoretical and practical implications.

The theoretical implications are as follows. First, the Poisson model with spatial filter provides a number of significant social-economic correlates of violent crime. The effects of said correlates on violent crime are consistent with previous studies in other settings. This finding not only adds credibility to SDT and RAT, but also supports that a combination of the two theories provides relatively comprehensive understanding of violent crime. Second, the GSV-based audit provides an alternative way to re-evaluate influential theories of crime. Significant environmental factors identified from GSV lend support to BWT and CPTED. Third, consistent with previous studies (e.g., He et al., 2016), the finding regarding the negative effect of immigrant concentration on violent crime lends empirical support to the "immigration revitalization" hypothesis proposed by Lee and Martinez (2002, p. 365).

The practical implications of our research can be summarized as follows. Environmental factors identified from GSV-based audit provide insights for CPTED initiatives, situational crime prevention efforts, and urban renewal projects. Police portal should be encouraged for neighborhoods presenting public signs of physical disorder. What's more, the identified socio-economic correlates also indicate that more police resources are required by neighborhoods with high percentage of rental and vacant residences. This study provides empirical test of GSV for the purpose of virtually identifying environmental correlates of violent crime. We find that GSV-based audit can reduce both human cost and travel expense, protect auditors from potential risks, and allow quantitative environmental data to be collected in a systematic way.

4.6 Conclusion

We investigate the relationship between residential built environment and violent crime in this chapter, through environmental audits with GSV. The contributions are as follows.

First, a model-based site selection method has proved its efficiency in determining “where” to audit. Second, model estimation with spatial filtering approach indicates that eigenvector spatial filtering can not only help to produce model estimates that are not biased by spatial autocorrelation, but also help to make further hypotheses based on an analysis of the systematic patterns in model residuals. Third, the GSV-based audit has proved effective in answering the question of “how” to audit. It can benefit future studies in cases of large-scale environmental assessment and dispersed distribution of study sites. Finally, the desktop and electronic version of RBEI can be reused in future environmental assessment of various purposes.

It bears mentioning that the findings of this study are limited to violent crime. Previous studies have empirically demonstrated the role of built environment in affecting disaggregated violent crimes such as homicide and robbery (e.g., Lasley, 1996; Webb and Laycock, 1992), and property crimes such as residential burglary (Breetzke, 2012) and auto burglary (Michael et al., 2001). Future studies are needed to replicate present analyses in those settings to evaluate the applicability of the proposed method.

The availability of historical GSV provides opportunities for other future studies such as time series study of built environment, comparative study of present landscape with historical environment. Imageries of different time nodes (e.g., August 2007, May 2009, June 2011, June 2012, June 2014, August 2015, and November 2015) are comparable. Similar to time series analysis of optical remote sensing data, environmental characteristics can also be observed from historical GSV imageries and can contribute to study the temporal dynamic variation of urban landscape. In addition to exclusively using GSV imageries, one can also take a mobile device and walk into the field. This allows the GSV-based virtual audit and on-site observation to be conducted simultaneously. The present urban landscapes can be directly compared with historical environment.

GSV brings opportunities and challenges simultaneously. Challenges of using GSV include: First, although the coverage of GSV imageries has been dramatically increased in recent years, its spatial completeness cannot be guaranteed in every street segment and every city. Future studies planning to conduct virtual audit using GSV need to preliminarily examine the availability of GSV in their study sites first. Second, although historical GSV imageries are time stamped, time interval between historical imageries varies upon individual study areas. Third, previous studies suggest that the reliability of fluid physical features (e.g., the presence of litter and graffiti) can be relatively lower than more stable features (e.g., block watch signs, trees on the property). Maintaining the contemporaneity of GSV data and crime data can help to reduce potential bias results from temporal instability of physical features.

Reference

Anderson, J. M., MacDonald, J. M., Bluthenthal, R., & Ashwood, J. S. (2012). Reducing crime by shaping the built environment with zoning: An empirical study of Los Angeles. University of Pennsylvania Law Review, 161, 699-756.

Andresen, M. A. (2006). Crime measures and the spatial analysis of criminal activity. British Journal of Criminology, 46(2), 258-285.

Austin, D. M., & Sanders, C. (2007). Graffiti and perceptions of safety: A pilot study using photographs and survey data. Journal of Criminal Justice and Popular Culture, 14(4), 292-316.

Braga, A. A., & Bond, B. J. (2008). Policing crime and disorder hot spots: A randomized controlled trial. Criminology, 46(3), 577-607.

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

Breetzke, G.D. (2012). The effect of altitude and slope on the spatial patterning of burglary. Applied Geography. 34, 66-75.

Burton, E. J., Mitchell, L., & Stride, C. B. (2011). Good places for ageing in place: development of objective built environment measures for investigating links with older people's wellbeing. BMC Public Health, 11: 839. doi: 10.1186/1471-2458-11-839.

Chen, M., Lin, H., Hu, M., He, L., & Zhang, C. (2013).

Real-geographic-scenario-based virtual social environments: integrating geography with social research. Environment and Planning B: Planning and Design, 40(6), 1103-1121.

Chun, Y. (2014). Analyzing space–time crime incidents using eigenvector spatial filtering: an application to vehicle burglary. Geographical Analysis, 46(2), 165-184.

City of Edmonton (2000). Safety audit guide for crime prevention. Edmonton, Canada: City of Edmonton. Accessed 24 June 2016. http://www.edmonton.ca/residential_neighbourhoods/PDF/SCACSafetyAuditGuide.pdf.

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

Clarke, R. V. (1995). Situational crime prevention. Crime and Justice, 91-150.

Clifton, K. J., Smith, A. D. L., & Rodriguez, D. (2007). The development and testing of an audit for the pedestrian environment. Landscape and Urban Planning, 80(1), 95-110.

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

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

Cornish, D. B., & Clarke, R. V. (2003). Opportunities, precipitators and criminal decisions: A reply to Wortley's critique of situational crime prevention. Crime Prevention Studies, 16, 41-96.

Cozens, P., & Love, T. (2015). A review and current status of crime prevention through environmental design (CPTED). Journal of Planning Literature, 30(4), 393–412.

Crowe, T. D. (2000). Crime prevention through environmental design: Applications of architectural design and space management concepts. Butterworth-Heinemann.

Curtis, J. W., Curtis, A., Mapes, J., Szell, A. B., & Cinderich, A. (2013). Using google street view for systematic observation of the built environment: analysis of spatio-temporal instability of imagery dates. International Journal of Health Geographics, 12: 53. doi: 10.1186/1476-072X-12-53.

Day, K., Boarnet, M., Alfonzo, M., & Forsyth, A. (2006). The Irvine–Minnesota inventory to measure built environments: development. American Journal of Preventive Medicine, 30(2), 144-152.

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

Doran, B. J., & Lees, B. G. (2005). Investigating the spatiotemporal links between disorder, crime, and the fear of crime. The Professional Geographer, 57(1), 1-12.

Edwards, N., Hooper, P., Trapp, G. S., Bull, F., Boruff, B., & Giles-Corti, B. (2013).

Development of a Public Open Space Desktop Auditing Tool (POSDAT): A remote sensing approach. Applied Geography, 38, 22-30.

ESRI Inc. (2016). http://www.esri.com/software/arcgis/arcgisengine. Accessed 29 June, 2016.

Ewing, R., Handy, S., Brownson, R. C., Clemente, O., & Winston, E. (2006). Identifying and measuring urban design qualities related to walkability. Journal of Physical Activity and Health, 3, S223-S240.

Farrington, D. P., & Welsh, B. C. (2002). Effects of improved street lighting on crime: a systematic review. London: Home Office.

Gau, J. M., & Pratt, T. C. (2010). Revisiting broken windows theory: Examining the sources of the discriminant validity of perceived disorder and crime. Journal of Criminal Justice, 38(4), 758-766.

Google Inc. (2016). https://www.google.ca/maps/streetview/. Accessed 29 June, 2016.

Greenberg, M., & Schneider, D. (1994). Violence in American cities: young black males in the answer, but what was the question? Social Science & Medicine, 39(2), 179-187.

Griew, P., Hillsdon, M., Foster, C., Coombes, E., Jones, A., & Wilkinson, P. (2013). Developing and testing a street audit tool using Google Street View to measure environmental supportiveness for physical activity. International Journal of Behavioral Nutrition and Physical Activity, 10:103. doi: 10.1186/1479-5868-10-103.

Griffith, D. A. (2002). A spatial filtering specification for the auto-Poisson model. Statistics & Probability letters, 58(3), 245-251.

Groff, E., Weisburd, D., & Morris, N. A. (2009). Where the action is at places: examining spatio-temporal patterns of juvenile crime at places using trajectory analysis and GIS. In Putting Crime in Its Place (pp. 61-86). New York: Springer.

Hannon, L. E. (2005). Extremely poor neighborhoods and homicide. Social Science Quarterly, 86(5), 1418-1434.

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

He, L., Chen, M., Lu, G., Lin, H., Liu, L., & Xu, J. (2013). An approach to transform Chinese historical books into scenario-based historical maps. The Cartographic Journal, 50(1), 49-65.

He, L., Páez, A., & Liu, D. (2016). Persistence of crime hot spots: an ordered probit analysis. Geographical Analysis. doi: 10.1111/gean.12107.

He, L., Páez, A., Liu, D., & Jiang, S. (2015). Temporal stability of model parameters in crime rate analysis: An empirical examination. Applied Geography, 58, 141-152.

Hoehner, C. M., Ivy, A., Brennan Ramirez, L. K., Handy, S., & Brownson, R. C. (2007). Active neighborhood checklist: a user-friendly and reliable tool for assessing activity friendliness. American Journal of Health Promotion, 21(6), 534-537.

Jacobs, J. (1961). Death and life of great American cities. New York: Vintage.

Jeffery, C. R. (1971). Crime prevention through environmental design. Beverly Hills, CA: Sage Publications.

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. Science, 322(5908), 1681-1685.

Kellett, L., Peter, L., & Moore, K. (2008). The City of Saskatoon's local area planning program: A case study. Social Indicators Research, 85(1), 159-167.

Kuo, F. E., Bacaicoa, M., & Sullivan, W. C. (1998). Transforming inner-city landscapes trees, sense of safety, and preference. Environment and Behavior, 30(1), 28-59.

Lasley, J.R., 1996. Using traffic barriers to "design out" crime: A program evaluation of LAPD's Operation Cul-De-Sac. Report to the National Institute of Justice. Fullerton, CA: California State University.

Lee, M. T., & Martinez Jr, R. (2002). Social disorganization revisited: Mapping the recent immigration and black homicide relationship in northern Miami. Sociological Focus, 35(4), 363-380.

Loukaitou-Sideris, A., Liggett, R., Iseki, H., & Thurlow, W. (2001). Measuring the effects of built environment on bus stop crime. Environment and Planning B: Planning and Design, 28(2), 255-280.

MacDonald, J. M., & Lattimore, P. K. (2010). Count models in criminology. In A. R.

Piquero & D. Weisburd (Eds.), Handbook of Quantitative Criminology (p. 683-698). New York: Springer

Malby, S., & Davis, P. (2012). Monitoring the impact of economic crisis on crime. Vienna, Austria: United Nations Office on Drugs and Crime.

Malleson, N., & Birkin, M. (2012). Analysis of crime patterns through the integration of an agent-based model and a population microsimulation. Computers, Environment and Urban Systems, 36(6), 551-561.

Marzbali, M. H., Abdullah, A., Razak, N. A., & Tilaki, M. J. M. (2012). The influence of crime prevention through environmental design on victimisation and fear of crime. Journal of Environmental Psychology, 32(2), 79-88.

McMillen, D. P. (2003). Spatial autocorrelation or model misspecification? International Regional Science Review, 26(2), 208-217.

McMillan, T. E., Cubbin, C., Parmenter, B., Medina, A. V., & Lee, R. E. (2010). Neighborhood sampling: how many streets must an auditor walk? International Journal of Behavioral Nutrition and Physical Activity, 7(1), 7-20.

Michael, S. E., Hull, R. B., & Zahm, D. L. (2001). Environmental factors influencing auto burglary a case study. Environment and Behavior, 33(3), 368-388.

Microsoft Inc. (2016). https://www.visualstudio.com/. Accessed 29 June, 2016.

Mills, J. W., Curtis, A., Kennedy, B., Kennedy, S. W., & Edwards, J. D. (2010). Geospatial video for field data collection. Applied Geography, 30(4), 533-547.

Minnesota Population Center. (2011). National historical geographic information system: Version 2.0. Minneapolis, MN: University of Minnesota. https://www.nhgis.org/. Accessed 18 June, 2016.

Moniruzzaman, M., & Páez, A. (2012a). A model-based approach to select case sites for walkability audits. Health and Place, 18(6), 1323-1334.

Moniruzzaman, M., & Páez, A. (2012b). Accessibility to transit, by transit, and mode share: application of a logistic model with spatial filters. Journal of Transport Geography, 24, 198-205.

Moniruzzaman, M., & Páez, A. (2016). An investigation of the attributes of walkable environments from the perspective of seniors in Montreal. Journal of Transport Geography, 51, 85-96.

Odgers, C. L., Caspi, A., Bates, C. J., Sampson, R. J., & Moffitt, T. E. (2012). Systematic social observation of children's neighborhoods using Google Street View: a reliable and cost-effective method. Journal of Child Psychology and Psychiatry, 53(10), 1009-1017.

ODPM/Home Office. (2004). Safer places: The planning system and crime prevention. London: HMSO.

Oliver, M., Doherty, A. R., Kelly, P., Badland, H. M., Mavoa, S., Shepherd, J., Kerr, J., Marshall, S., Hamilton, A., & Foster, C. (2013). Utility of passive photography to objectively audit built environment features of active transport journeys: an observational study. International Journal of Health Geographics, 12:20. doi: 10.1186/1476-072X-12-20.

Pain, R., MacFarlane, R., Turner, K., & Gill, S. (2006). ‘When, where, if, and but’: qualifying GIS and the effect of streetlighting on crime and fear. Environment and Planning A, 38(11), 2055-2074.

Parmenter, B., McMillan, T., Cubbin, C., & Lee, R. E. (2008). Developing geospatial data management, recruitment, and analysis techniques for physical activity research. Journal of the Urban and Regional Information Systems Association, 20(2), 13-19.

Perkins, D. D., Meeks, J. W., & Taylor, R. B. (1992). The physical environment of street blocks and resident perceptions of crime and disorder: Implications for theory and measurement. Journal of Environmental Psychology, 12(1), 21-34.

Perkins, D. D., Wandersman, A., Rich, R. C., & Taylor, R. B. (1993). The physical environment of street crime: Defensible space, territoriality and incivilities. Journal of Environmental Psychology, 13(1), 29-49.

Pikora, T. J., Bull, F. C., Jamrozik, K., Knuiman, M., Giles-Corti, B., & Donovan, R. J. (2002). Developing a reliable audit instrument to measure the physical environment for physical activity. American Journal of Preventive Medicine, 23(3), 187-194.

Reynald, D. M. (2009). Guardianship in action: Developing a new tool for measurement. Crime Prevention & Community Safety, 11(1), 1-20.

Rosenfeld, R., & Fornango, R. (2007). The impact of economic conditions on robbery and property crime: The role of consumer sentiment. Criminology, 45(4), 735-769.

Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The collaborative image of the city: mapping the inequality of urban perception. PloS ONE, 8(7), e68400.

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

Sampson, R. J., & Raudenbush, S. W. (2004). Seeing disorder: Neighborhood stigma and the social construction of “broken windows”. Social Psychology Quarterly, 67(4), 319-342.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. Science, 277(5328), 918-924.

Shobe, H., & Banis, D. (2014). Zero graffiti for a beautiful city: the cultural politics of urban space in San Francisco. Urban Geography, 35(4), 586-607.

Spicer, V., Reid, A. A., Ginther, J., Seifi, H., & Dabbaghian, V. (2012). Bars on blocks: A cellular automata model of crime and liquor licensed establishment density. Computers, Environment and Urban Systems, 36(5), 412-422.

Steinbach, R., Perkins, C., Tompson, L., Johnson, S., Armstrong, B., Green, J., Grundy, C., Wilkinson, P., & Edwards, P. (2015). The effect of reduced street lighting on road casualties and crime in England and Wales: controlled interrupted time series analysis. Journal of Epidemiology and Community Health, 69(11), 1118-1124.

Stucky, T. D., & Ottensmann, J. R. (2009). Land use and violent crime. Criminology, 47(4), 1223-1264.

Thompson, S. K., & Gartner, R. (2007). Effective crime prevention in Toronto, Canada. Enhancing urban safety and security: global report on human settlements 2007. Accessed 24 June 2016 from http://unhabitat.org/wp-content/uploads/2008/07/GRHS.2007.CaseStudy.Crime_.Toronto.pdf.

Vanwolleghem, G., Van Dyck, D., Ducheyne, F., De Bourdeaudhuij, I., & Cardon, G. (2014). Assessing the environmental characteristics of cycling routes to school: a study on the reliability and validity of a Google Street View-based audit. International Journal of Health Geographics, 13(1), 1.

Webb, B. and Laycock, G. (1992) Reducing crime on the London underground: An evaluation of three pilot projects, Crime Prevention Unit: Paper no. 30, London: Home Office.

Welsh, B., & Farrington, D. (2008). Effects of improved street lighting on crime: a systematic review. Campbell Systematic Reviews, 4(13).

Wilcox, P., Quisenberry, N., Cabrera, D. T., & Jones, S. (2004). Busy places and broken windows? Toward defining the role of physical structure and process in community crime models. The Sociological Quarterly, 45(2), 185-207.

Wilson, J. Q., & Kelling, G. L. (1982). Broken windows. Atlantic Monthly, 249(3), 29-38.

Wolfe, M. K., & Mennis, J. (2012). Does vegetation encourage or suppress urban crime? Evidence from Philadelphia, PA. Landscape and Urban Planning, 108(2), 112-122.

Chapter 5 Conclusions and future research

This dissertation set out to investigate the social and environmental factors that influence violent crime based on crime hot spot identification and characterization. Together with Chapter 2, 3, and 4, this dissertation provides a comprehensive understanding of hot spots of violent crime from perspectives of both social context and physical built environment. This final chapter summarizes the main contributions of the dissertation and directions for future research.

5.1 Contributions

This dissertation yields several main contributions as follows.

1) A unified modeling framework

Chapter 2 applies a unified modeling framework - seemingly unrelated regression - to assess the stability of model parameters in crime rate analysis, and adds to existing literature by empirically studying the appropriateness of using single-year and multi-year average measure. Compared to ordinary least squares, seemingly unrelated regression is able to provide more efficient and precise estimates of model parameters and standard errors by comprehensively employing all the information in the equation system and considering the interdependent unmeasured causes which can similarly influence violent crime in different years.

2) The appropriateness of averaging crime rates over multiple years

In previous studies, the main criterion for averaging crime data across multiple years is the presence of annual fluctuation, that is, if it is observed by using the difference of means test, crime data of multiple years should be averaged. However, the findings in Chapter 2 reveal that the criteria to support the averaging of crime data should be not only the presence of year-by-year volatility in crime data, but more importantly, the temporal stability of relationships between socio-economic characteristics and crime. If crime processes are stable over a period of time, and the effects of socio-economic correlates on crime of different years are significantly stable, it is unnecessary to average crime over multiple years.

The difference of means test in Chapter 2 finds that crime rates in 1998 (lower), 1999 (lower) and 2001 (higher) are significantly different from the five-year average crime rate. However, the Wald statistic test suggests strong parameter stability between five single years and the average of five years. This implies the importance of taking social indicators into consideration when selecting the form of the dependent variable.

3) Combination of the two ruling spatial theories of crime

Chapter 3 and 4 uses a combination of the two ruling spatial theories of crime - social disorganization theory and routine activity theory – to identify correlates of violent crime. The effects of identified correlates on violent crime are consistent with previous studies in other settings. The findings in these two chapters give evidence that a combination of the two ruling spatial theories of crime provides an applicable framework for understanding the spatial and temporal dimension of violent crime hot spots.

4) Findings support immigration revitalization perspective

The negative effect of immigrant concentration in Chapter 2 and 3 gives credence to the "immigration revitalization" perspective (see Lee and Martinez 2002, p. 365), which argues that "immigration revitalizes poor areas and strengthens social control due to strong familial and neighborhood institutions and enhanced job opportunities associated with enclave economies, and thereby decreases crime" (Lee and Martinez 2002, p. 366). Furthermore, since this is a spatially lagged variable in the present analysis, this suggests that the effect may lead to "diffusion of benefits" beyond the neighborhood. More importantly, this finding extends our understanding concerning this relationship by confirming that the negative effect is not only cross-sectional, but also lasting over longer periods of time.

5) Using cluster membership variable to capture the persistence of hot spots

In Chapter 3, the measure of cluster membership has proved convenient to capture the magnitude of the hot spot persistence over several years. By overlaying several cluster maps, the cluster membership can be obtained by counting the number of times a block group belongs to clusters. The cluster membership variable is regressed against the full set of socio-economic variables by using an ordered probit model. And the model identifies a series of correlates of hot spot persistence during the period.

6) Correlates of longitudinal persistence of violent crime hot spots

The results in Chapter 3 confirms that not only cross-sectional violent crime patterns can be accounted for by cross-sectional socio-economic deprivation and residential mobility, but also the decennial variation of socio-economic deprivation and residential mobility tend to impact the longitudinal persistence of violent crime hot spots. This extends our conventional knowledge regarding the effects of socio-economic deprivation and residential mobility. Strategic insights can be obtained from the findings to inform focused crime reduction policies.

7) A model-based method to select audit sites

In Chapter 4, a model-based site selection method has proved its efficiency in determining target sites for audit. A Poisson regression model with spatial filtering is used to identify socio-economic correlates of violent crime. Parting from the hypothesis that omission of built environmental factors results in systematic residual pattern, target sites for audit can be determined by analyzing residual spatial pattern as retrieved by an eigenvector-based spatial filter.

8) Using Google Street View to identify the environmental correlates of violent crime

Chapter 4 uses Google Street View to systematically identify physical features associated to violent crime, after socio-economic correlates have been controlled for in the Poisson regression model with spatial filtering. The procedure is fairly general, and can be applied to other cases where on-site audits are challenging or unfeasible. A desktop data collection tool is developed to facilitate virtual audit data collection. With some modification, the tool can be retooled for multi-purpose environmental audits. The GSV-based virtual audit has proved effective in cases of large-scale environmental assessment and dispersed distribution of study sites.

This study provides empirical test of GSV for the purpose of virtually identifying environmental correlates of violent crime. We find that GSV-based audit can reduce both human cost and travel expense, protect auditors from potential risks, and allow quantitative environmental data to be collected in a systematic way.

9) Results can help to inform crime prevention efforts

The results obtained from seemingly unrelated regression in Chapter 2, ordered probit regression in Chapter 3, and Poisson regression with spatial filtering in Chapter 4 can inform crime prevention policies. The environmental factors identified from GSV-based audit provide insights for CPTED initiatives, situational crime prevention efforts, and urban renewal projects. Police portal should be encouraged for neighborhoods presenting public signs of physical disorder.

5.2 Future research

It bears noting that the findings in this dissertation are limited to violent crime. Hence, several natural avenues for future research can be discerned.

1) Replicating present analyses in other types of crime

The socio-economic and environmental correlated identified in this thesis is limited to violent crime at the aggregate level. More specifically, the temporal stability of model parameters found in Chapter 2 is limited to violent crime rate. The temporal persistence of violent crime hot spots as well as its correlates found in Chapter 3 is limited to violent crime hot spots. The environmental correlates identified in Chapter 4 are limited to violent crime count as well. Although there are similarities between correlates of diverse types of crime, different types of crime have diverse ecological and individual causes (Felson, 2014). It is reasonable to assume that socio-economic correlates of robbery are different from correlates of murder, and crime opportunities for forcible rape can also different from opportunities for aggravated assault. Therefore, with the proposed methods in this thesis, future research needs to substantiate these findings in the case of disaggregated types of violent crime and property crime.

Given that there are complex mechanisms associated with crime, future study also needs to account for the interaction effects between diverse socio-economic and environmental correlates of crime when conducting regression analysis.

2) Alternative measures of population at risk

The census residential population is used as the population at risk when calculating crime rates in Chapter 2. Although this is a common practice, recent studies also suggest alternative measures of the population at risk such as the ambient population (Andresen, 2006a). It was found that the use of ambient population has an impact on the results of crime analysis (Andresen, 2011), although Harries (1991) suggested that this improvement is most often not worth the costs in both time and money.

It bears noting that ambient population can only capture population at risk for some types of crime (Andresen, 2011), and alternative measure of population at risk may be required only by some types of crime when calculating crime rates. For crime rates of some property crimes, census residential population is able to capture the population at risk. It may be appropriate to use ambient population for violent crime, because violent crime is referred to as crime against persons and requires at least the presence of offenders and victims.

3) Replicating present analyses using finer-scale spatial unit

This dissertation used census block group as spatial unit, because it is the finest scale for which the U.S. Census Bureau tabulates and publishes sample data. In cases that census data is not used in the analysis, other finer-scale spatial unit (e.g., census block, face block, and street segment) can be used.

In terms of the choice of spatial unit, although some found that it has no impact on the result of crime analysis, see for example, Ouimet (2000), Wooldredge (2002), and no impact on “the power of the statistical tests performed” (Ouimet, 2000, p. 150), others suggested the use of smaller spatial units (Andresen, 2011). The use of small spatial unit provides stronger coefficients than the larger spatial unit (Ouimet, 2000). Future research can replicate present analyses using finer-scale spatial unit and test whether finer-scale spatial unit improves analysis.

4) Hot spot persistence at different time scales

The findings regarding persistence of crime hot spots in Chapter 2 are limited to the scale of nine years. Future research need to explore hot spot persistence at different time scales (particularly longer time periods). Future research also need to examine the persistence of other disaggregated types of violent crime, identify neighborhood structural predictors that influence the persistence of individual types of crime, in order to compare the consistency and difference between aggregated and individual types of violent crime.

5) Using Google Street View to study urban environment

The availability of historical GSV provides opportunities for future research such as time series study of built environment, comparative study of present landscape with historical environment. GSV Imageries of different time nodes (e.g., August 2007, May 2009, June 2011, June 2012, June 2014, August 2015, and November 2015) are comparable. Similar to time series analysis of optical remote sensing data, environmental characteristics can also be observed from historical GSV imageries and can contribute to study the temporal dynamic variation of urban landscape.

What’s more, a combination of virtual audit and on-site audit can be used in future research. One can take a mobile device and walk into the field to conduct virtual audit and on-site observation simultaneously. The present urban landscapes can be directly compared with historical environment.

References

This main reference only lists references from Chapter 1 (Introduction) and Chapter 5 (Conclusions and future research). Other references remain self-contained within each chapter.

Dustmann, C., & Fasani, F. (2016). The effect of local area crime on mental health. The Economic Journal, 126(593), 978-1017

Elliott, D. S., & Ageton, S. S. (1980). Reconciling race and class differences in self-reported and official estimates of delinquency. American Sociological Review, 45, 95-110.

Fajnzylber, P., Lederman, D., & Loayza, N. (2002). What causes violent crime? European Economic Review, 46(7), 1323-1357.

Felson, R. B. (2014). Back to basics: Gender and the social psychology of aggression. In The Oxford Handbook of Gender, Sex, and Crime, edited by R. Gartner and B. McCarthy. New York: Oxford University Press.

Mair, J. S., & Mair, M. (2003). Violence prevention and control through environmental modifications. Annual Review of Public Health, 24(1), 209-225.

Sparks, C. S. (2011). Violent crime in San Antonio, Texas: An application of spatial epidemiological methods. Spatial and Spatio-temporal Epidemiology, 2(4), 301-309.