SPATIAL AND STRUCTURAL DETERMINANTS OF RESIDENTIAL BURGLARY RATES IN KITCHENER-WATERLOO

By

Martin Anstey
Wilfrid Laurier University

Bachelor of Arts
University of British Columbia, 1995

THESIS
Submitted to the Department of Geography and Environmental Studies
in partial fulfilment of the requirements
for the Master of Environmental Studies degree
Wilfrid Laurier University
1997

0-612-24372-9

ABSTRACT

Traditional ecological analyses of burglary have a number of shortcomings. First, break-ins occurring within a neighbourhood have been regarded as a solely the product of local residents. Second, the relationship between burglary and its determinants has been considered constant across space. Third, the data used in ecological analysis has been treated as statistically independent. The purpose of this study is to expand upon previous ecological studies through an analysis of residential burglary in Kitchener-Waterloo, Canada. Theory from environmental criminology (a discipline concerned with the spatial aspects of crime) is used in conjunction with spatial data analysis techniques and GIS technology to model the variation of breaking and entering across the city's enumeration areas. The results suggest a heterogeneous spatial structure underlying the determinants of burglary rates. Within a four kilometre radius of the CBD, the crime is very localized; its prevalence in a neighbourhood best predicted by the offender characteristics of the local area. Within the area located beyond this inner zone, significant predictor variables relate to the attractiveness of an EA in terms of opportunities, suggesting that the crime is committed by individuals journeying from other areas. Throughout the entire study area higher population densities were consistently found to be the greatest deterrent to burglary rates recorded at this scale.

Table of Contents

List of Figures

List of Tables

Appendices

CHAPTER 1: INTRODUCTION

1.1 Introduction

Burglary commonly occurs throughout all parts of the city, yet some areas are more predisposed to it than others. Previous studies have shown that the extent to which a neighbourhood experiences this crime is in part a function of the kinds of people who live there and the type of built environment in which they reside. In recent years the relative location of a neighbourhood within the city has also been found to be a key determinant of its burglary rate.

An exploration of the causes underlying the distribution of neighbourhood burglary rates can benefit from the application of three newly emerging areas related to geography. The first is environmental criminology (EC), a hybrid discipline of geography and criminology whose purpose is to construct *theory* on the spatial aspects of criminal activity. The second area, spatial data analysis (SDA), is the field concerned with the analytical *techniques* applicable to data referenced to a location in space, such as burglary rates recorded for city neighbourhoods. The third area of interest is geographical information systems (GIS), a *technology* used to store, analyze and display spatial data such as aggregate crime data.

EC, SDA and GIS each have clear contributions to the analysis of crime data sets. Yet there has been little development in their integration as a complete methodological approach to the exploration and modelling of the intraurban distribution of criminal activity. This is particularly unusual in light of the widespread introduction of GIS technology to crime research, and the efforts made to integrate GIS and SDA within the wider field of geography.

The missing link of using SDA methods to explore and model crime data is part of a discipline-wide neglect of the special nature of geographic data. The unique characteristics of spatial data sets cause certain fundamental assumptions of many statistical techniques to be violated. This is a serious technical issue which raises questions about the validity of conclusions drawn from non-spatial areal studies of any kind. Consequently, the spatial analytic tools available to effectively model neighbourhood crime data have been largely ignored.

In addition to the statistical nuances of spatial data, previous research has tended to overlook the substantive spatial processes which account for the intraurban distribution of crime. The city has been viewed as a homogeneous entity across which the relationships among the key players in a criminal event remain the same. The possible existence of discrete subareas, within which the interaction of potential offenders and potential victims is significantly different, has been largely ignored by researchers. Of greater consequence has been the false assumption that the crime rate of a particular neighbourhood is solely a product of the people living within its geographic boundary.

The purpose of this research effort is to bring together the current theoretical knowledge of burglary from the field of environmental criminology, SDA techniques and GIS technology in order to effectively analyze burglary in Kitchener-Waterloo. This study focuses solely upon residential burglary as it is less geographically restricted than its commercial or institutional counterparts. It is also more common and more serious.

The specific intent of this study is to incorporate and evaluate the role of space in the intraurban distribution of this crime in Kitchener-Waterloo. This is to be done by paying particular attention to the role of space (in both a statistical and substantive sense) in the distribution of this particular form of human activity. In addition to constructing a statistical model, a second objective is to delineate relatively high risk areas within the city. It is hoped that the identification of these areas will assist the local police and planners in allocating resources to deal with this crime; and also possibly assist future researchers to select areas for a more locally-sensitive case study analysis.

1.2 Research Questions

The central purpose of this research is to answer the following question: What social and physical factors account for the intraurban variation in residential burglary rates in Kitchener-Waterloo at the enumeration area scale when space is explicitly incorporated into the analysis? This analysis will be quantitative by nature, so the main product will be a model explaining the spatial variation of residential burglary rates in terms of covariates referenced to a common spatial unit, as well as in terms of the spatial arrangement of these units. In order to answer this question, a series of component questions were formulated:

1. Which social and physical characteristics are significant predictors of residential burglary rates across Kitchener-Waterloo?
2. How significant is the role of spatial structure in the distribution of residential burglary in the study area?
3. How can GIS technology best contribute to such an analysis?

1.3 Research Objectives

Specific research objectives were determined in order to answer the research questions. They consist of the following:

1. Create a database of variables suggested as relevant by the literature using GIS data layers and census information, as well as the data set provided by the police.Explore the data set to understand its statistical and spatial patterns, as well as the interrelationships of the explanatory variables.
2. Construct a model of the factors associated with the intraurban variation of burglary rates in Kitchener-Waterloo which explicitly incorporates the spatial effects present within it.

3. Interpret the findings of the analysis. Suggest the implications for policing and planning, as well as any future research which could improve upon this study.

1.4 Thesis Organization

Chapter Two is a literature review of the two component topics of this study: the crime of residential burglary and the spatial analysis of aggregate data. It begins with a general introduction to residential burglary, outlining its significance to Canadian society and the need for research into this crime. This is followed by a summary of the previous research on the determinants of burglary's intraurban distribution; specifically, the characteristics of offenders, opportunities and their interaction over space. The final section outlines the need for, and describes the specifics of, the spatial analytical techniques and technology available to explore and model data such as burglary rates recorded for areal units.

Chapter Three outlines the creation of the conceptual framework for the subsequent analysis. It includes the methods of data collection used to translate the knowledge gained from the literature review into a modelling environment. It also summarizes some important issues related to the data which had to be addressed before the analysis could proceed.

Chapter Four is a narration of the analysis. It begins with simple exploratory techniques and proceeds to advanced spatial model construction. Included in this chapter is a review of some of the regression models available to analyze aggregate spatial data. It also contains an interpretation of the results of the analysis. This includes a review of the important variables both individually and acting in concert. The role of spatial structure is also discussed as a relevant factor in this study.

Chapter Five summarizes the research effort. It offers recommendations based upon these findings to the police, local planners and researchers of criminal activity. This chapter also includes a summary of the limitations of this study.

CHAPTER 2: LITERATURE REVIEW

2.1: Residential Burglary

> "A burglar is by common law a felon that, in the night, breaketh and entereth into a mansion house of another, of intent to kill some reasonable creature, or to commit some felony within the same, whether his felonious intent be executed or not."
> (Coke, 1797, vol. 3: 63 as quoted in Maguire and Bennet, 1982: 7)

Little has changed in the definition of burglary since this early attempt to codify it into English law. The origins of the word are suggested as being a compound of "burgh" (house) and "laron" (thief). "Burglary" itself was predated by the Anglo-Saxon crime of "hamsocn", defined by Hume as "the felonious seeking and invasion of a man in his dwelling-place" (Maguire & Bennet, 1982: 7). The Canadian *Criminal Code* defines "Breaking and Entering" (B&E) as the crime which occurs when a dwelling or other premise is illegally entered by a person who intends to commit an indictable offense (Chard, 1995: 2). Both burglary and B&E thus describe an identical crime, and will be used interchangeably throughout this study.

The emphasis of the various definitions of this crime is, and always has been, on the act of illegally gaining entry to a premise. Burglary to a *residence* is considered a particularly serious offense not so much due to the victim's loss of property than from the "breaking" of the sanctity of their dwelling place. This destroys the sense of security they associate with their home (Evans, 1989: 88). As a result, burglary to a dwelling is widely perceived as "mala in se", or intrinsically evil, and so, unlike many other infractions of the law, transcends the moral order of a particular place and time (Bursick & Gramswick, 1993: 19).

Although the secondary offenses committed by the offender once inside the home –typically vandalism and/or theft– are not considered serious crimes in and of themselves, they are costly.

According to the Insurance Bureau of Canada, losses due to thefts occurring in residential burglary incidents in 1993 amounted to over $373 million in insurance claims, with over a third of victimized households losing $1000 or more (Chard, 1995: 10). In addition to items being stolen, it is typical (71% of incidents) that property is damaged during the commission of this crime (Chard, 1995: 1).

While the police and public regard B&E as a serious crime (Evans, 1989: 92), the punishment for convicted burglars provided by the courts is relatively light, particularly for first time offenders. The maximum sentence for B&E in Canada is life in prison, yet offenders are rarely punished so severely. In 1994 only 40% of convicted juvenile burglars were sent to custodial service (average sentence of three months), while 45% were put on probation (average length of one year). For convicted adult burglars the punishment was more severe with 61% handed a prison sentence and 37% given probation (Chard, 1995: 14-16).

A second mismatch between the perceived seriousness of the crime and the societal response is evident in its under reporting, which is common among victims of this crime. One third of burglary victims do not report the incident to the police. Reasons cited are usually insufficient losses to warrant a call, little faith the stolen property would be found or the offender caught (Evans, 1989: 89). There are some grounds for this perception as in Canada only 16% of the victims who report a burglary incident to the police ever recover any stolen property, and in only 17% of reported incidents is a suspect ever charged by the police (Chard, 1995: 11-12).

After a rapid growth from the mid 1960s to the mid 1980s, rates of residential burglary have stabilized and even begun to decline over the last 10 years. However, breaking and entering remains a serious problem in Canada, which ranked third for this crime in a recent international survey (Chard, 1995: 6). It is a psychologically damaging experience for the victim. It is expensive to society. Its offenders are rarely caught or property recovered. It is important to study this crime to find out how it occurs in order to better allocate resources to prevent it. The focus of this thesis now turns to one approach to performing such research: environmental criminology.

2.2 Environmental Criminology

The partnership of geography and criminology dates back over 150 years, when crime was first revealed to exhibit an uneven distribution across the regions of England. The field of cartographic criminology began a tradition of studying the spatial distribution of crime (Herbert, 1989: 1). Simple comparisons were made between crime rates and the social and economic characteristics of places.

The Chicago School of Social Ecology of the 1920s and 1930s introduced quantitative methods to the analysis of the city. The sociologists Shaw and McKay explored ecological approaches to urban crime in particular. Aggregate social and demographic variables were examined in order to discover the correlates of delinquent activity. Neighbourhood crime rates and the location of offender residences were also mapped to describe the spatial pattern of delinquent behaviour. The Chicago school had a profound impact on the way crime was studied, providing a great deal of insight on its association with a number of social and economic factors. Yet this approach failed to produce an exhaustive explanation of crime's intraurban distribution.

Critics of the ecological school argued that this approach was incapable of producing any causal inferences due to the ecological fallacy (Davidson, 1993: 2). Conclusions drawn from an aggregate analysis were criticized as being inapplicable to individuals. However, the findings of the Chicago School were important "...albeit in subtle and more variable ways than the classical ecologists anticipated in the search for a general theory to explain the distribution of crime" (Davidson, 1981: 89).

One subtlety which would later emerge was the geographic context of crime. The ecologists believed that differences in crime rates were a product of the ability of local communities to regulate and control the behaviour of their residents. But this mistakenly assumed that crime is a product of forces operating at a neighbourhood scale (Bursick and Grasmick, 1993: 24). The influence of nearby neighbourhoods on a given area's crime rate was largely ignored.

During the 1950s ecological studies were technically enhanced through the introduction of factor analysis. The purpose of factoral studies was to define the underlying socioeconomic dimensions of the city and relate them to crime rates. Social area analysis pioneered by Shevky and Bell employed such methods to study social differentiation rather than the spatial differentiation of the Chicago school (Davidson, 1981: 73). They created constructs representing social structure (social rank, urbanization and segregation) and related these to crime rates (Shevky and Bell, 1955).

Although these multivariate techniques were computationally more sophisticated than their ecological predecessors, crime was still addressed essentially as a non-spatial phenomena. The results of aggregate analysis remained relatively unsubstantial and controversial. After these early sociological attempts, the analysis of crime rates undertook new methodological and conceptual positions (Herbert, 1989: 1).

One new approach developed in the 1970s was the discipline of environmental criminology, founded by the criminologists Brantingham and Brantingham. They contended that any crime consists of four dimensions: the law, the offender, the target, and the place. Environmental criminology is the study of place, "the discrete location in time and space at which the other three dimensions intersect and a criminal event occurs" (Brantingham & Brantingham, 1981a: 8). This perspective incorporated theory from many diverse fields, but its emphasis on location was particularly suited to benefit from previous research in geography.

There are three levels of analysis at which environmental criminology is studied. At one extreme is macro-analysis, where the units of interest are very large such as states, provinces, regions and countries. An example is the study of the violent subculture of the southeastern United States and its influence on the prevalence of homicide in that region. At the other extreme is micro analysis which looks at specific crime sites; for example, research to determine why corner lot houses have a propensity to be burglarized. In between these two extremes is meso-analysis where geographically small aggregates (census tracts, police precincts, etc.) are the unit of interest. Meso-scale environmental criminology differs from previous aggregate analysis in its focus on the social, physical *and* locational characteristics of areas which make them more or less conducive to criminal activity.

This focus on place and space has received ever-increasing attention by researchers during the last twenty years, resulting in a great deal of new theory. One approach, for example, involves the study of offender travel behaviour through so-called "journey to crime" studies. Such spatial behaviour has been found to be of fundamental importance to the distribution of crime, suggesting that its neglect by traditional ecological studies was a serious theoretical flaw.

Research in meso-scale environmental criminology has, on occasion, incorporated explicitly geographic techniques such as spatial autocorrelation measures to improve upon the explanatory power of traditional ecological variables (for example, Brown, 1982; Costanzo et al, 1986). However, such attempts have limited themselves to description rather than the full exploration and exploitation of the spatial structure inherent in crime data sets. The full use of the spatial analytical methods has yet to applied to the analysis of crime. As a result, the body of knowledge gained from research in meso-scale environmental criminology has yet to be fully capitalized upon.

During the same period that environmental criminology emerged as a practical means to analyze crime patterns, there was a surge in the use of computing technology by law enforcement agencies. One of these new technologies, Geographic Information Systems, have now been introduced to police departments "everywhere" (Christie and Shields, 1996: 28). There has also been a vast increase in the number of police databases which are inherently spatial because they are addressed-referenced.

This rapid increase in geographical information and information technology has occurred within a number of fields. Consequently, there has been a growing interest in the analytical methods available to explore and model spatial data, and on the means to integrate such methods with GIS technology (Anselin and Getis, 1992). An interesting example of this integration within the field of environmental criminology is a "geographic profiling" program developed by Dr. Kim Rossmo at Simon Fraser University. This software combines data on the location of individual crime sites with journey-to crime theory to create a probability surface of the location of the offender's home or workplace. The resulting computer map is used to narrow the search area for the investigating officers. This program is remarkably accurate at deducing the whereabouts of criminals because geographic behaviour is so predictable and so central to the criminal event (Grescoe, 1996: 50).

Environmental criminology provides an excellent theoretical framework for the new techniques and technologies applicable to the analysis of spatially referenced crime data; moreover, it "provides a useful framework for geographers and a link with the wider field of study [of criminology]" (Herbert, 1989: 4). At the aggregate scale of analysis it has assisted in the incorporation of spatial dimension missing in traditional ecological studies, although this area still requires development. The purpose of this thesis is to further this area of research.

2.3 Determinants of Residential Burglary

The burglar can be considered as a special type of consumer who, like all consumers, acts in a relatively rational manner in face of a set of constraints. The purpose of this section is to identify these constraints and the individual burglar's considerations in dealing with them. Having described the offender-victim system at a macro and micro scale, the literature review then reports on the factors relevant to an analysis at the meso-scale: the characteristics of offender areas, opportunity areas, and the role of space in their interaction.

Hagerstrand identifies three types of constraints which restrict the activities of all people: authority constraints, capability constraints and coupling constraints (Hagerstrand, 1970). An adaptation of these three to criminal activity is provided by Rengert. He interprets authority constraints to mean the power relationships within our society. The offender is constrained by the police and courts who have the authority to arrest and detain them. The ultimate authority constraint is imprisonment, where the burglar is spatially constrained from committing further break-ins (Rengert, 1989: 170).

Secondly, the offender is limited by capability constraints. These include the limitations on their movement over space due to transportation costs and time limitations. The burglar, like all consumers, has a finite amount of time and money available to search for the best "product".

The third constraint encountered by the offender is the coupling constraint. In order for a successful break-in to occur, the skill and motivation of the burglar has to be coupled with an ideal set of circumstances. They first have to find an unoccupied household and then be able to break-in without being noticed. If they are given any indication that this is not the case, they generally are forced to abandon their attempts at committing this crime (Poyner, 1983: 34).

Acting within this system of constraints, the individual burglar searches for a specific set of circumstances in order to maximize their returns. Interviews with convicted burglars indicate that both the neighbourhood and the particular dwelling within it provide certain cues to an offender motivated to commit this crime. These are the perceived risk, reward, ease of entry and distance to travel to the site (Bennett, 1989: 190). Clearly these relate to Hagerstrand's constraints: the authority constraint creates the perceived risk; the coupling constraint relates to the ease of entry; and the capability constraint to the distance to travel. The reward consideration can be seen as the motivation to overcome these constraints (see Figure 2.1).

At an intermediate scale of analysis, the determinants of intraurban residential burglary rates are a complex fusion of these macro and micro dynamics. Research into these factors has been the purpose of a vast amount of literature from a variety of disciplines. A means to organize this material is provided by Rengert who classified the determinants of crime's spatial distribution into three groups:

(1) The location of crime-prone populations;
(2) the location of opportunities; and
(3) the relative accessibility of potential offenders to opportunities (as quoted in Brown, 1982: 248).

This is a useful taxonomy, providing the framework for the remainder of the literature review as well as the analysis. However, there are several issues that have to be addressed if Rengert's classification is to be used.

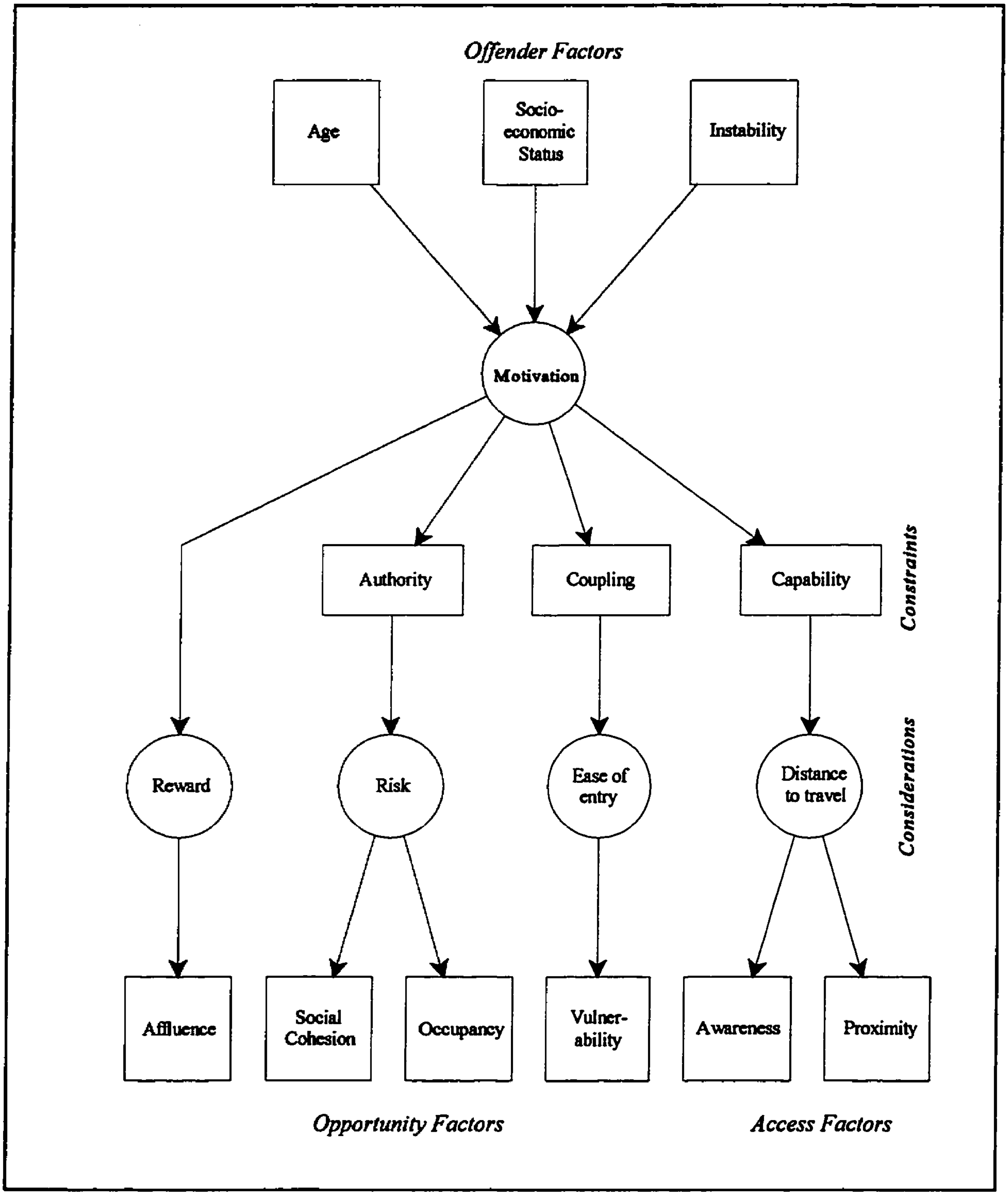


Figure 2.1 Conceptual Model of the Burglary Event

First it is necessary to identify what precisely constitutes a "crime-prone population", an "opportunity", and its "relative accessibility". In doing so, it is important to recognize that at an aggregate scale many of the factors associated with burglary are interdependent (Bursick, Jr. and Grasmick, 1993: 60), and as such, not easily assigned exclusively to any one of these groups. Often what makes a neighbourhood an attractive target also makes it less criminogenic.

Secondly, in applying Rengert's simple taxonomy to an aggregate analysis, it must be kept in mind that the causes of burglary are entirely scale-dependent; what may be relevant to burglary aggregated to enumeration areas may not be relevant to census tracts. An areal unit which appears to be homogeneous may contain a variety of sub-units –particularly if the original unit was created for administrative purposes (Brantingham and Brantingham, 1981: 22).

The remainder of this section summarizes the determinants of residential burglary. The first topic, offender factors, examines the causes of the necessary precondition for a burglary event: the motivated offender. These are the individuals who are willing to defy authority constraints. The second topic is an exploration of what constitutes "opportunities". These relate to Hagerstrand's authority and coupling constraints, and consequently to the risk, reward, and ease of entry considerations of the individual burglar. The final topic "relative accessibility" examines the spatial component of this crime associated with Hagerstrand's capability constraint.

2.3.1 Offender Factors

A willingness to break-in to someone else's home is only present in a particular subset of the population. As types of people tend to cluster together in cities, "the motivation to commit crimes is unevenly distributed across space" (Brown, 1982: 248). This section outlines the most common socio-demographic characteristics associated with the people who commit residential burglary in an attempt to determine the location of potential offenders within the city.

As with most crimes, burglary is largely committed by males. In Canada they constitute over 96% of those accused of breaking and entering (Hendrick, 1995: 9). In this study gender is not considered an offender factor in and of itself; however, when attempting to estimate the number of potential offenders based upon age and income, only the figures for males were collected.

Age

Burglary is predominantly a crime committed by young men. Over 80 per cent of those charged with breaking and entering in Canada are under the age of 26. Young offenders (those under 18) are particularly prevalent in this crime, comprising almost a third of those charged with this offense (see Table 2.1). Other studies cite juveniles as comprising an even greater proportion of the burglar population. For example, an analysis of Metropolitan Toronto found that when focusing exclusively on residential break-ins (not commercial or institutional), juveniles comprised half of the known offenders (Waller and Okihiro, 1978: 23).

It should be noted, however, that younger burglars tend to operate closer to home and are less sophisticated in their commission of the crime, making them more likely to be caught than older, more experienced felons (Maguire & Bennet, 1982: 22).The same study by Waller and Okihiro also noted that many persons in adult correctional facilities admitted to having previously broken into homes, although very few of them were incarcerated for this crime (Waller and Okihiro, 1978: 3).The "dark figure" of undetected offenders may overemphasize the youthful and localized nature of this crime (Maguire & Bennet, 1982: 22).

Table 2.1 Total Accused of Breaking and Entering in Canada by Age, 1994 (from Canadian Crime Statistics, 1994).

Age Group	% of Charged
12 - 17	32.5
18 - 25	48.0
26 +	19.5

Socioeconomic status

Burglary is predominantly a crime committed by lower socioeconomic status (SES) people (Davidson, 1981: 59). This was found to be the case for England, America, Canada, Australia and New Zealand –the countries containing all of the study areas found in the literature review. For many, being economically disadvantaged provides the necessary motivation to commit this crime. There are a number of SES-related variables: housing conditions, crowding and income being the most commonly used. In this study, these variables are considered as facets of the same problem experienced by the poor.

A related factor, ethnicity, is particulary important to understanding the distribution of burglary in American cities, where the racial composition of a neighbourhood is the dominating predictor of burglary rates (Byrne and Sampson, 1986a: 4). Yet, even the early pioneers of ecological studies were quick to point out that a neighbourhood's racial composition is inextricably correlated with other socioeconomic variables, making it difficult to isolate its specific contribution (Bursick and Grasmick, 1993: 26). As a result of this limitation, combined with Kitchener-Waterloo's lack of any extensive ethnic areas, this factor was disregarded from the subsequent analysis.

Instability

A third offender factor is related to instability. This is largely associated with the family status of the offender. For younger offenders, the family unit within which they belong influences their motivation to commit crimes such as burglary. Families experiencing alcoholism, drug abuse and/or domestic violence are more likely to produce juvenile delinquents, although data limitations prevent the incorporation of these characteristics into an aggregate analysis. Areas with a large proportion of one-parent families also tend to produce a disproportionate amount of young criminals (Dunn as quoted in Brown, 1982: 248), and this variable can be analyzed in aggregate form. Previously it was noted that juvenile delinquents comprise a significant portion of the offender group; however, only 6% of young men in Canada are prone to criminal behaviour, and many of these come from "poor and disorderly communities" (The Globe and Mail, December 17, 1996).

For adult offenders, the instability factor relates to the extent of their personal attachments. Neighbourhoods with high burglary rates frequently have a greater proportion of lodgers, single men and recent movers (Waller & Okihiro, 1978: 51). Whether instability serves to attract burglars because of the anonymity provided by this environment, or reflects a group of people which is more criminogenic is debatable. It is likely that instability acts as both an offender and opportunity factor.

2.3.2 Opportunity Factors

Opportunity factors relate to the highly complex decision-making preceding the perpetration of a break-in event and so are more elusive to quantification than offender factors which deal simply with the motivation to commit this crime. This qualitativeness makes them particularly difficult to delineate from one another at an aggregate scale. However, opportunity factors are widely recognized as more significant predictors of neighbourhood burglary rates than offender factors (Waller & Okihiro, 1978: 53). Consequently, the bulk of current research is focused on the environments in which criminal activity occurs rather than the environments from which criminals emerge (Herbert, 1989: 8).

Social cohesion

Community residents who are connected together through social ties act to prevent criminal activity by maintaining an informal surveillance over their local area (Bursick and Grasmick, 1993: 66). Such social cohesion is a particularly effective deterrent to the burglar, a criminal who avoids the risk of being identified as a stranger (Bennett, 1989: 181).

Social cohesion is a function of other neighbourhood characteristics such as rates of home ownership, family status, migration rates and housing stock characteristics. As a result it is often difficult to isolate the impact of this factor from others. There is also a strong assertion that social cohesion is in fact an offender factor as "...crime is less likely to occur in neighbourhoods that are socially cohesive and stable, as such areas produce few offenders" (Waller and Okihiro, 1978: 51). Thus it is difficult to separate the social cohesion opportunity factor from the instability offender factor. The distinction made in this study is that instability relates primarily to the relationships occurring within the household while social cohesion relates to the relationships between households.

Occupancy

At the scale of the individual dwelling, signs of occupancy of the target or of its neighbours is the deterrent most frequently cited by burglars as it clearly makes the crime much more risky

(Poyner, 1983: 34). The amount of time a dwelling is left unoccupied is partly a function of the household members' socio-demographic characteristics. The majority of people are required to spend a significant and predictable amount of time away from their home, either to be at work or school. Occupancy can thus be analyzed in an aggregate form as "...one would expect the rates of criminal victimization to be highest in those neighbourhoods in which a relatively significant portion of the population has predictable, non-discretionary time blocks outside of the home, thereby creating an absence of capable guardians" (Bursick and Grasmick, 1993: 69).

Occupancy levels are also related to the average number of persons per household. Two-parent families and households with large families are less likely to be burglarized than other types of dwelling places (Waller and Okihiro, 1978: 54). Conversely, one person households have been found to experience higher victimization rates than others (Sampson, 1986: 25).

Clearly neighbourhoods of small households and/or households with frequently absent residents are more attractive to burglars than ones which maintain a relatively large daytime population. Again, this variable's association with neighbourhood social and demographic characteristics inextricably links it to other determinants of residential burglary (Waller and Okihiro, 1978: 51).

Affluence

There is a great deal of conflicting evidence regarding the role of affluence as an opportunity factor. A significant amount of research indicates that the poorest neighbourhoods experience the highest burglary rates (for example, Evans and Oulds, 1984; Brantingham and Brantingham, 1981b;

Davidson, 1980). In these studies, break-ins in low income areas were generally found to be committed by local residents.

However, there is a great deal of evidence indicating that affluent neighbourhoods also experience high burglary rates. Waller & Okihiro found that Toronto's two types of high risk areas were poor inner city neighbourhoods and affluent areas within the suburbs (Waller and Okihiro, 1978). This resulted in a distance decay pattern from downtown Toronto with the exception of some local "peaks" in the more affluent suburbs. An earlier study of Boston revealed a similar spatial pattern (Repetto, 1974).

A British study found that high income and low income areas are more likely to be burglarized than middle income ones –particularly high income areas located near low income areas (Maguire & Bennet, 1982: 20). This suggests that the importance of a neighbourhood's level of affluence as an attractor to offenders is largely dependent upon its relative location within the city. Thus, an analysis of this variable's role as a determinant of residential burglary must incorporate a spatial context in order to provide meaningful results. This topic is further discussed in section 2.3.3.

Vulnerability

Research findings show that certain types of cityscape are more conducive to criminal activity than others (Harries, 1980: 92). This vulnerability is associated with land use factors, either in the form of the immediate physical environment or at the neighbourhood scale (Harries, 1980: 93). Research on vulnerability to burglary is primarily done at the scale of the individual dwelling. Micro-scale environmental criminological analysis compares the characteristics of burglarized households with non-burglarized households using variables which reflect the four considerations of the individual burglar outlined on page 13 (for example, Bennett, 1989).

The pioneer work of this type of research is Newman's 'defensible space' theory on the criminogenic characteristics of the physical environment. Newman was motivated by the rapid rise in crime in the developed world which began in the mid-1960s:

> "Within the present atmosphere of pervasive crime and ineffectual authority, the only effective measure for assuring a safe living environment is community control. We are advocating a program for the restructuring of residential developments in our cities to facilitate their control by the people who inhabit them" (Newman, 1972: 204)

His key argument was that architectural design could be used to improve the residents' control of their environment, such as through the construction of barriers (both real and symbolic) to outsiders and also through giving greater consideration to surveillance during the design process (Newman, 1972).

Defensible space theory has been criticized as "exaggerated and over extrapolated" (Harries, 1980: 105); that compared to the fundamental causes of crime such as poverty and racism "...the issue of micro-environmental manipulation to effect behaviour modification seems trivial and irrelevant" (Harries, 1980: 103). However, there is a great body of research indicating that vulnerability does influence criminal behaviour –particularly the burglar's. Specifically, for the individual burglar, "...locational attributes influence perceptions of risk and rewards of crime" (Davidson, 1993: 7). Low vulnerability areas may present a number of coupling constraints, steering motivated offenders away from one neighbourhood towards another. Therefore, a reduction of vulnerability in a given area may not reduce the city's crime rate, but it will displace it from one neighbourhood to another.

One dimension of vulnerability at the aggregate scale is the neighbourhood's landuse. This has been found to have a significant influence on neighbourhood burglary rates. One study found that exclusively residential areas report much lower burglary rates than those bordering or containing non-residential built-up areas (Maguire and Bennett, 1982: 39). Furthermore, the deeper one travels into residential areas, the lower the preponderance of this crime. A related study of the influence of street networks on burglary found that the mean rate for the border blocks of residential enclaves was four times that of interior blocks (Harries, 1980: 96).

Research has also found a preponderance of offenses near main roads (Maguire & Bennet, 1982: 166). This may be because dwellings located near main roads are not protected within the confines of a residential enclave. They are also more likely known to potential offenders, a topic which is further discussed in section 2.3.3.

A second dimension to vulnerability, the density of the built environment, has long been debated by researchers. One perspective is that high density serves to increase crime rates (Harries, 1980: 81). For example, an analysis of Toronto census tracts found that high density areas experienced substantially higher burglary rates than areas of low density, single detached housing (Waller & Okihiro, 1978: 16).

Other theorists such as Jacobs argue that density instead serves to reduce crime because it increases the informal surveillance occurring within an area (Jacobs, 1961). Evidence to support this notion is provided by Maguire and Bennet who found that single detached housing was particularly hard hit by burglars (Maguire & Bennet, 1982: 22).

The density controversy may be indicative of a dependency upon the scale of analysis. For example, a high rise may experience low burglary rates because it is difficult to gain access to aboveground apartments; yet a high density district of the city may contain a greater degree of poverty and, as a result, crime. Waller and Okihiro found that when they compared apartment burglary rates with house burglary rates for Toronto's census tracts there was very little correlation (Waller and Okihiro, 1978: 16). Furthermore, they found apartment burglary rates were different from house burglary rates in that they were best predicted by local offender variables rather than local opportunity variables (Waller and Okihiro, 1978: 52). This suggests that an apartment burglary is likely to be perpetrated by an offender residing either within or very near it, while a house burglary is likely the act of an individual operating at a wider spatial scale.

Another reason for this controversy may be that it is in fact medium density dwellings which are most at risk. Statistics compiled for Canada show that semi-detached houses, row houses and duplexes experience the highest burglary rates, while single detached and high density housing experience significantly lower rates (Hendrick, 1995: 9). Yet this same agency attributes Canada's high ranking as third in the world for burglary (behind Australia and the United States) to the preponderance of low density development in Canadian cities, which is not conducive to surveillance and security (Chard, 1995: 6). An additional objective of this study is thus to identify the role of density in this study area at this particular scale of analysis.

Other Opportunity Factors

There were four other opportunity factors identified in the literature: police presence, community watch programs, neighbourhood reputation and proximity to offender areas. The latter was incorporated into this study, but its spatial aspects warranted a separate section. The rationale for the exclusion of the first three is now described.

The degree of police presence would be very difficult to incorporate as an opportunity factor related to Hagerstrand's authority constraint. Specifically, it would be difficult to interpret because an increase in the local crime rate tends to result in increased police effort within the community (Hakim as quoted in Brown, 1982: 248). Thus it is anticipated that the local police patrol high burglary risk areas of the city more frequently than low risk areas. If this is the case, a variable measuring police presence in an EA would likely reveal a spurious positive correlation with burglary rates.

Neighbourhood Watch and other self-policing programs are likely to be inextricably linked to the socioeconomic and family status dimensions of a neighbourhood already included in this analysis. Furthermore the effectiveness of these programs has been called into question in recent years (Davidson, 1993: 3). A major study found that self-policing programs have no significant long-term impact on neighbourhood crime rates (Bennett, 1990: 158). These criticisms, coupled with the difficulty of quantifying this variable using an areal unit of analysis, caused this opportunity factor to be omitted from the research.

Areas in the city that have a reputation for high-crime rates tend to attract increased criminal activity (Davidson, 1984: 64). This "reputation" factor was not incorporated into this study for two reasons. First it is difficult to see the advantage for a burglar in deliberately seeking out high risk areas as the level of anxiety of local residents would not be conducive to their particular offense. Secondly, a "bad" reputation is an extremely qualitative characteristic, making it difficult to reduce to a series of variables –other than the existing variable of burglary rate.

Finally, the proximity of a neighbourhood to offender areas is also recognized as an important determinant. It is usually treated as simply another opportunity factor (for example, Waller and

Okihiro, 1978; Maguire and Bennet, 1982). However, the relative accessibility of a neighbourhood is much more complicated than simply its linear distance to potential offenders. The complexities underlying a neighbourhood's relative accessibility are the subject of the next section.

2.3.3 Access Factors

Environmental criminology has produced a great deal of innovative research on the geography of human movement at the level of the crime scene. For example, it has been found that right-handed criminals tend to flee to the left, but move to the right when they encounter obstacles; they discard evidence to the right; and stay near outside walls when hiding in large buildings (Grescoe, 1996: 51). The predictable nature of human spatial behaviour is also important to an understanding of criminal activity at the intraurban scale of analysis. This section outlines findings from research on burglar spatial behaviour applicable to an aggregate study. In way of summary, two access factors are identified for use in this analysis: awareness and proximity.

The Journey to Crime

Traditional ecological studies analyzed the crime rate of a given neighbourhood using only offender data recorded for the same areal unit. Yet it is now known that the majority of property crimes occurring within a neighbourhood are committed by people who reside elsewhere (Bursik, Jr. and Grasmick, 1993: 60). The distance travelled by offenders to targets is 'sufficient to affect quite markedly the location of offense and offender areas' (Maguire & Bennet, 1982: 27). This "journey to crime" (a.k.a. "criminal commute") follows a predictable pattern and, as such, is of fundamental importance to the analysis of burglary rates recorded for areal units.

"Spatial behaviour and behaviour in space offer themes well understood in some areas of geographical research but not at all well explored in the context of crime" (Herbert, 1989: 5). One area, marketing geography, makes use of the distance decay concept to explain the spatial patterns of supply and demand within the city. This relationship is formalized as the spatial demand curve which states that with increased distance between a consumer household and a store, there will be a decrease in the quantity purchased at that location by the particular household (see Figure 2.2). The decrease is attributable to the increased cost in time and money for the consumer to shop at the more distant location. The spatial demand curve "plays an important role in determining the spatial structure of retailing in the metropolis" (Jones and Simmons, 1990: 36).

A distance decay curve can also be used to explain the spatial structure of residential burglary within a city. Despite obvious differences in their motivations, a burglar can be regarded as a special kind of consumer who is trying to acquire the best product while accruing the fewest costs. In addition to the same transportation costs incurred by consumers (in time and money), potential offenders face the additional costs of becoming acquainted with unfamiliar areas (Brown, 1982: 248).

There is one important difference between consumer and burglar spatial behaviour. Burglars try to avoid being recognized by their neighbours, causing them to travel a short distance from their residence before seeking a target (see Figure 2.2). Consequently, the requirement of familiarity coupled with the need for anonymity causes burglars to frequently target dwellings located on the fringes of their own neighbourhood (Brantingham and Brantingham, 1981: 46).

There is a great deal of empirical evidence to support the journey to crime concept. The exact parameters of the distance decay relationship vary between studies, being in part a function of the city and country, as well as the type of offender. However, all the journey to crime studies found for the literature review identified burglary as a fairly local crime, with most events occurring within one or two kilometres of the offender's home.

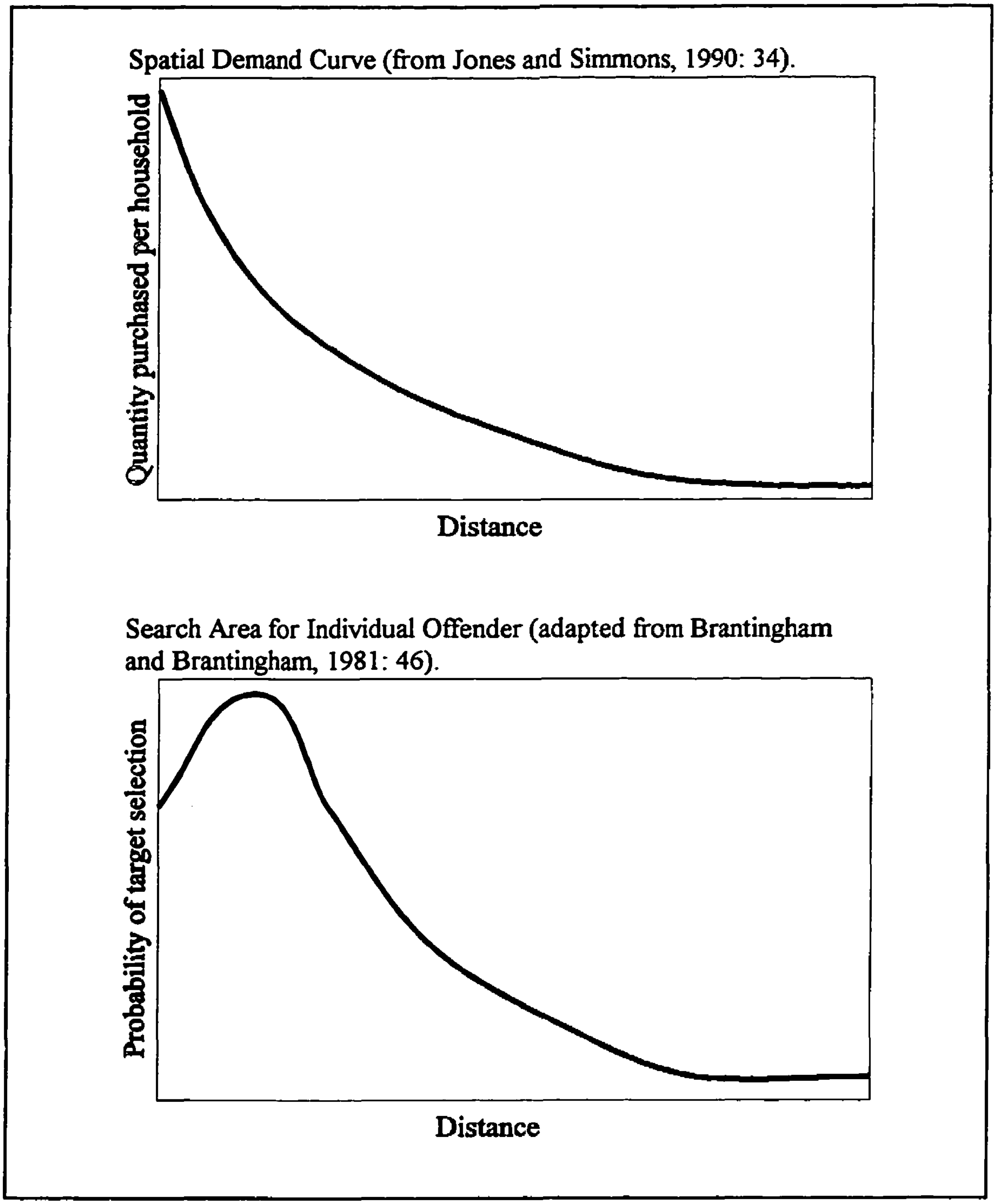


Figure 2.2 Consumer and Offender Travel Behaviour

The importance of the journey to crime has undergone some criticism. Harris argues that unlike robbery, where recognition avoidance is paramount, burglary requires careful target study and selection, which does not necessitate inter-neighbourhood travel (Harris, 1980: 86). Conversely, another perspective is that the automobile renders intraurban distances insignificant as "[offenders] have readily accessible modes of transportation, which allow them to move rapidly from one area to another, crossing census tract boundaries and spreading out over the metropolis" (Felson, 1986: 127). The negation of distance made possible by the automobile threatens to render the journey to crime concept archaic; moreover, it casts doubt on the relevance of any spatial behavioral aspects of this crime. In the next section a detailed examination of the different types of burglars is used to demonstrate the continued value of the journey to crime concept.

The Burglar Hierarchy

The search area distance decay curve is a succinct way of describing the burglar's criminal commute. However, in practise, the spatial behaviour of offenders is much more complex than this relationship could describe. In order to explain the intricacies of offender travel behaviour, it is necessary to identify the different types of burglars.

The burglar population can be divided into a three-tier hierarchy of increasing sophistication (see Table 2.2). This hierarchy corresponds roughly to the age breakdown outlined in Table 2.1. There are substantial differences in target selection and spatial behaviour as one progresses through the hierarchy, and thus each type of burglar has a distinct spatial pattern.

Table 2.2 The Burglar Hierarchy

CHARACTERISTICS	OPPORTUNISTS	SEMI-PROFESSIONALS	PROFESSIONALS
Age	12 - 17	18 - 24	25 and over
Awareness Space	Very limited	Substantially expanded	Very extensive
Journey to Crime	Very short, likely to be within 1 km	Longer commute but still likely within 2 km	Very high mobility. Intraurban distance not a factor.
Choice of targets	Known dwellings and spontaneous opportunities	On the fringes of their own area and semi-spontaneous opportunities	Affluent, low density targets.
Spatial pattern	Very local	Local, but at some distance from their residence	High income areas in the suburbs.

Most burglars are opportunistic amateurs. These tend to be adolescents going through a delinquent stage of their lives (Maguire and Bennett, 1982: 161). For very few of them is this a first stage in their criminal career, and consequently, they are the target of most crime prevention initiatives (Herbert, 1989: 5).

In studying the young, opportunistic burglar, it is important to understand the role of limited rationality. Traditional studies assume an entirely rational offender whose objective is to seek out an attractive target, travel to it at an appropriate time, and then break-in and enter after careful consideration of the situation. However, it is often the case that a spontaneous opportunity presents itself, motivating an individual who is otherwise not seeking to commit a crime to do so; for example, a neighbour's house window left wide open during the work day. The offender's decision is often made very quickly when presented with a spontaneous opportunity (Bursik & Grasmick, 1993: 61).

A young offender is likely to encounter a spontaneous opportunity during their daily routine activities. Therefore, a break-in is most likely to occur close to the areas which they frequent (home, school, work or recreation). This contributes to the very short journey to crime pattern of juvenile delinquents –usually specified as within a kilometre radius of their homes (see Baldwin and Bottoms, 1976; Repetto, 1974).

The second group, the semi-professionals, tend to be unattached, disadvantaged men aged 18 to 24. They are more experienced than the first group in how to break-in and how to unload stolen property, as they have acquired criminal skills through social interaction and physical observation (Rhodes and Conly, 1981: 169). Many of these criminals are repeat offenders and consequently are known to the police; however, this is still a part-time occupation for most of this offender group (Maguire and Bennett, 1982: 39).

Spontaneous opportunities remain a factor in the semi-professional's criminal behaviour. However, this term may be better described as semi-spontaneous opportunities as at this stage in the hierarchy burglars are "...individuals with high incentives for crime who go about their daily routines keeping an eye peeled for opportunities" (Rengert, 1989: 166). Spontaneous opportunities are thus more likely to occur in areas where a large number of semi-professional burglars reside. This may account for the tendency of crimes such as burglary to cluster in low income neighbourhoods, despite knowledge of more profitable areas in the city (Bursik & Grasmick, 1993: 61).

At the semi-professional stage of the hierarchy, the burglar begins to expand their awareness space through the exploration of the areas just beyond the limits of their routine activity space (Brantingham and Brantingham, 1981: 46). Therefore, neighbourhoods close to where an offender resides and close to where they work are likely to become incorporated into their target selection decision-making.

Further to this point, many adult burglars travel through the city to get from home to work and recreation. Thus it is likely that areas located next to or near major traffic arteries are also within the awareness space of a greater number of offenders. When one considers that semi-professional burglars are those who are ever seeking targets in their daily routines, a neighbourhood's adjacency to, or containment of, a major traffic artery likely causes an increase in its risk to burglary. Quantitative evidence to support the role of traffic arteries in the distribution of residential burglary is provided in a British study (Maguire and Bennet, 1982: 39).

Despite the increased awareness space, and thus activity space, of the semi-professional, the distance decay curve remains a good representation of this criminal's spatial behaviour. This is because "most criminals like to operate in areas they have some knowledge of but actually prefer areas they are very familiar with" (Rengert, 1989: 167). Previous research suggests that non-professional adult offenders generally select targets within two kilometres of their homes (see Costanzo, 1986; Evans, 1989).

The third group, the professionals, is much smaller than the first two; however, they commit many hundreds of burglaries over their criminal careers (Maguire and Bennett, 1982: 58). Professionals are career criminals usually in their mid twenties or older. They are highly mobile and selective in their choice of targets. This group is the dominant offender group in commercial burglaries, but is far less prominent in residential burglary. Those that are almost exclusively target affluent neighbourhoods (Maguire and Bennett, 1982: 40).

A further comparison with consumer spatial behaviour can be made with professional burglars. People looking to make day-to-day purchases will generally travel to the most convenient store. Those looking to purchase shopping goods (department store-type purchases) are willing to travel further than the nearest location in order to maximize their return through exposure to a wider variety of goods and/or better prices (Jones and Simmons, 1990: 36). The professional burglar is also purposively looking to maximize the reward for their efforts. This criminal has made a career out of identifying low-risk, high-return opportunities throughout the city, and are highly mobile. Consequently, at this stage in the hierarchy, the burglar's awareness space is very extensive. (Brantingham and Brantingham, 1981: 45). This combination of awareness and mobility causes affluent homes, regardless of their location within the city, to run a higher risk of being broken into. The critics of the journey to crime concept are likely thinking of the professional in their negation of distance as a factor in this crime.

These findings suggest that are in fact a series of journey to crime patterns, each relating to the various tiers of the burglary hierarchy (see Figure 2.3). The journey to crime is thus still relevant to the majority of break-ins, although significantly less important in professional burglaries. This has important consequences for a spatial analysis such as proposed for this study. This topic will be discussed in Section 2.4.1.

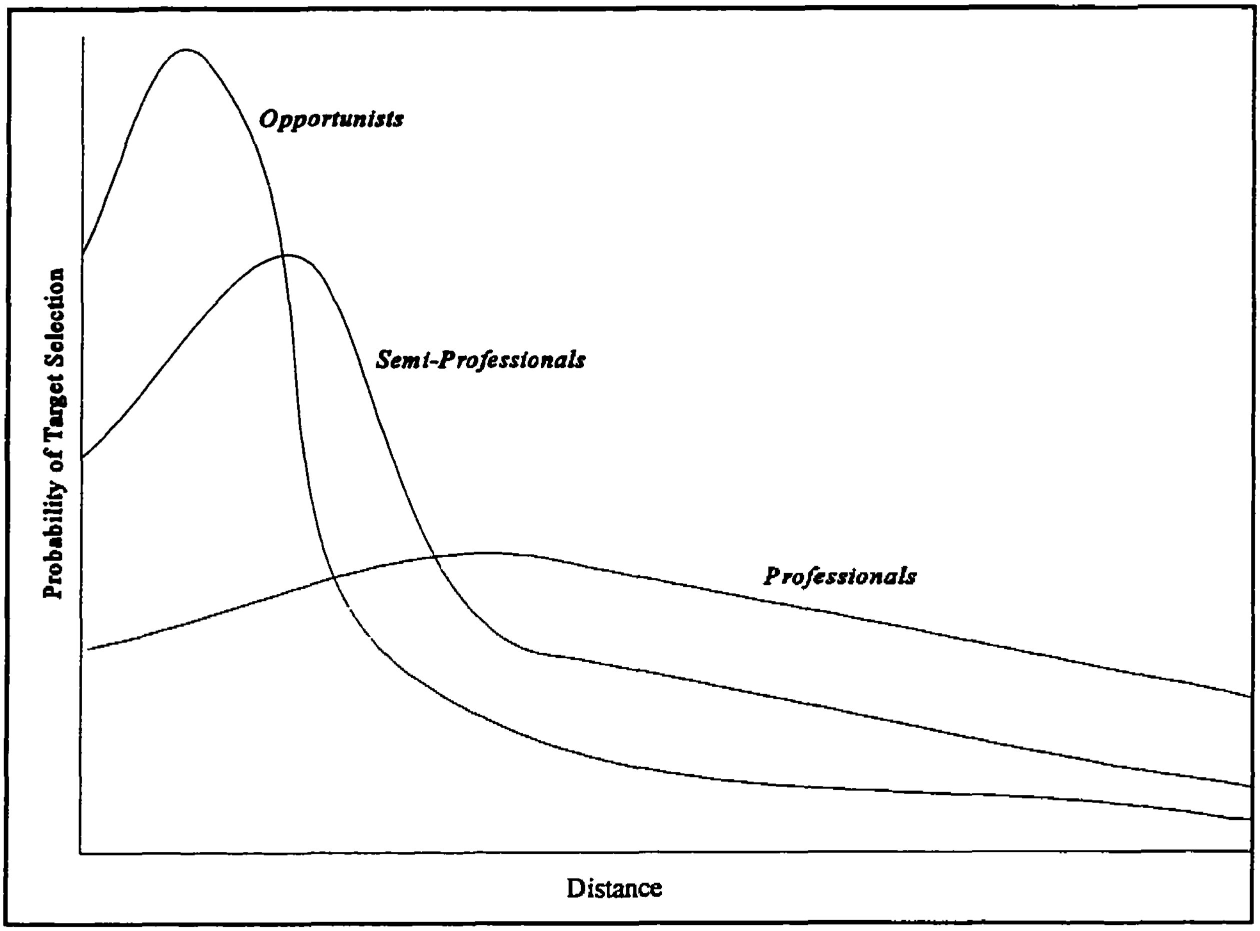


Figure 2.3 The Different Journeys to Crime

The Gradient Hypothesis

A second expansion on the journey to crime concept is to recognize the influence of city structure on offender decision-making behaviour. The traditional gradient hypothesis, based upon Burgess' zonal model of urban ecology, proposes an inverse linear relationship between crime and distance to the central business district. Central areas were felt to experience higher rates of delinquency because they contained more of the social and physical conditions associated with this behaviour (Stahura and Huff, 1986: 55). In this study it is posited that areas close to the downtown are also more likely to be within the awareness space of a potential offender due to their travel to the central core for work, shopping or recreational purposes. "The central business district is the second most used area after the home neighbourhood. Therefore, the locations of crime sites are also expected to be skewed in the direction of the central business district..." (Rengert, 1989: 166). It is anticipated for this study that once other variables have been introduced to the analysis which capture the social and physical characteristics of central neighbourhoods, there will remain a centripetal dimension at work reflecting this geography of visibility.

Awareness and Proximity

The role of offender spatial behaviour in the distribution of this crime is best demonstrated in Waller and Okihiro's analysis of Metropolitan Toronto, although they did little to incorporate the spatial dimension into their statistical analysis. They found offender areas also tended to be opportunity areas. This suggests the actions of opportunists and the short trips made by semi-professionals to the edge of their own neighbourhoods. They also found that neighbourhoods located close to public housing also suffered higher rates. This suggests the role of a journey to crime behaviour of burglars to areas outside of the confines of their own neighbourhoods. At a larger scale, they found that the gradient hypothesis appeared to be applicable to Metropolitan Toronto burglaries, as rates tended to decline with distance from the city centre. Noteworthy exceptions were affluent, low density suburbs, which tended to have high burglary rates –suggesting the work of professional burglars (Waller & Okihiro, 1978).

In summary it is evident that the intraurban distribution of residential burglary is partly a function of two access factors: awareness and proximity. The degree to which a potential offender is cognizant of a particular target and willing to travel to it is a function of their age and experience as a burglar, their mobility, the target area's visibility in respect to the city's thoroughfares, as well as its relative location within the city. Potential offenders and opportunities are unevenly distributed across the city, as is their degree of interaction, which has consequences on the spatial structure of the crime across the city. Specifically, this is likely to result in discrete zones of criminal behaviour across which the significance of the previously outlined offender and opportunity factors shifts. In spatial statistical terms this implies a spatially heterogeneous phenomena, or the presence of "spatial regimes", a topic discussed further in the next section.

2.4 Spatial Data Analysis

In order to test the concepts described in section 2.3, a data set representing the determinants was compiled. The analysis of such spatial information is not as straightforward as it is for non-geographic data. The purpose of this section is to identify the unique characteristics of spatial data and the subsequent need for specialized tools for their analysis.

Haining describes three different types of problems encountered in the analysis of spatial data. First are those that result from the particular data set, such as missing or unreliable values. The second type relate to the problems which stem from statistical assumptions not being met. The third type of problem relates to the introduction of modelling assumptions relating to the particular subject matter theory underlying the research (Haining, 1990: 4). The latter includes both scientific evidence produced in previous research as well as the analyst's intuition regarding the relationships at work (Haining, 1990: 54).

The first set of problems, regarding the data itself, is not unique to spatial data analysis and thus is left to the next chapter. The other types of problems, as they relate to this study, are illustrated through a description of the difficulties of applying the least squares regression equation to this particular data set. This leads into a discussion of spatial effects, two attributes of spatial data which differentiate it from other kinds of data. In conclusion the two stages of spatial data analysis are described.

2.4.1 The Ordinary Least Squares Regression Model

The objective of this study is to explain the spatial variation of residential burglary rates in terms of covariates referenced to a common spatial unit, as well as in terms of the spatial arrangement of these units. This relationship is to be described formally as a vector of parameters in a regression equation. The ordinary least squares (OLS) form of regression is commonly used in geographic research into crime.

In this model, the dependent variable at location i (y_i) has a functional relationship with a set of explanatory variables measured at the same location ($x_{1,i},\ldots x_{k,i}$). This relationship is expressed as a regression equation:

$$y_i = \beta_0 + \beta_1 x_1 + \ldots \beta_k x_k + e_i \qquad i = 1,\ldots n \tag{1}$$

where $\beta_0,...\beta_k$ are regression coefficients, $x_{1 \text{ to } k}$ are the explanatory variables, and e_i the error (the unexplained variance).

The variance is unknown and has to be estimated from the residuals of the fitted model. This is calculated as:

$$\delta^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n - p} \tag{2}$$

where y_i is the observed value at area i, $\hat{y}$ is the regression estimate of that value, n is the number of observations, and p the number independent variables in the equation.

The residuals are also used to measure the fit of the model. This is done by calculating the coefficient of determination, which measures the proportion of the total variance explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{3}$$

Inferences about the phenomena of interest based on regression results rely on several assumptions. First that for any fixed value of any of the independent variables, the distribution of the dependent variable is normal, with mean $\mu_{y/x}$ and a constant variance of δ^2. Secondly that the dependent variable's value at any given observation is independent of the values of other observations. Thirdly that a linear relationship exists between the variables under study. Finally that the residuals are normally distributed, independent random variables with a mean of 0 and a variance of δ^2.

The difficulty in applying this model to geographic data is that it is likely that the error terms are not independent, but are in fact correlated with themselves over space. Furthermore it is unlikely that the variance will be constant either. Therefore, the confidence intervals for the regression coefficients will be invalid and, as such, so will the calculation of the overall model fit.

2.4.2 Spatial Effects

The causes of dependency can be attributed to two characteristics of spatial data known as spatial effects. These consist of spatial autocorrelation and spatial heterogeneity. Spatial effects act as a hindrance to analysis because they prevent the use of common statistical methods such as least squares regression. Yet they are also valuable information in and of themselves, and if properly accounted for, are able to provide new insight unattainable through non-spatial methods.

Spatial autocorrelation is the extent to which a variable is correlated with itself over space. Of particular importance is positive spatial autocorrelation which refers to a clustering of similar values. Its presence in areal data has two primary causes. The first relates directly to Haining's "statistical assumptions" set of problems. This occurs when data relating to a particular phenomena is collected for areas which do not correspond to its spatial scale or spatial partition. This tends to result in neighbouring areas being similar in values simply because of the nature of their spatial arrangement. An example of this might be air pollution measures recorded for individual neighbourhoods. Spatial autocorrelation may also result from a spatial process at work. Such substantive autocorrelation relates to Haining's "modelling assumptions" problems. For example, neighbourhood rates of a contagious disease may be spatially autocorrelated because high-rate areas likely spread the illness to neighbouring areas through spatial interaction.

Whatever the particular cause or causes of spatial autocorrelation in a variable, it serves to eliminate the degree of independence between the observation units. This threatens to reduce the amount of evidence in the data set that the analyst can use to make inferences. Yet spatial autocorrelation is also "an important part of the discovery process by which one improves on a current base of knowledge" (Miron, 1984: 201). Specifically, spatial autocorrelation measures can be seen as uniquely geographic information which, if properly incorporated into an analysis, provide greater explanatory power. For example, spatial autocorrelation resulting from a spatial process can be formally represented in a regression model by the estimation and interpretation of an autoregressive parameter.

In this study it is assumed that any spatial autocorrelation found in the dependent variable, burglary rate, is not the result of a substantive spatial process such as contagion. If burglary does exhibit positive spatial autocorrelation, it is more likely due to the clustering of its determinants. With such "nuisance" autocorrelation, the concern becomes simply to eliminate the dependency from the regression model rather than incorporate and interpret it. When the objective is to eliminate spatial autocorrelation, it is the residuals (the unexplained variance) of the model which become the subject of interest. Uncorrelated error terms indicate that the model has been properly specified (O'Loughlin, 1986: 68).

There are a number of causes for regression residuals to be spatially autocorrelated. First, there may be an important spatially autocorrelated explanatory variable missing from the analysis. Its absence will likely be reflected in the pattern of residuals. Secondly, it may be that the data was not adequately transformed in order to linearize the relationship among variables, causing a systematic over- or under representation of the relationship, which may translate into a spatial pattern. Thirdly, there may be a systematic error measurement of the dependent variable. This is of particular importance to this study, as the level of reporting of burglary incidents may be unevenly distributed over space. A fourth cause of spatially autocorrelated residuals is the influence of extreme values, or outliers, on the parameter estimates which may cause the significant under- or over-prediction of less extreme cases. Finally, their may be two or more discrete areas within the study area (spatial regimes) within which the relationships between the variables are substantially different, causing the residuals of a global model to cluster accordingly. This is the case of the second type of spatial effect, spatial heterogeneity.

Spatial heterogeneity may result from the influences of processes operating at larger scales which cause sub-areas of the study area to respond differently to the same circumstances. For example, "the housing market of an urban area is spatially heterogeneous (measured by price differences for identical housing units or components) because supply and demand conditions vary between areas" (Haining, 1990: 23). These supply and demand conditions are also applicable to burglary, as there are different types of offender who exhibit markedly different spatial behaviour (demand), and whom respond differently to potential opportunity (supply). For example, the professional burglar operates over a wider spatial scale than the opportunist, which may result in a shift in parameter estimates in areas distant from clusters of opportunistic offenders.

Spatial heterogeneity results in different slopes and/or intercepts in the regression equation for each of the sub-areas. It relates to the association of variables (the regression model's partial correlation coefficients) regardless of the presence of other variables. Therefore, a true heterogeneous spatial structure is one which isn't simply lacking an important variable, and cannot be "fixed" through the introduction of a binary variable representing regime membership. Thus the construction of a regression model not explicitly designed to account for this effect would result in correlated residuals.

Components of Spatial Data Analysis

There are a number of techniques available to both explore spatial data and to correctly model it through the incorporation of spatial effects. The field which encompasses these methods is referred to as Spatial Data Analysis (SDA). It is comprised of two stages: Exploratory Spatial Data Analysis (ESDA) and Confirmatory Spatial Data Analysis (CSDA). ESDA is performed using a flexible approach that is responsive to the patterns that successive analysis uncovers (Haining, 1990: 50). The exploratory stage is used primarily to describe the data, suggest theory and the appropriate models to use in an analysis.

CSDA analyses a problem with a model-driven approach. While the role of ESDA is to suggest theory, CSDA is used to estimate a formal statistical model and then confirm it through hypothesis testing. There are a number of confirmatory models available to perform geographic studies which account for spatial effects. In deciding upon one it is necessary to consider the data set as to which model makes the most intuitive sense as well as is the most statistically sound. This topic is discussed further in section 4.2.1.

Spatial data analysis has been greatly enhanced in recent years due to the widespread introduction of Geographic Information Systems (GIS). This technology allows for a faster and more flexible use of ESDA, both before and after the CSDA stage. The next section of the literature review looks at the traditional role of GIS in the analysis of spatially-referenced crime data as well as its potential contribution to this particular study.

2.5 Geographic Information Systems and Crime Analysis

Information systems have been used to assist in the analysis of crime data for over 25 years. The pioneer system, PATRIC (Pattern Recognition for Investigating Crime), was first implemented by the Los Angeles Police Department during the early 1970s. Its purpose was simply to record and display the location of criminal events. A number of programs of increasing sophistication have since been developed for crime analysis purposes (Maltz et al, 1991: 26).

An integral component of current police information technology is GIS, which can now be found in the majority of North American police departments (Christie and Shields, 1996). Its ability to store and display data related to the location of criminal events remains its primary value to these organizations. Analysis is generally restricted to simple mapping and querying (Miller, 1993).

Yet the widespread implementation of GIS has resulted in a growing interest in the methods of using the technology to explore and analyze spatially referenced crime data sets (Davidson, 1993: 8). This is an interesting development as criminology is a discipline "in which spatial thinking has played a very minor role in the past" (Goodchild, 1995: 46). Recently research has been made into the use of artificial intelligence to find spatio-temporal patterns of crime sites (Hernandez, 1991). GIS has also been used to determine the influence of street lighting on criminal behaviour (Berry, 1993: 133). Also, theory from environmental criminology has also been operationalized through the creation of geographic profiling software described in section 2.2.

However, there is little evidence that GIS is being used in conjunction with more advanced spatial analytical techniques to explore the crime's intraurban distribution. A goal of this thesis is to demonstrate how a GIS-SDA linkage provides a powerful means to understand the patterns and causes of criminal behaviour such as burglary. The specific benefits of integrating GIS into this study are now detailed.

The most obvious advantage of incorporating GIS into this analysis is its advanced visualization capabilities. In addition to displaying the variables of interest, GIS software is able to visualize spatially disaggregated statistics, such as model residuals and local indicators of spatial association –information that is difficult to interpret in a spreadsheet format.

The second greatest contribution of GIS technology relates to the quantification of spatial relationships between areal units –an essential first step in performing SDA. In the past this information was laborious to quantify because it involved measuring the relationship of each observational unit to all others. Furthermore, there are a number of ways to define the spatial relationship of areal units, including binary contiguity, length of common border, distance between centroids, or a combination of these three. Building a series of such matrices is a very time consuming task.

This can be greatly facilitated through the use of GIS because the topological data contained within its existing data structures can be exported into an SDA module which uses this data to construct spatial weights matrices. This module can also be created to work within a GIS using its macro language or a common programming language. The details of the approach taken in this particular study will be outlined in the next chapter.

A third advantage of GIS to crime analysis is their ability to relate different spatial data sets (Berry, 1993: 133). They make possible the creation of non-traditional variables through the manipulation and query of spatial data layers. An example of this process is the creation of an alternative population density measure, which is outlined in section 3.3.2. Related to this benefit is the ease of which lagged versions of variables can also be created.

Thus in summary GIS and SDA form a natural partnership with the field of environmental criminology. Environmental criminology provides the necessary theory for variable selection, analysis and interpretation. SDA provides a statistically sound means to explore and analyze spatially referenced crime data sets. GIS can improve the inputs and the outputs of a traditional analysis; moreover, it facilitates (and often makes feasible) the entire research effort. The next chapter describes the operationalisation of this methodological approach.

CHAPTER 3: METHODOLOGY AND DATA

This chapter provides the transition from the theoretical knowledge-gathering of the literature review to the modelling of the data discussed in the next chapter. First the methodological framework of the study is described. Next a conceptual framework is constructed to provide a structure for the ensuing analysis. This section discusses a major problem associated with aggregate analysis, the ecological fallacy. The study area selection and data collection procedures are then described. Issues related to missing or inaccurate data are then discussed.

3.1 Methodology

The three components of this study are spatial data analyis, GIS and theory from environmental criminology. The three components were used dynamically throughout the entire analysis (see Figure 3.1). The theory was used in the selection and construction of variables, as well as to guide the analysis. GIS and SDA was incorporated through the use of software packages. Throughout the analysis, MapInfo GIS was linked to SpaceStat, a spatial statistical package, through the import and export of data using a common file format (loose coupling). ARC/INFO was also used when advanced GIS functionality was required (such as buffering and overlaying), as well as to provide topological data to the SDA package.

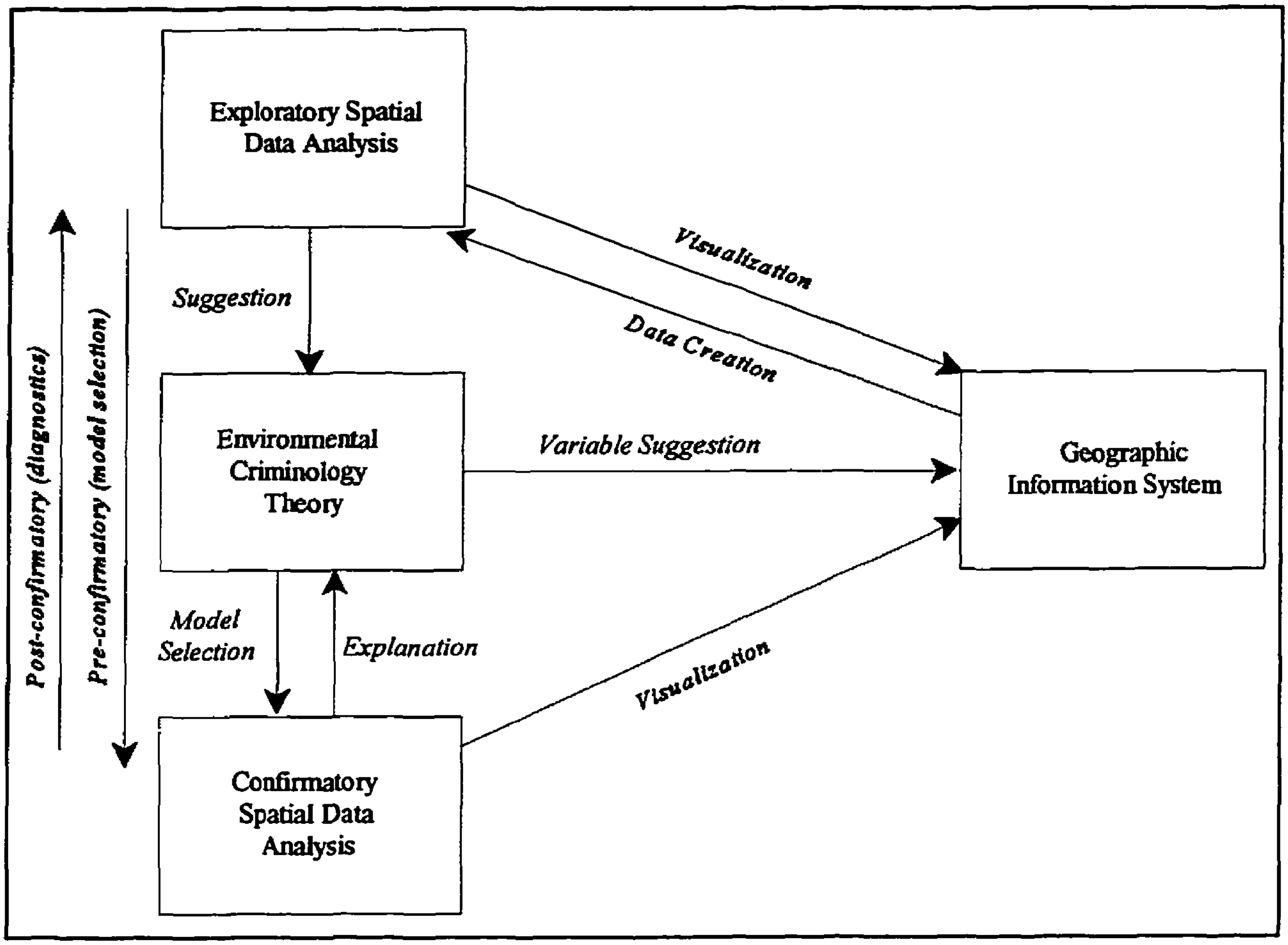


Figure 3.1 Methodology for the Analysis (adapted from Anselin and Getis, 1992)

3.2 Conceptual Framework

The literature review established that aggregate residential burglary rates are a function of a number of structural and spatial variables which can be classified into offender, opportunity and access factors. These factors are the product of the complex interaction of macro-scale constraints and the micro-scale considerations of the individual burglar (see Figure 2.1). Meso-scale studies must derive meaning on an activity performed by an individual (or small group) through the analysis of aggregates. This poses the danger of the ecological fallacy (a.k.a. the fallacy of decomposition or disaggregation).

In this study, meso-scale determinants were conceived as a neighbourhood's *potential* to both produce burglars based upon their offender characteristics, and to attract them based upon their opportunity and access characteristics. The motivation for an individual to commit burglary is a function of the determinants mentioned in the literature review such as gender, age, stability and socioeconomic status. Areas with a disproportionate amount of people with the right criminogenic combination of these factors thus have a greater propensity to produce offenders. This propensity can be estimated using aggregate data.

A similar approach can be used to measure the potential opportunity within each area. A burglary is a result of an individual's choice of an individual dwelling based upon their decision regarding signs of occupancy, perceived rewards and ease of entry. At an aggregate scale, attractive dwellings are likely to be more prevalent in some neighbourhoods than others. Furthermore, there are neighbourhood-scale influences at work on the individual's decision such as population density and land-use.

Access can be seen as each neighbourhood's spatial attractiveness in terms of the number of offenders who are aware of it and willing to travel the distance to it. Awareness and proximity are a function of the different types of burglars –each of which has been shown to have a distinct spatial pattern. As different types of offenders are likely to be unevenly distributed across the study area, it is important to create offender variables which capture this spatial behaviour.

Thus it may be possible to quantify the dimensions underlying criminal activity in terms of how they relate to aggregates by taking a probabilistic approach. The difficulty, however, lies in extricating the contribution of each determinant of this crime measured at this scale. Ecological studies have long been criticized for confusing correlation with causation. For example,

> "...it is possible to identify what may be called *correlates* of crime, but the correlates are always imperfect. To say that poverty is often associated with crime is not to say that poverty *causes* crime. If that were the case, poverty would be linked to crime universally -every poor person would be criminal. But only a minority of the poor, or victims of discrimination, or the insane, or the retarded, become criminals, suggesting that criminality is a probabilistically arrived-at condition. Thus predictions about criminality in an area or in relation to a group of individuals are high risk propositions" (Harries, 1980: 5).

This risk pertains to mistaking correlations between aggregate variables with correlations calculated for individuals (Harries, 1980: 7).

Davidson provides three options to deal with false or spurious explanations of crime resulting from the application of the ecological method. First, simply don't work with or trust the results of an aggregate analysis. Or secondly, use techniques to isolate within the aggregate correlation the individual correlation from the aggregation bias. In this way a researcher can understand the relationship between the levels of analysis. Or one can remain sceptical about inferences made from an ecological analysis until they are confirmed by studies at the scale of the individual (Davidson, 1981: 87-89).

The first option was felt to be over-critical of aggregate scale analysis. While ecological analysis are unable to provide definitive answers, they can suggest associations between determinants and burglary rates. These suggestions may lead to further, more detailed research. Furthermore, false or spurious inferences involving projection from smaller to larger units ('fallacies of aggregation) have also been found to exist (Shannon, 1986: 30), so that there may not be any ideal scale for crime analysis. Finally, the observation units used in this analysis were considered sufficiently small and homogeneous to permit meaningful conclusions to be drawn.

The second solution of applying some form of multi-level model was considered to be outside the scope of this study. Furthermore, a condition of receiving the burglary data from the local police was to maintain the anonymity of the individual victimized households, which prevents the analysis of individual crime sites.

The approach made in this study is to view the results as only one component of a total explanation of the distribution of this crime. A full understanding of the spatial and structural determinants of burglary rates could only ever be achieved through an analysis over a variety of scales and through a variety of methodological approaches.

3.3 The Study Area

Kitchener-Waterloo, Ontario was chosen as the study area for a number of reasons. First it facilitated the collection of data to be able to meet the police representatives in person. Secondly it is an ideal study area for this form of human activity because it is largely isolated from the influence of other communities. Although situated in the densely populated Quebec-Windsor corridor, the nearest large cities (Toronto, Hamilton and London) are located between 50 to 100 kilometres away. The small cities of Cambridge and Guelph are located nearby, and likely house some of the offenders responsible for the study area's burglaries. Yet considering the local nature of this crime (except for professional burglaries), the vast majority of the break-ins recorded for the analysis can be assumed to be committed by persons residing within the study area. A final advantage to using Kitchener-Waterloo as a study area is that it was possible to investigate areas of interest in person, which contributed a great deal of local understanding to the analysis.

Breaking and entering rates are not published for individual cities. However, data is available for the property crime rates recorded for police jurisdictions. A comparison of ten jurisdictions in southern Ontario reveals that the Waterloo Region, within which the study area is located, experiences relatively low property crime rates compared to its nearest neighbours (see Figure 3.2).

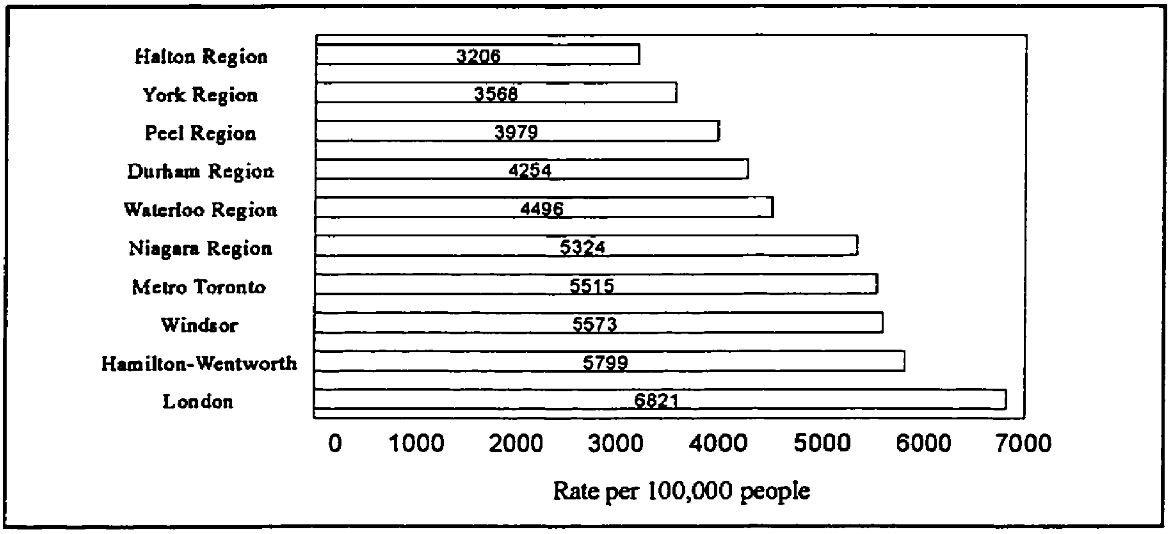


Figure 3.2 Property Crime in Selected Ontario Police Jurisdictions, 1994 (from Chard, 1995)

Although B&E data was obtained for the entire jurisdiction of the Waterloo Regional Police, the study area was restricted to the built up area of Kitchener and Waterloo. The exclusion of the outlying areas maintained an urban focus for the research. Rural burglaries are associated with different factors than urban ones (Maguire and Bennett, 1982). In order to create useable spatial connectivity matrices, it was also necessary to exclude the city of Cambridge, which is connected to the study area by only a thin urbanized corridor consisting primarily of industrial parks.

Some of the EAs located within the study area boundary were also excluded from the analysis. First there were commercial, industrial or institutional EAs which contained no residences and hence no residential burglaries. Non-residential areas were accounted for when measuring the degree of mixed land-use present within or bordering each neighbourhood. Secondly, there were 14 EAs for which there was no census data. These were small, multi-level residential buildings which had been assigned their own EA. The burglary incidents which occurred within these EAs were excluded from the analysis. While this does affect the study to an extent, it was felt to be a better choice than the alternatives. Estimating the missing census data or assigning it a global mean would introduce an unacceptable degree of inaccuracy. Simply leaving the entries blank would not suffice either, as the spatial statistical software used in this study cannot accept data files containing missing values.

3.4 Data Collection

After constructing a conceptual model and delimiting the study area, the information gained from the literature review was transformed into useable data for statistical analysis. There were four sources used to create the inputs variables: the burglary data set, 1991 Census data, existing GIS data layers for the area, and paper maps requiring digitization. Many of the variables were subsequently lagged (the average of neighboring areas' values was calculated) in order to capture the spatial interaction occurring between EAs.

3.4.1 The Dependent Variable

The data necessary to build the dependent variable *police-recorded residential burglary rate* was obtained from the Waterloo Regional Police. It consists of the addresses of over 2400 dwellings for which a break-in event was reported in 1995. This set does not include calls which the police deemed as highly questionable (such as for insurance fraud purposes). Of these reports, 1626 fell within the study area. The remainder occurred in either Cambridge, the outlying rural areas, or the EAs lacking census data. Also included in the data set were the date and time the burglary was reported to the police, as well as the estimated time frame within which the event occurred.

Maintaining the anonymity of the individual victims was an important issue. This restricted the analysis exclusively to areal units, necessitating the aggregation of the individual addresses to their respective EAs (of which there are 294). This scale was chosen as the best compromise between maintaining the confidentiality of the individual victims while ensuring that the units of analysis were geographically small and homogenous.

One method of data aggregation involves using a GIS street network file to perform address matching, followed by a point-in-polygon search using an enumeration area boundary file. During the address-matching phase, the GIS interpolates the locational coordinates of a particular household along a street segment. The point-in-polygon algorithm then assigns this point location to an EA based upon the coordinates of the census unit's boundary vectors. Thus the calculated locational coordinates (and consequently the assignment to an EA) is only an estimate, introducing a possible error in the aggregation process. This method is inadequate for a rare event such as burglary, as it would only take a small number of false assignments to undermine the statistical analysis.

As an alternative, geocoding was performed manually using a Vernon street directory. In this way it was possible to determine those addresses which could not be confirmed to fall within a given EA's boundary. On-site investigations were then made to determine the exact location of these sites.

The additional time required to manually perform this operation was certainly outweighed by the increased precision of the final product.

Once each event was assigned to its proper EA, a database query was made to count the number of burglaries per area. In some cases, multiple break-ins had occurred at a single address throughout the year. This raised the question of whether the inclusion of these events when determing the number of burglaries per area is warranted. Due to the aggregate nature of this study, it was decided to keep all the burglary events. If a certain number of break-ins did occur within the geographic boundaries of an EA during 1995, whether at the same particular household or not, the final result is that the neighborhood itself was the site of that number of burglary events.

Enumeration areas vary widely in the number of dwellings they contain; in this case, ranging from 20 to 844 with a mean of 301. Thus it would be impossible to make comparisons between neighbourhoods without first rating the data. Unlike many other crimes, the calculation of burglary rates is straightforward as "researchers have emphasized the need to relate the number of crimes to the amount of opportunity: thus, for residential crime, the number of burglaries must be related to the number of residences" (Waller and Okihiro, 1978: 15). For this reason, the sole denominator used to calculate the burglary rate was the number of dwellings within each EA.

One important issue had to be resolved before the burglary variable could be rated. The data set provided by the police was recorded for the year 1995, while the census information including the denominator, number of dwellings, was collected in 1991. This temporal mismatch threatened to inflate the burglary rates of EAs where development had occurred during the interim years.

A partial solution to this problem involved calculating the number of dwellings constructed between 1991 and 1995 by comparing the 1995 Vernon street directory with the 1991 edition. Any new listings were assigned to their respective EA, thus increasing the denominator for that particular area. This made for more accurate burglary rates, yet it could not resolve the problem that other census data used in the study were less reliable for rapidly growing areas, particularly ones with few households in 1991. This is one limitation of the study for which there is no solution –save for re-doing the analysis using 1996 Census data when it becomes available.

A second problem relating to the compilation of the burglary rate is the possibility of systematic under reporting, a bias which could make burglary rates dependent upon the type of neighbourhood (Evans, 1989: 90). Findings of the 1993 General Social Survey suggest that only 67% of the residential burglary incidents in Canada are reported to the police (Chard, 1995: 12). Yet burglary is one of the most likely crimes to be reported (Evans, 1989: 91). Fortunately for this study, previous research indicates that crime reporting rates are based more on the type of offence than type of victim; there is no evidence of any systematic bias in reporting rates based on income, ethnicity, gender or age (Evans, 1989: 91). Victim surveys used to generate a more accurate knowledge of the true frequency of criminal events, show similar patterns to official police data (Maguire and Bennet, 1982: 28). Therefore, it is generally recognized that residential burglary data are sufficiently reliable for geographic analysis (Evans, 1989: 93).

3.4.2 Explanatory Variables

The explanatory variables were constructed by using census data, GIS data layers, and through a merger of the two. Census data was downloaded from WLU Info's Census data base. GIS data layers relating to land use (schools, industrial zones, and major roads) were digitized from paper maps. Existing GIS layers included the EA topology and the study area's street network file.

The result of the data collection operation is a series of inputs for the spatial regression models to be built later in the study (see Table 3.1). They represent some of the major factors associated with residential burglary proposed in the literature. These include both classical ecological variables and spatial variants of these traditional inputs. Also collected were connectivity matrices necessary for the spatial analysis.

Table 3.1 Explanatory Variables Used in the Analysis

OFFENDER FACTORS	OPPORTUNITY FACTORS	ACCESS FACTORS
YOUTH *m1517*: % male aged 15 to 17 *m1824*: % male aged 18 to 24	**SOCIAL COHESION** *rented:* % of households that are rented	**AWARENESS** *km_cbd*: kilometres to the CBD *hi_schl*: distance to nearest high school *mjrd_100*: % of area within X metres of a major roads
SOCIOECONOMIC STATUS *lt15000*: % of male adults earning less than $15000* *lmavginc*: average income of males aged 15+*	**OCCUPANCY** *avghhsz*: average household size* *avgmntrs*:average number of maintainers* *infant*:% aged 0 to 4* *retired*: % aged 65 and over*	**PROXIMITY** *hwy_exit*: distance to nearest highway exit *m1517_1*: % of pop. within X km male aged 15 to 17 *m1824_15:* % of pop. within X km male aged 18 to 24
INSTABILITY *loneprnt*: % of households headed by one parent* *hhsz1*: % of households one person only*	**AFFLUENCE** *logvalue*: avg dwelling value* *bedrooms*:avg bedrooms per dwelling*	
	VULNERABILITY *lnewdens*: population density of buffered street* *year_blt:* average year of construction* *repairs*: % of dwellings requiring repairs* *industry*: (0/1) contains or borders an industrial area	

**= lagged variables also included*

Offender Variables

The youth factor was incorporated by calculating the proportion of residents in each EA which were males aged 15 to 17 (*m1517*) and males aged 18 to 24 (*m1824*) –the prime offender cohorts. This variable's utility as an explanatory variable is likely to be minimal due to journey to crime behaviour. Its primary function is as a control variable to compare the classical ecological approach to the spatial approach used in this study.

Two variables were used to incorporate socioeconomic status : % of males earning less than $15,000 a year (*lt15000*) and average income of males aged 15 and over (*lmavginc*). The latter variable was log transformed in order to conform to a normal distribution necessary for the parametric statistical tests used in this study. These two variables were chosen because each can be seen as a different measure of poverty; moreover, combined they give a rough indication of the skewness of income within each EA.

The final offender factor, instability, was quantified by incorporating two variables: % of families that are headed by one parent (*loneprnt*) and % of households with one person only (*hhsz1*). The first variable measures the proportion of broken homes in an area, which is relevant to the younger offender's level of stability. The second variable measures the proportion of residents lacking the stabilizing effect of marriage and/or children (in terms of illegal activity).

Opportunity Variables

In this study, social cohesion is considered as the extent of ties between residents. The most relevant measure for this, % of residents who had moved within the last year, was not available from WLU Info and so could not be incorporated into the analysis. As a proxy, % of dwellings which are rented (*rented*) was used, although this variable has many obvious correlations with other variables used in this study.

Occupancy is a function of the number of people in a dwelling place and the extent of their routine activities. The first aspect was quantified through average household size (*avghhsz*). The latter is measured by % of people under five (*infant*), as these individuals are more likely to be at home during the daytime with a parent or guardian; % of people over sixty-four (*retired*), who have less time blocks committed to activities outside of the home; as well as the average number of maintainers per dwelling (*avgmntrs*) –maintainers are those people responsible for the costs of maintaining the household.

Affluence was measured by average dwelling value (*logvalue*), a logged variable, and average number of bedrooms per dwelling (*bedrooms*). These are visible indicators of wealth, which are more relevant as opportunity factors than the actual income of the residents, which is used as an offender variable.

Finally, vulnerability was measured by the variables average year of construction (*year_blt*), % of dwellings requiring minor or major repairs (*repairs*), and population per buffered street coverage (*lnewdens*). The variable *year_blt* was created by calculating the weighted average of the dwelling year of construction data found in the Census. It is used to determine if older areas attract more burglary activity than newer areas. The repairs variable is used to quantify the association between the level of dilapidation in a given EA with its burglary rate. The final variable, *lnewdens*, is a substitute for the population density per unit^2 normally used in ecological studies. A comparison of the cities' road network and the census unit topology revealed that many border EAs contain a significant amount of non-urban space (agricultural areas, forests, etc.). Even small, centrally located

EAs are often dominated by a park. Population density measures based upon these areas would not properly reflect the density of their built environment.

A method to improve the population density variable involved using the buffer and intersect functions contained within ARC/INFO. The single-line street network layer was buffered to 25 metres, a figure that was felt to approximate the extent of a typical city property from the road. The choice of buffer size may be a conservative estimate of the true population density because in the more affluent areas, property lines likely extend much further back from the road. As a consequence, the influence of this variable may be somewhat understated. These buffered areas were clipped in ARC/INFO. The new population densities were calculated using these clipped areas rather than the entire EA (see Figure 3.3). This variable was logged to obtain a normal distribution.

A final aspect of vulnerability, landuse, was problematic to define. Initial attempts involved quantifying the presence of non-residential land use within each EA. However, there was a difficulty in defining these areas for the intent of this study (for example, whether to include institutional areas, small retail areas along King Street, etc.). It would be difficult to justify the necessary time to construct a single representative variable for this characteristic. Therefore, as an alternative, a binary variable was created representing whether an EA bordered or contained an industrial area large enough to be identified as a discrete area on the local MapArt street map. This introduced a degree of subjectivity to the collection of this particular variable, but this is likely to be unavoidable, regardless of the time and effort put into this procedure.

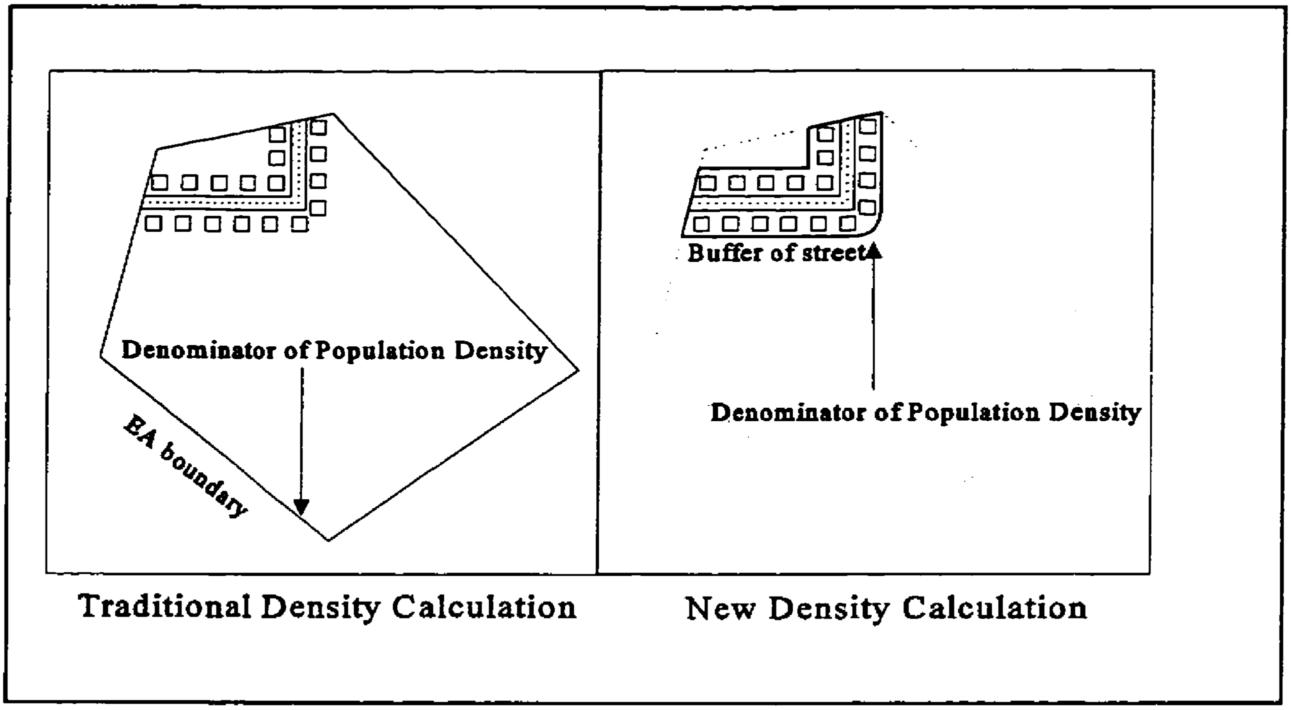


Figure 3.3 Measures of Density

Access Variables

Data for the two components of access (awareness and proximity) were collected primarily through the use of GIS. Awareness was operationalized using variables measuring the neighbourhoods' relative location within the city as well as its relationship to the study area's major roads. Proximity consisted of variables which capture the journey to crime at an aggregate scale. A number of different approaches were made to quantify proximity before a satisfactory method was found.

Awareness - Relative Location

In the literature review it was established that centrally located neighbourhoods are more likely to be known by offenders from across the urban area due to their proximity to major attractors such as work, recreation and shopping sites. In this study, distance to the central business district (*km_cbd*) was measured as the straightline distance separating the EA centroid and the corner of King and Queen streets, the point chosen as representative for the CBD. This site is immediately surrounded by high population densities, offices, industrial sites, as well as shopping and recreational centres.

It could be argued that Kitchener-Waterloo is in fact a bi-modal or even multi-nodal metropolitan area, as there are a large number of shopping, office and industrial nodes located in the suburbs. However, it was felt that Kitchener's city centre lies at the only true inner city in the region. Furthermore, the transportation system is centred upon this part of the city.

It was also posited that EAs located near high schools would also be more likely to be within the routine activity space of people in the prime offender cohort: males aged 15 to 17. This was operationalized by calculating the linear distance separating each EA centroid from the nearest high school (*hi_schl*).

Awareness - Major Roads

EAs either bordering or containing major traffic arteries were posited to enclose a greater proportion of high visibility dwellings and, consequently, suffer higher burglary rates. GIS was used to quantify this relationship. Using ARC/INFO, the major roads data layer was buffered from 100 to 300 metres at 50 metre intervals. An "Intersect" operation was performed to determine the proportion of each EA located in the buffered area. These area measurements were transformed into proportions by dividing by the total area. Statistics for each EA were then compiled and exported to the statistical software for further analysis. A possible refinement to this method would be to overlay the buffered major road layer onto the buffered street segment layer, rather than the EA layer, to estimate the proportion of the built area which is located in high visibility areas (as per the population density variable). However, at this stage in the study, the value of using GIS to improve the explanatory variables was not yet known.

A series of statistical tests was performed on the entire range of buffered major road layers in order to identify the optimum buffer size. This involved first identifying extremely high and low risk burglary rate EAs, a method whose mechanics are discussed in the next chapter. These two groups were then used to run t-tests on each buffer size. From Table 3.2 it is clear that the larger the buffer, the better the major road variable is at discriminating high burglary risk areas from low burglary risk areas. A notable exception is the smallest buffer (100 metres) which is bordering on significance. This preliminary evidence suggests that areas located near major roads do indeed experience higher burglary rates.

Table 3.2 The Identification of an Appropriate Buffer Size for Major Roads Using High and Low Risk EAs

Major Road Buffer Size	% of EAs entirely within buffer	T-test signif (ERGs)	R to burgrate (unadjusted for area)	R to burgrate (adjusted for area)	Spearman's (unadjusted for area)	Spearman's (adjusted for area)
100 m	2	0.12	0.02	0.02	0.05	0.05
150 m	4.8	0.16	0.01	0.01	0.02	0.03
200 m	8.5	0.10	0.02	0.02	0.01	0.02
250 m	11.6	0.06	0.03	0.03	0.01	0.02
300 m	18.7	0.05	0.03	0.04	0.01	0.02

To further investigate this relationship, a Pearson's correlation test was run using all of the burglary rates. The highest R value was again the largest buffer (300 metres). However, this is likely more a function of the areal unit size. Many EAs fall completely within the 300 metre buffer, resulting in a non-normal data distribution, which in turn affects the correlation measures. Using the non-parametric Spearman's correlation test, and accounting for areal unit size, revealed that the 100 metre buffer is the most significant, although the correlation is very weak. Based upon these statistics, it was decided to use the 100m buffer (*mjrd_100*) as an explanatory variable for the regression models.

Proximity - Incorporating the Journey to Crime

A number of different attempts were made to quantify the journey to crime at an aggregate scale. The first option considered was to include lagged versions of the two age-gender cohorts in the model. However, this course was reasoned to be inappropriate as it would add to the complexity of the analysis by introducing a number of new variables for each lag.

The second approach is to use a population potential model:

$$\sum \frac{P_{ij}}{D_{ij}^{x}} \tag{4}$$

This is a simple gravity model with P_{ij} representing the raw number of the age-gender cohort of interest divided by a distance measure D_{ij} raised to power x. Distance acts quickly to negate the effect of a given population group, necessitating the introduction of raising distance to some exponent. The potential is calculated for all EAs to all EAs. SpaceStat provides the necessary module to build these models using centroid coordinates and the population variable of interest. Three models were constructed: a simple linear distance, squared distance and cubed distance.

The population potential model was found to have a severe limitation. Its origins are from physics, where it is used to describe potential work. There is no direct correspondence between the gravitational pull of two entities with people's movement patterns over space. As a result potential measures are arbitrary. This in and of itself may not be limiting, as accessibility in this study is a relative rather than absolute measure. The problem lies in operationalizing the model: in order to account for the short distances separating small, neighbouring EAs, it was necessary to rescale the distance matrix to avoid having a fractional denominator. There was no possibility to rescale the population figures accordingly without an uneven shift in relative values. Alternatively, changing the distance units from kilometres to metres resulted in extremely small population potentials for most EAs. Consequently, there was no clearly defined way to calculate a meaningful measure of the offender potential for each area in this study.

As an alternative, concentric zones were constructed around the centroid of each EA. These were used to calculate the number of potential offenders at various distance bands. These figures were then divided by the total population within the band to calculate the proportion of people who were in the age-gender cohorts of interest. The literature review of journey to crime travel behaviour suggested a wide range of cut-off points. Therefore, t-tests were used to identify the most appropriate band size. Extremely high and low risk areas were again used to identify the best cut-off point.

The t-tests for males aged 15 to 17 revealed that the proximity of a neighbourhood to male adolescents is in fact negatively related to burglary rates –all else being equal (see Figure 3.4). Several reasons are possible for this relationship. First, this particular age cohort is still living with their parents. Thus they are likely to be located in family-oriented communities. In this city, these areas tend to be socially cohesive and are home to fewer low income people than average. Secondly, dwelling places in family-oriented communities are more likely to be found within the protection of a residential enclave. As outlined in section 2.3.2, burglars tend to avoid these areas. Finally, as was mentioned in section 2.3.1, only 6% of Canadian adolescents are likely to become involved in serious criminal activity of any kind. Therefore, without accounting for other factors, the youth dimension is likely overwhelmed by the socioeconomic and demographic characteristics of the neighbourhoods in which they reside.

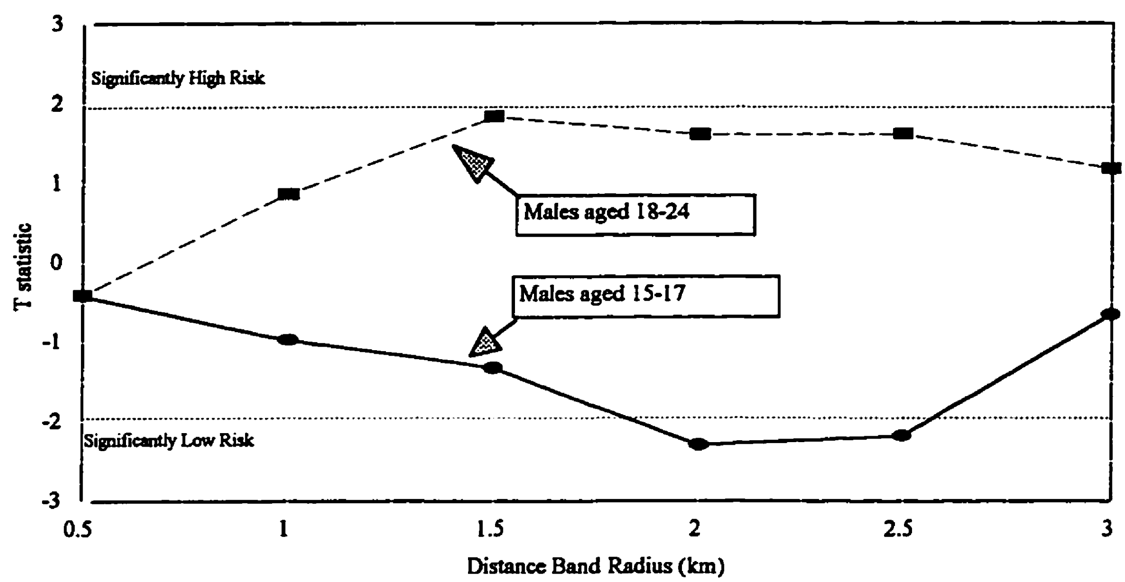


Figure 3.4 T tests on Distance Bands by Extreme Risk Group

The t-tests for males aged 18 to 24 were more revealing. They show that the 1.5km cut-off best emphasizes this dimension of the crime in this particular study area, which corresponds with many of the results found in the literature review. A variable was created measuring the % of residents with 1.5 kilometres that were male aged 18 to 24 (*m1824_15*). The 1 kilometre band was chosen for the males 15 to 17 cohort (*m1517_1*) in anticipation that the introduction of other variables would draw out its true influence. All of the other band size variables were subsequently dropped.

In order to capture the impact of the professional burglars, the distance of each EA's centroid to the nearest highway exit was also calculated (*hwy_exit*). It was expected that a neighbourhood's proximity to a highway increases its accessibility to mobile offenders. This variable was used in an exploratory sense rather than to confirm any pre-existing theory on offender travel behaviour.

3.4.3 Spatial Weights Matrices

Having collected the necessary explanatory variables, there remained the task of quantifying the spatial relationships of the observation units. The geographic analysis of aggregates differs from the traditional ecological approach in its incorporation of spatial relationships. This is done through the creation of a spatial weights matrix, W. This is a square matrix of the same dimension as there are cases. It summarizes the connectivity of all areal units to all others. The relationship of a unit to itself (found on the diagonal) is usually set to zero.

There are a number of ways of abstracting spatial relationships. The most common method is simple binary contiguity. In this approach a "1" is assigned to any neighbourhoods which share a contiguous border and a "0" assigned to all others. This was done by exporting an AAT (Arc Attribute Table) resident to all ARC/INFO layers to SpaceStat's *Weights from AAT* function. This routine reads the AAT information to deduce which areas share a common arc.

A second approach is to export the PAT (Polygon Attribute Table) after reducing it to three columns: Left Polygon Identifier, Right Polygon Identifier and Perimeter Length. This can be read into the *Weights from PAT* function to calculate connectivity as a function of the length of common boundary.

A final approach is to calculate the centroids of each area in a GIS and export these to SpaceStat. The statistical software's *Create Distance Matrix* routine builds an NxN matrix of distances from all areas to all others. A mask can then be applied to this matrix to convert all distances smaller than a specified cut-off value to "1" and assign a "0" to all others. It was problematic to apply distance measures in this study due to the large EAs located on the study area boundary. The centroids calculated for these areas were located over 2.5 km from their nearest neighbours. Therefore, in order to create a useable spatial weights matrix –where all neighbourhoods are connected to at least one other– some of the smaller, more centrally located EAs would be defined as having 50 or more neighbours.

Contiguity matrices were computed for both binary and perimeter contiguity. However, the perimeter measures were subsequently dropped from the study. Using two different weights matrices to create lagged variables produced an unmanageable amount of data. Furthermore, early exploration of the influence of each approach to measuring connectivity revealed that there was insufficient differences in the results to warrant the use of two separate measures.

3.5 Data Issues

Before the data set could be analyzed there were a number of issues which needed to be resolved. For many of the methods used in the analysis it is necessary to have normally distributed variables. The burglary rate variable follows a highly positively skewed distribution: the 30 hardest hit EAs contain 10% of the study area dwellings but 24% of the burglaries. This problem is to be expected as "the spatial analysis of proportions is hindered by the intrinsic non-normality and heteroscedasticity of these variables (since they can in fact be conceived as parameters of a binomial distribution)" (Anselin, 1992: 17). Furthermore, many areas (12.2%) reported no break-ins in 1995. In this case a log or square root transformation is not applicable to create a normal distribution.

A Freeman-Tukey transformation was felt to be a suitable approach. This incorporates the original count data as well as the denominator. The formula is

$$Y_i = \sqrt{\frac{1000\ C_i}{N_i}} + \sqrt{\frac{1000\ (C_i+1)}{N_i}} \tag{5}$$

where Y_i is the transformed burglary rate, C_i is the count of burglary events in a given EA, and N_i the number of dwellings. This transformation acts to both normalize the data and remove the dependence of the variance on the mean. Summary statistics and histograms reveal this to be a successful conversion as the burglary rate now approaches normalcy (see Appendix A). The new variable was labelled *ft_burg*.

In order to deal with heteroscedasticity it is possible to use weighted least squares, which downweights the contribution of units with a small denominator (in this case, number of dwellings). However, this method was not used as it means additional complexity in the interpretation of the spatial models. Furthermore, the use of the Freeman-Tukey transformed burglary rate was felt to be sufficient to account for the dependence of the variance on the mean.

A second issue is the number of EAs with missing census data. Many of the smaller EAs lacked data for three census variables used in the analysis: 15 (5%) lacked data for the average dwelling value variable; an additional 11 EAs (4%) lacked data for this variable as well as average income for males over 15, and % of males over 15 years earning less than $15,000. It was necessary to estimate these missing values based upon their relationship to complete variables, as well as to neighbouring EAs with known values.

The first attempt to estimate the missing values involved using least squares regression. However, such global interpolation was confounded by the fact that most of the EAs with missing values are apartment buildings, which are likely to be substantially different than the "average" neighbourhood. This also affected the estimation of an autoregressive parameter to improve the model. Regression was further hindered by the lack of correlation between the complete and incomplete variables.

An alternative approach to statistical interpolation involved developing a more intuitive estimate of the missing values (see Figure 3.5). First the EA of interest was designated as either primarily apartment or non-apartment. If it was an apartment and there were other high rise EAs with known values within a kilometre radius, then these values were used to estimate the missing variables. If there were no high rise neighbours within a kilometre radius, the missing variables were estimated as the weighted average of their contiguous neighbours multiplied by constant. The constant was calculated as the global average for apartment EAs divided by the global average for non-apartment EAs. The constants for average income for males over 15, % of males over 15 years earning less than $15,000, and average dwelling value were 0.774, 1.134, and 0.818 respectively. The local estimates were then added to the least squares estimate, and the sum divided by two to give a final value.

The significant difference in the estimated housing values between the OLS and local approach is shown in Table 3.3, indicating that the interpolation process was not very successful. However, considering the relatively small number of EAs with missing values, this estimation was felt to be a superior alternative to simply excluding these areas.

Table 3.3 Interpolation of Missing Dwelling Values

EA	APT (Y/N)	NEIGHBOURS	W_APT $	W_NAPT $	INTERPOLATION	LOCAL EST	OLS EST
9316	Y	NON-APT	-	112154	(W_Avg * 0.889)	100	68
38021	N	NON-APT	-	150231	W_Avg.	150	130
38063	Y	NON-APT	-	179934	(W_Avg * 0.889)	160	121
38104	Y	APT	120000	-	W_Avg	120	66
38107	Y	APT	94264	-	W_Avg	94	88
38162	Y	NON-APT	-	122539	(W_Avg * 0.889)	109	57
38169	Y	NON-APT		122539	(W_Avg * 0.889)	109	56
38212	Y	NON-APT		168746	(W_Avg * 0.889)	150	140
38251	N	NON-APT		171337	W_Avg	171	167
38309	Y	NON-APT		136762	(W_Avg * 0.889)	122	109
38371	Y	NON-APT		189428	(W_Avg * 0.889)	168	126
89022	Y	NON-APT		114760	(W_Avg * 0.889)	102	95
89057	Y	NON-APT		88872	(W_Avg * 0.889)	79	94
89212	Y	NON-APT		311785	(W_Avg * 0.889)	277	108
89213	Y	NON-APT		311785	(W_Avg * 0.889)	277	114
89316	Y	NON-APT		204245	(W_Avg * 0.889)	182	80

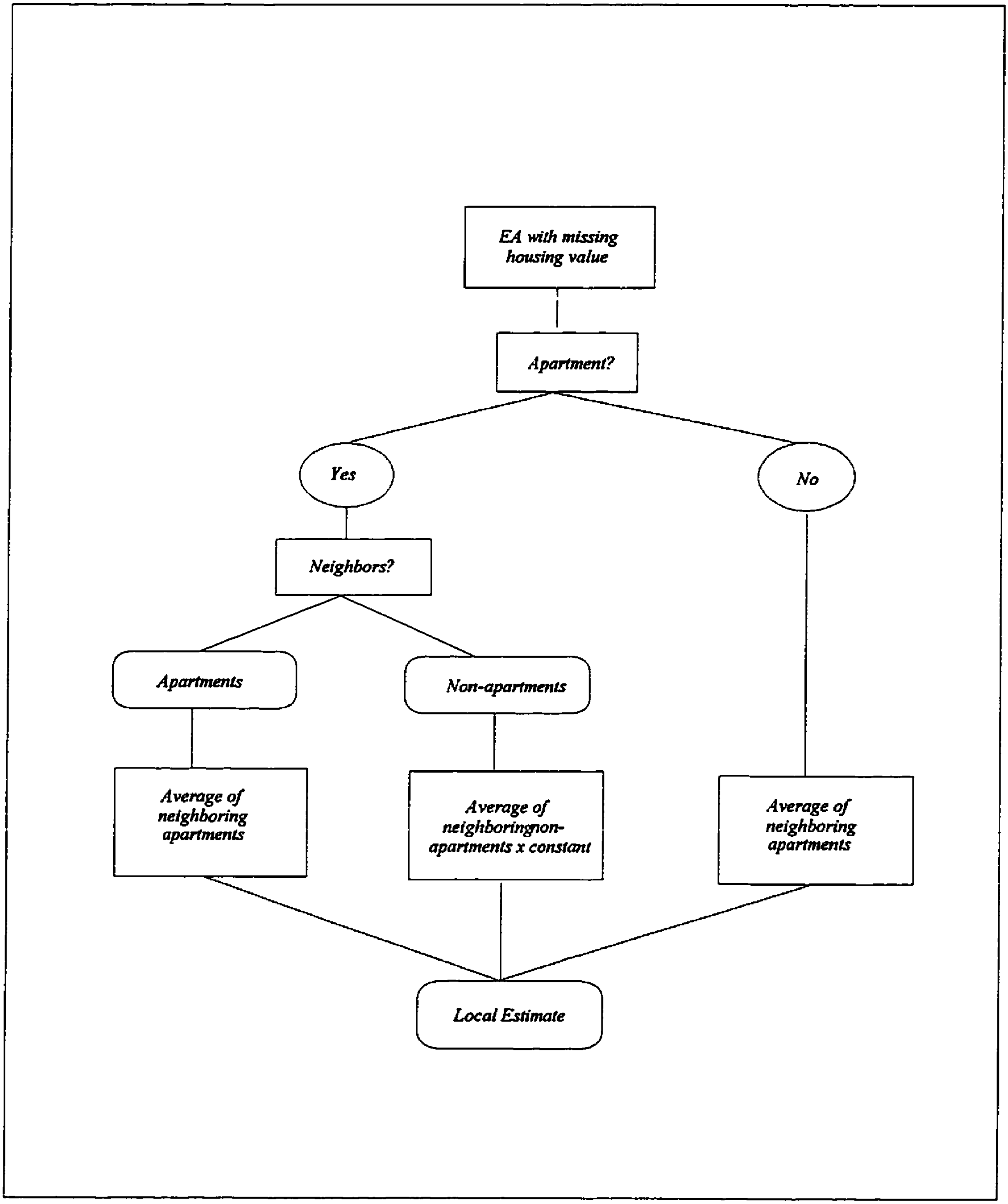


Figure 3.5 Interpolation Method Used for Missing Values

CHAPTER 4: ANALYSIS AND INTERPRETATION

This chapter narrates the statistical analysis of the data set described in the previous chapter. Three separate exploratory analyses were run before the final confirmatory models were constructed. The first phase involved a traditional ecological study which includes the use of the spatial variables measuring access as well as lagged variables. The second phase involved identifying significantly high or low risk EAs, based upon the number of households they contain. These extreme risk groups (ERGs) were then examined in a bivariate and multivariate analyses to determine the explanatory variables which best differentiated between these two extremes. This was done as an alternate form of analysis, and also to detect any missing variables important to the regression models. The third phase involved identifying the underlying spatial structure of the data set through applying spatial statistics as well as partitioning the data set into a series of spatial regimes. This culminated into the confirmatory analysis, which includes the selection of formal spatial regression models, analysis and residual analysis.

4.1 Exploratory Spatial Data Analysis

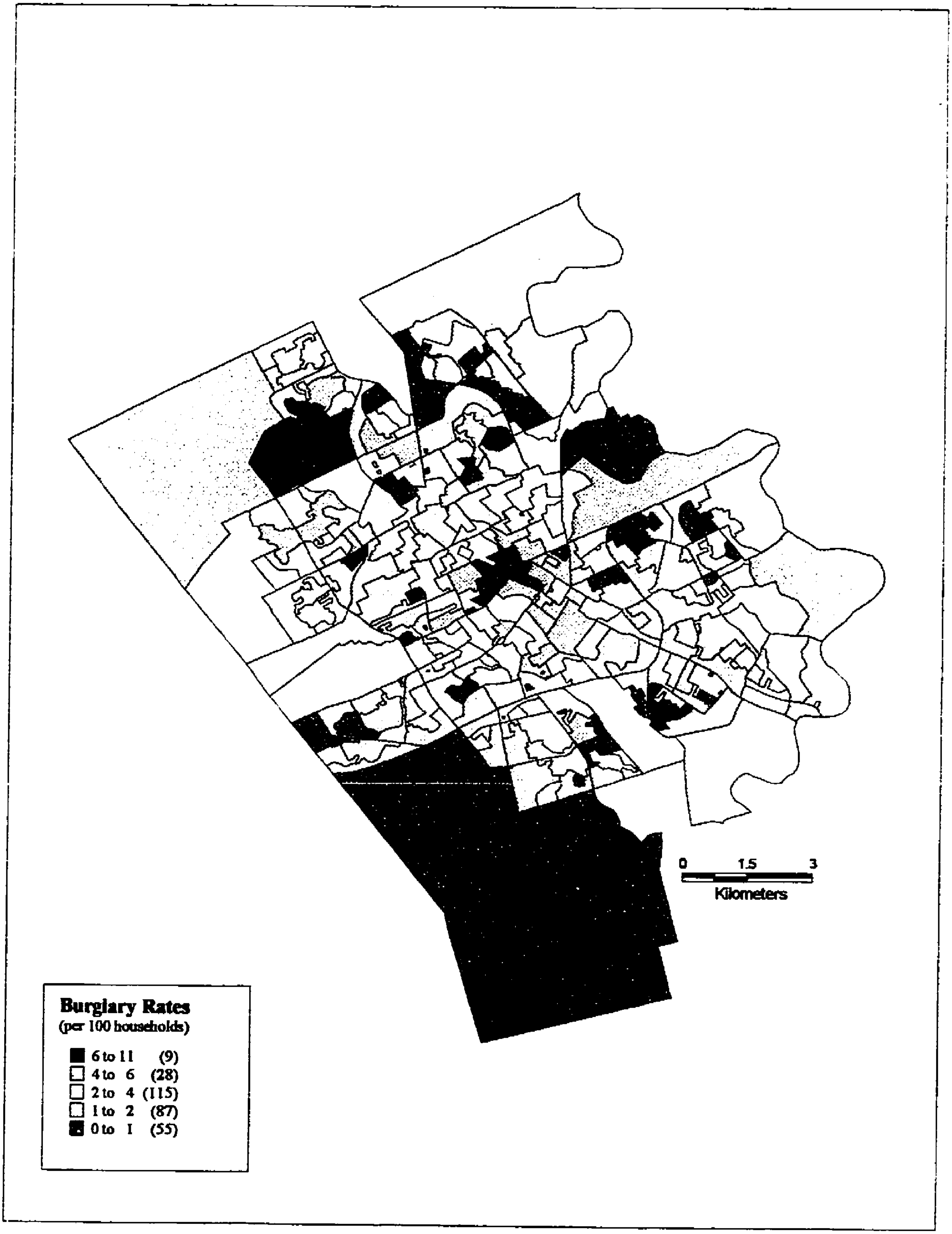
4.1.1 Ecological Analysis

The first step taken in the exploratory spatial data analysis was a simple aspatial summary of the data set. Descriptive statistics were compiled to create a brief summary of the matrix (see Appendix B). These aspatial measures were used in conjunction with histograms to determine if any of the variables required transformation for the parametric tests used in the analysis. In order to have normally distributed data many of the variables underwent a log transformation. For several other variables, such as *rented* and *mjrd_100*, it was not possible to identify a suitable transformation, and so they were left in as is. A compromise had to be made between having ideally distributed data and abstracting the meaning of the numbers by transforming them to new values.

The burglary rates were then mapped (see Figure 4.1). It is evident that there are two clusters of EAs with high burglary rates: a central cluster concentrated around the intersection of King and Victoria streets, and a suburban cluster in the northwestern section of the study area (Lakeshore Village and Erbsville). There are also a number of other high risk areas distributed throughout the southeastern quarter of the study area (Rockway, Centreville and Kingsdale West), although they do not exhibit the same degree of clustering. EAs with exceptionally low burglary rates are dispersed throughout the entire study area (Bridgeport, Colonial Acres, Westheights, Williamsburg, Heritage Park and Rosemount). Several other low-risk EAs were located within the central high risk area. The generally clustered distribution of burglary at this scale suggests that this is a non-random event, and hence is dependent upon spatial and structural variables –not simply the population distribution.

Two statistical techniques common to traditional ecological studies are bivariate correlation and multiple regression. The ecological analysis stage of this study involved using these two methods to model the variation of this crime. A correlation matrix was created to look at the bivariate relationship between all of the variables (see Appendix C). It should be noted that this is a measure of the linear relationship between two variables, which may not necessarily be the case. Therefore scattergrams were also used to determine any other kinds of bivariate relationships at work. Only one relationship between the explanatory variables and the dependent variable revealed any non-linear trend (*km_cbd*). This relationship is discussed further on in this section.

Figure 4.1: Burglary Rate Map



The burglary rate was correlated with other variables in both its transformed and untransformed state in order to identify any significant changes in relationships which may result from the modification of the dependent variable. None were found.

From the correlation matrix it becomes clear that several of the variables are highly correlated. Average number of maintainers, average household size, and households with one person are all correlated at over $|0.80|$. Furthermore, they are also highly correlated with average number of bedrooms, a measure of affluence. A decision was made to drop average household size and average number of maintainers from the analysis. These two variables are measures of occupancy which are already represented by the infant and retired variables. Including them in the subsequent multivariate analysis would introduce an unacceptable degree of multicollinearity. By maintain bedrooms and households with one person, a greater amount of information was retained because they are measures of affluence and instability, respectively.

From the correlation with *ft_burg*, we can see that for the entire study area the transformed burglary rate is associated with offender, opportunity and access variables. Areas with larger, more valuable homes were found to experience significantly higher burglary rates, as were those with many dwellings in need of repairs. Other significantly positively correlated variables are male population over 15 years earning less than $15,000 per annum, and the proportion of the population within 1.5km of the centroid who are male aged 18 to 24. Burglary is negatively associated with areas containing higher population densities, newer housing, rental housing, one person households and persons older than 65. However, with the exception of population density, none of the above mentioned variables hold a strong linear association with B&E rates.

A stepwise ordinary least-squares regression ran with all the variables identifies two significant variables:

$$\underset{}{ft\ burg} = \underset{(0.00)}{0.675} - \underset{(0.00)}{0.123 lnewdens} + \underset{(0.00)}{0.005 wloneprnt} \qquad \textbf{(6)}$$

In the regression model the log of the population density remains significantly negatively related to burglary rates, while a previously insignificant variable, the lagged version of the proportion of households headed by one parent, becomes significantly positively related. From this equation, we can infer that the density of the immediate physical environment acts as a significant deterrent to burglars, and having accounted for density, areas with less stability (measured as those with many single parent families) "export" burglars to nearby sites.

The model fit (adjusted R^2 of 0.20; F statistic of 36.74) is fairly weak, particularly when one considers that the inappropriate use of OLS with such data tends to inflate goodness-of-fit statistics. The residuals follow the same spatial pattern as burglary rates in general: a cluster around the downtown, along with a loose confederation of EAs on the cities northwest and southern fringes. Consequently, they show a very high level of spatial dependency: the standardized score for the Moran's I is 4.64; the Lagrange Multiplier (error) 19.82, and the Lagrange Multiplier (lag) 20.5. The extent of the autocorrelation in the residuals indicates either missing variables or an unaccounted for spatial structure in the data set. Clearly the traditional ecological analysis is insufficient and inappropriate for this study. The next phases in the analysis are aimed to improve upon these results.

4.1.2 Extreme Risk Group Analysis

The police data set contained records of the 1626 residential burglary events occurring among the 88397 potential targets (dwelling places) within the study area. This results in an overall rate of 1.84% –a relatively rare event. When rare events are being studied and/or populations are highly variable, it is useful to convert the data to Poisson probabilities (Bailey and Gatrell, 1995: 302). This method involves calculating whether a given enumeration area has a significantly high or low burglary rate compared to the expected number for an EA with that many dwellings following a Poisson distribution –the most appropriate for rare event data (Hirschfield et al, 1991: 163). The formula for converting rates to Poisson probabilities is as follows:

$$p_i = \begin{cases} \sum_{x \geq y_i} \dfrac{u_i^x e^{-u_i}}{x!} & y_i \geq u_i \\ \sum_{x \leq y_i} \dfrac{u_i^x e^{-u_i}}{x!} & y_i \leq u_i \end{cases} \tag{7}$$

where u_i is the Poisson mean value, x the observed number of events, and e the base of the natural logarithm.

The Poisson scores were used to delineate high and low risk areas based upon those areas which were over 0.95 in significance or below -0.95. There were 47 low EAs and 46 highs. An ordinal variable was subsequently created, identifying group membership. The high and low EAs were labelled extreme risk groups (ERGs) and separated for further analysis.

First the ERGs were mapped (see Figure 4.2). It is clear that while the probabilities follow the same general pattern of the rates, there are important differences. There are a number of EAs which transform either from an average rate to an extreme high or low probability; or, conversely, which shift from a seemingly high or low rate to an unexceptional probability. The most significant change appears to be in the area around the central core, where the clustering of high risk probabilities is even more pronounced than high risk rates.

The second step of the analysis of the ERGs involved implementing an independent samples t-test to determine which variables differed substantially between each extreme group. Table 4.1 summarizes the results. The highly significant (signif < 0.05) variables were *bedrooms*, *lnewdens*, *yearblt*, and *repairs*, as well as the lagged variables *wloneprt* and *wyrblt*. *Industry* was the least significant variable in the t-tests. However, it was more appropriate to run a Chi-square test on this binary variable. The test did not reveal any significant relationship between risk group membership and the presence of industry.

These results point to a number of relationships at work, but one must bear in mind that they are examined in isolation, without accounting for the influence of other variables. The t-tests suggest that high risk areas are different from low risk areas in that they have significantly more bedrooms per dwelling, are of lower density and contain a greater proportion of older and dilapidated housing. They also tend to neighbour areas with older, dilapidated housing, as well as lone parent families. Interestingly, unlike the lagged version of *loneprnt*, the unlagged variable is entirely insignificant (significance of 0.97), which suggests that offenders are travelling a few blocks from home before selecting a target.

As a follow-up to the t-tests, a discriminant analysis was run to look at the impact of the individual variables when acting in concert. The primary objective in running this ESDA stage was to study the spatial distribution of misclassified ERGs in order to determine *a priori* any explanatory variables missing from the final analysis.

Figure 4.2: Extreme Risk Group Map

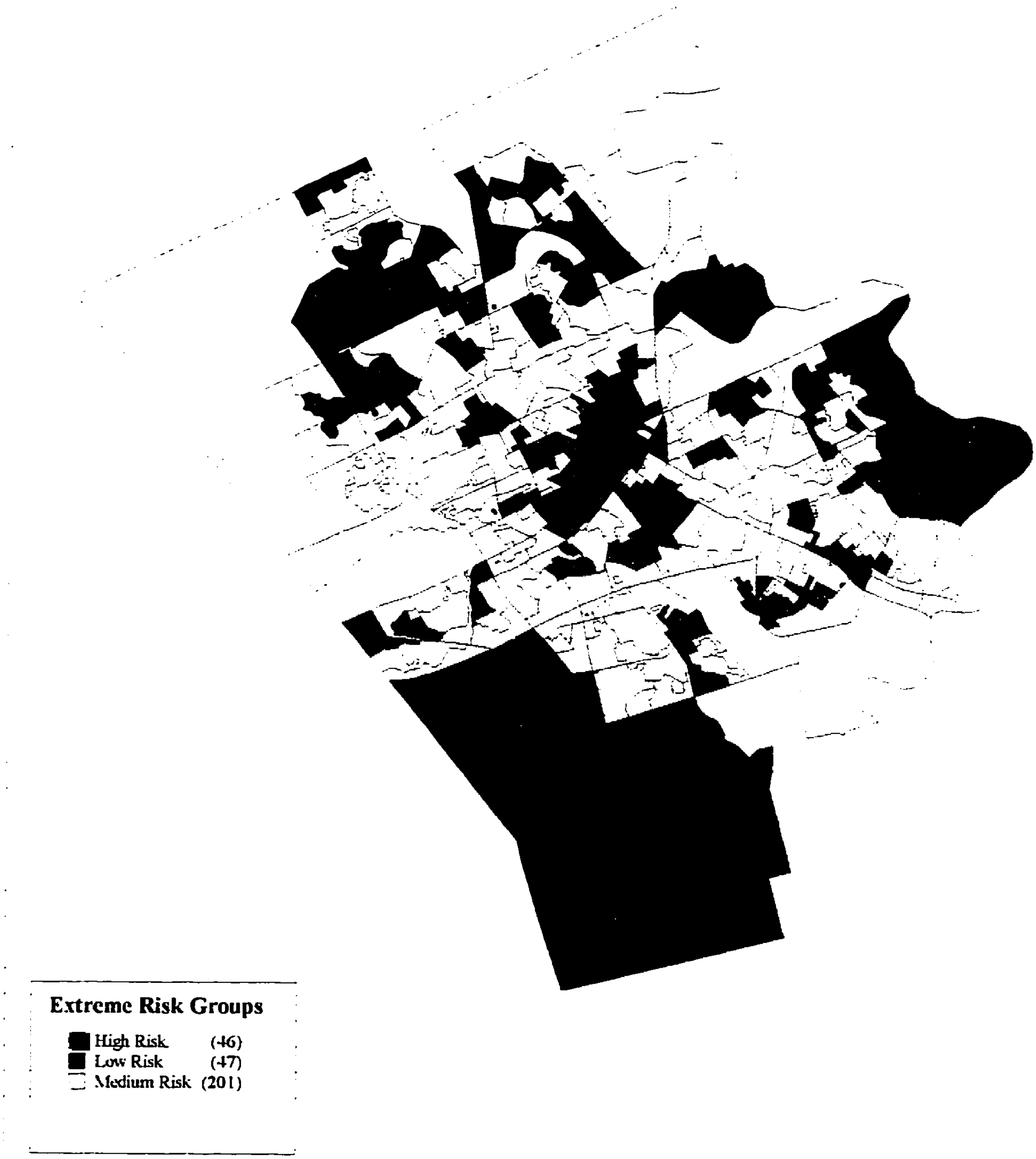


Table 4.1 T test and Discriminant Analysis Results

VARIABLE	BIVARIATE (T-test Statistic)	MULTIVARIATE (SCDF Coefficients)
Offender		
wloneprnt	2.66	.60
Opportunity		
bedrooms	2.07	.74
lnewdens	-2.80	-.53
repairs	2.98	
wrepairs	2.78	
wyrblt	-2.59	
yearblt	-4.12	-.43
Access		
mjrd_100		.45
m1824_15		.40

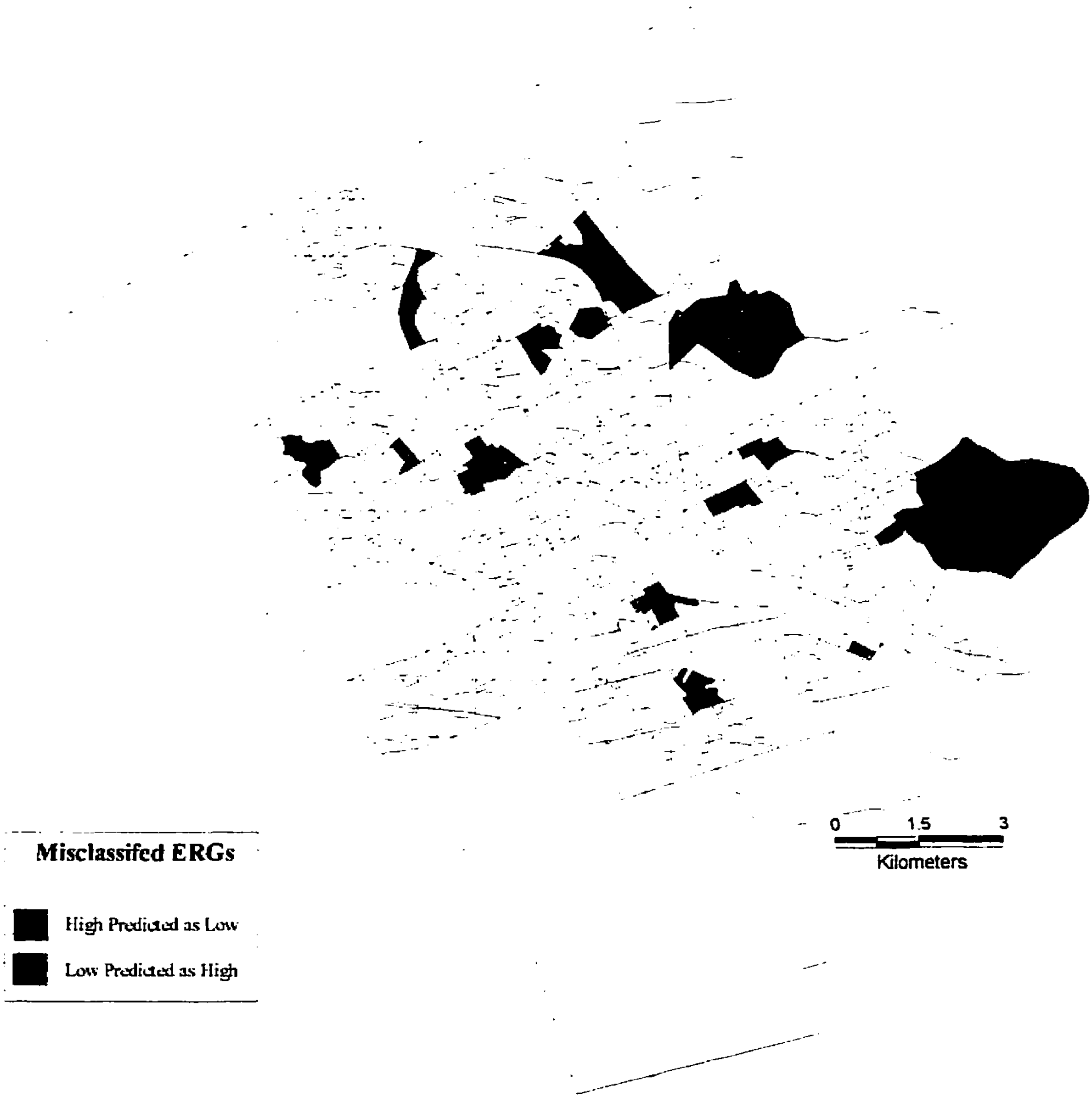
A stepwise model using the Wilks' lamda method was used, with prior probabilities assuming all groups as equal (which, in this case, they essentially were). The covariance matrix chosen was within groups. The canonical discriminant functions evaluated at the group means (group centroids) were -.8516 for the low risk areas and .8701 for the high risk areas. From the standardized canonical discriminant functions (SCDFs) we can see that high risk areas are different from low risk areas in that they contain older, larger dwellings, built at lower densities, located close to a major road and a cluster of males aged 18 to 24. The classification accuracy was 78.5%. This is a modest result considering that by chance alone it would be 50%.

Including lagged variables increases the canonical correlation coefficient from .6065 to .6565 but without a corresponding increase in the per cent correctly classified. This suggests that the existing misclassified EAs are distinctive and that the addition of the lagged variable (*wloneprt*) simply further underscores these differences. The misclassified EAs were mapped to uncover any spatial patterns which may provide clues to their uniqueness (see Figure 4.3). There does not appear to be any obvious pattern as both types of misclassified EAs were found in non-clustered arrangements across the study area in a variety of neighbourhoods.

Another attempt to determine the cause of the DA misclassification involved comparing the misclassified ERGs with the spatially partitioned OLS models' residuals (the topic of the next section). The models' residuals were standardized into Z-scores and compared to the two forms of misclassification (see Appendix D). These cases were not overly influential in the regression models suggesting that if there is indeed a missing variable it does not have a profound influence on the ecological analysis of the EAs.

A final application of the ERGs was to further study the relationship between distance from the CBD and burglary rates –the gradient hypothesis. Whereas the correlation coefficient of -.01 suggested no relationship at work, a scatterplot of *km_cbd* and *ft_burg* did reveal a "v" shaped curve. The EAs were divided among concentric zones from the city centre delineated as 0-2km, 2-4km, 4-6km and 6+km (see Figure 4.4). A Chi-square test was run to see if the relationship depicted in the box plot follows a distinct pattern.

Figure 4.3: Misclassified Extreme Risk Group Map



The resulting χ^2 statistic of 50.54 is highly significant, proving that the zone-risk relationship deviates from a random situation. Further study reveals that high risk areas are substantially over represented in the inner core (0-2km) and the periphery (6+ km), further suggesting a non-linear relationship between distance to CBD and burglary rates. This finding clearly shows that any model constructed of the data would have to incorporate spatial complexities not captured by the gradient hypothesis.

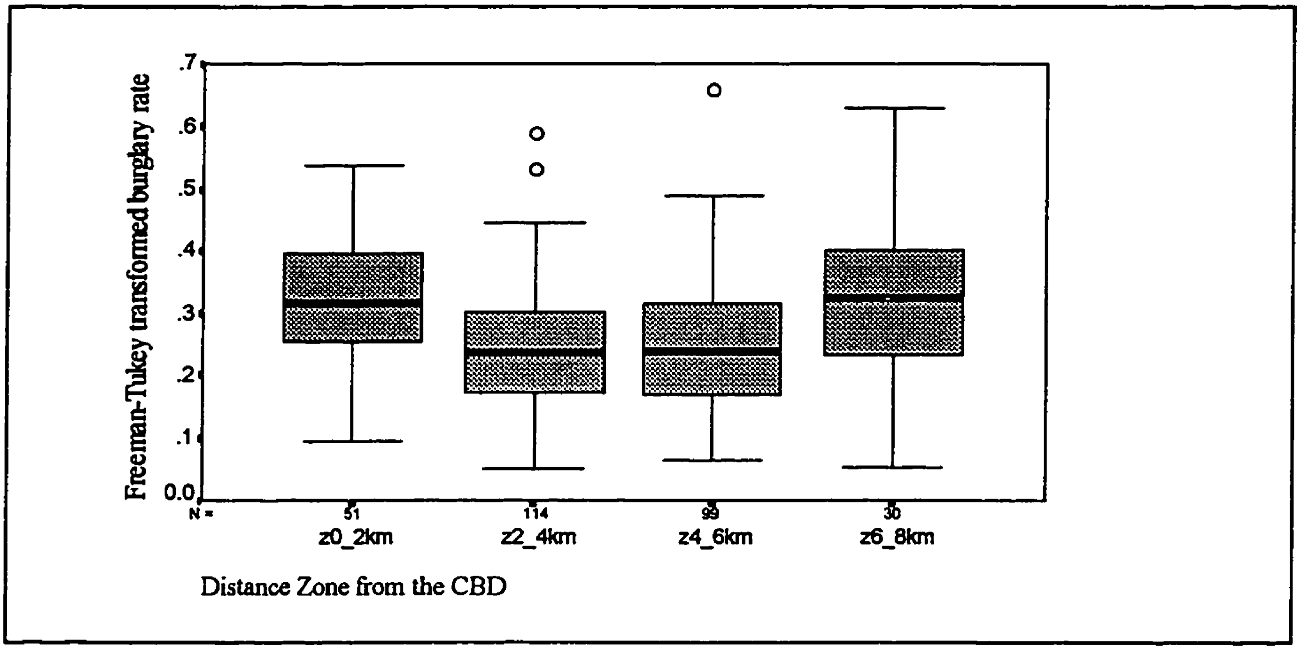


Figure 4.4 Boxplot of Transformed Burglary Rate by Zone

4.1.3 Spatial Analysis

The first two stages of the exploratory analysis show that only a modest result can be achieved using a traditional ecological approach supplemented by spatially modified variables. Furthermore, that the model's weakness cannot be attributed to a missing variable. The significant autocorrelation of the least squares model's residuals indicates that the spatial structure in the data set has not yet been properly accounted for. The purpose of the third exploratory stage is to further study this structure. It is comprised of descriptive spatial statistics and the exploration of spatial regimes.

Descriptive Spatial Statistics

The first statistical exploration of the burglary variable's spatial distribution involved finding a global trend in the burglary variable. This is commonly done by fitting a trend surface model to the data:

$$Z = \alpha + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy \; ... + \epsilon \qquad (8)$$

This is similar to an ordinary least squares regression model except that the zonal centroids (the x and y variables) are used to create a polynomial surface of the data. This is not entirely appropriate for areal data, as it assumes that no variation occurs within the areal unit. Nor does it account for the influence of the variation in areal unit size –which is quite significant in this study.

An alternative way to detect a first order trend in areal data is to perform a window average transformation of the original data values. The data is smoothed by reassigning the value of a particular area by a weighted average of itself (x_i) and its neighbours (x_j):

$$(x_i + \Sigma_j w_{ij} x_j) \; / \; (1 + \Sigma_j w_{ij}) \qquad (9)$$

The weights matrix (w_{ij}) is row standardized when performing this operation. While this does not provide a concise numerical summary of the data or any goodness-of-fit statistics, it does serve to clarify the spatial distribution of the variable of interest. The window average method has the advantage of smoothing data which may be unreliable due to a small number of counts (such as burglary events). From Figure 4.5 it becomes clear that there are three distinct high risk areas in the study area: the northwestern suburban neighbourhoods, the central core and, to a lesser extent, the southeastern suburban belt.

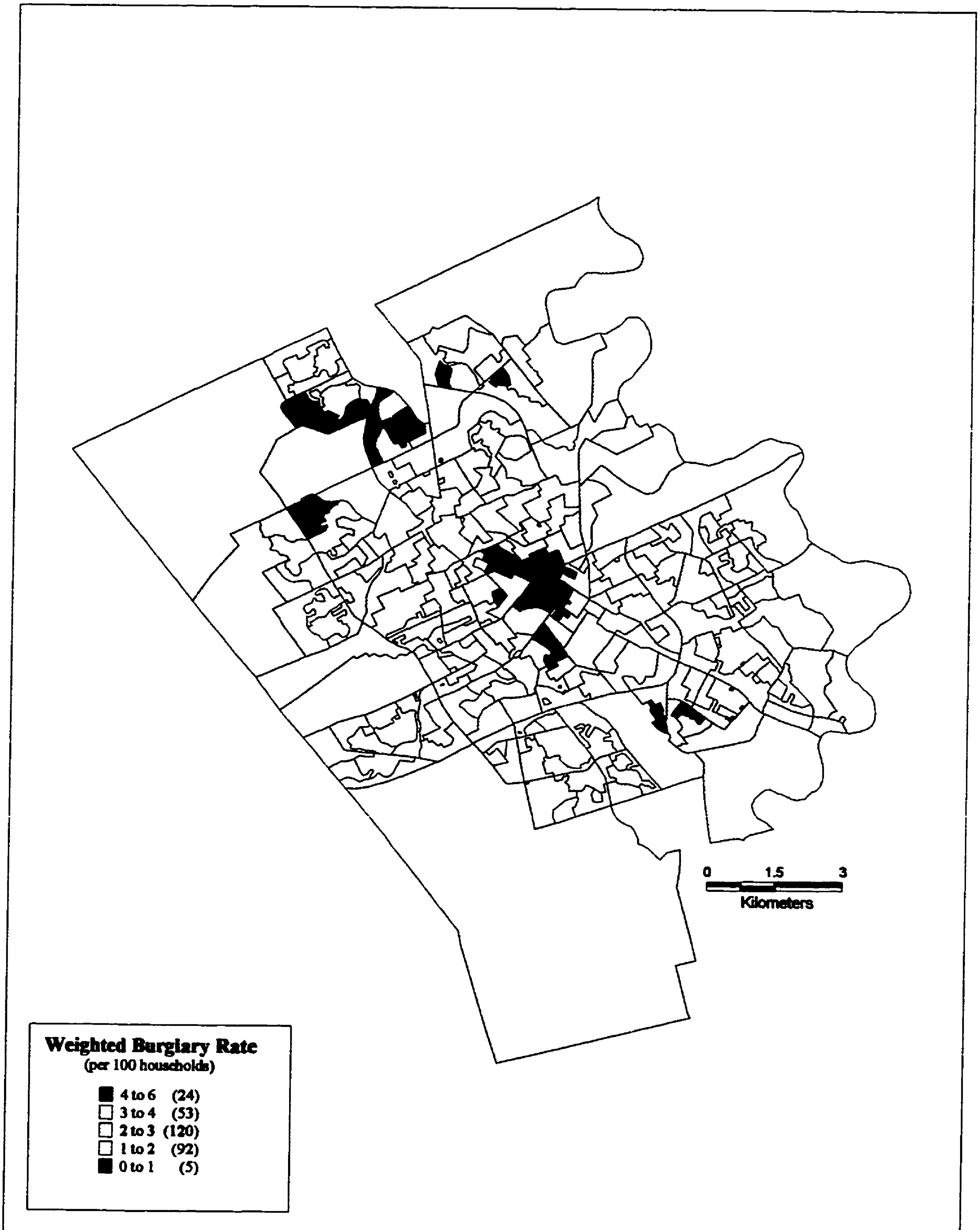
A further exploration of the spatial distribution of the dependent variable is to study the extent of the second order spatial effects (or spatial autocorrelation). This can be done by measuring autocorrelation at a global scale in conjunction with a local measure of association. These two measures, combined with the knowledge of the global trend, provide information on the scale at which burglary occurs.

Correlograms were created for all of the variables was made using the Moran's I statistic:

$$\frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})\ (y_j - \bar{y})}{\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right) \left(\sum\sum_{i \neq j} w_{ij}\right)} \tag{10}$$

where w_{ij} is the connectivity of area i and j, y_i and y_j are observations for areas i and j, the mean is $\bar{y}$. This formula assumes that a row standardized weights matrix is being used. A statistic was calculated up to nine lags, the furthest possible distance. The procedure used in this study was to assume each variable followed a normal distribution (which, asymptotically is a safe assumption when dealing with over 30 observations). A second global measure of autocorrelation, Geary's c, is very similar to Moran's I and considered a weaker approach, so was not calculated.

Figure 4.5: Window Average of Burglary Rate Map



A number of interesting patterns are revealed in the correlograms (see Appendix E). The *year_blt* variable is an example of a variable following an isotropic global trend. EAs are significantly positively correlated with other EAs up to four lags away, and then negatively related to EAs from six to nine lags away. Conversely, the *lnewdens* variable shows a much weaker level of auto-correlation and only to a distance of two lags, suggesting more of a patchwork distribution. Burglary rates fall somewhere in between these two extremes, revealing a structure of significant positive auto-correlation over several lags. This suggests that EAs generally experience similar burglary rates to their neighbours. This could be caused by the general trend in the data shown in Figure 4.5, but the tailing off after three lags indicates the presence of heterogeneity rather than a broad regional trend (Bailey and Gatrell, 1995: 273). This is likely caused by the journey to crime spatial process which describes the short-distance travel behaviour of most offenders.

Local indicators of spatial association (LISAs) provide observation-specific measures of spatial autocorrelation. These allow for global measures such as Moran's I to be decomposed, providing a useful means to finding significant pockets of local nonstationarity (Anselin, 1995: 98). They are also useful for detecting the influence of spatial outliers (areas which are significantly different from their immediate neighbours).

Using the local Moran, it was possible to identify areas which are surrounded by a cluster of similar values (high local Moran value), as well as areas surrounded by dissimilar areas (low value). This allows one to see if the degree of autocorrelation is stationary over the study area. If this is not the case, it is likely that the study area comprises discrete spatial regimes. The Local Moran's I statistic is calculated as follows:

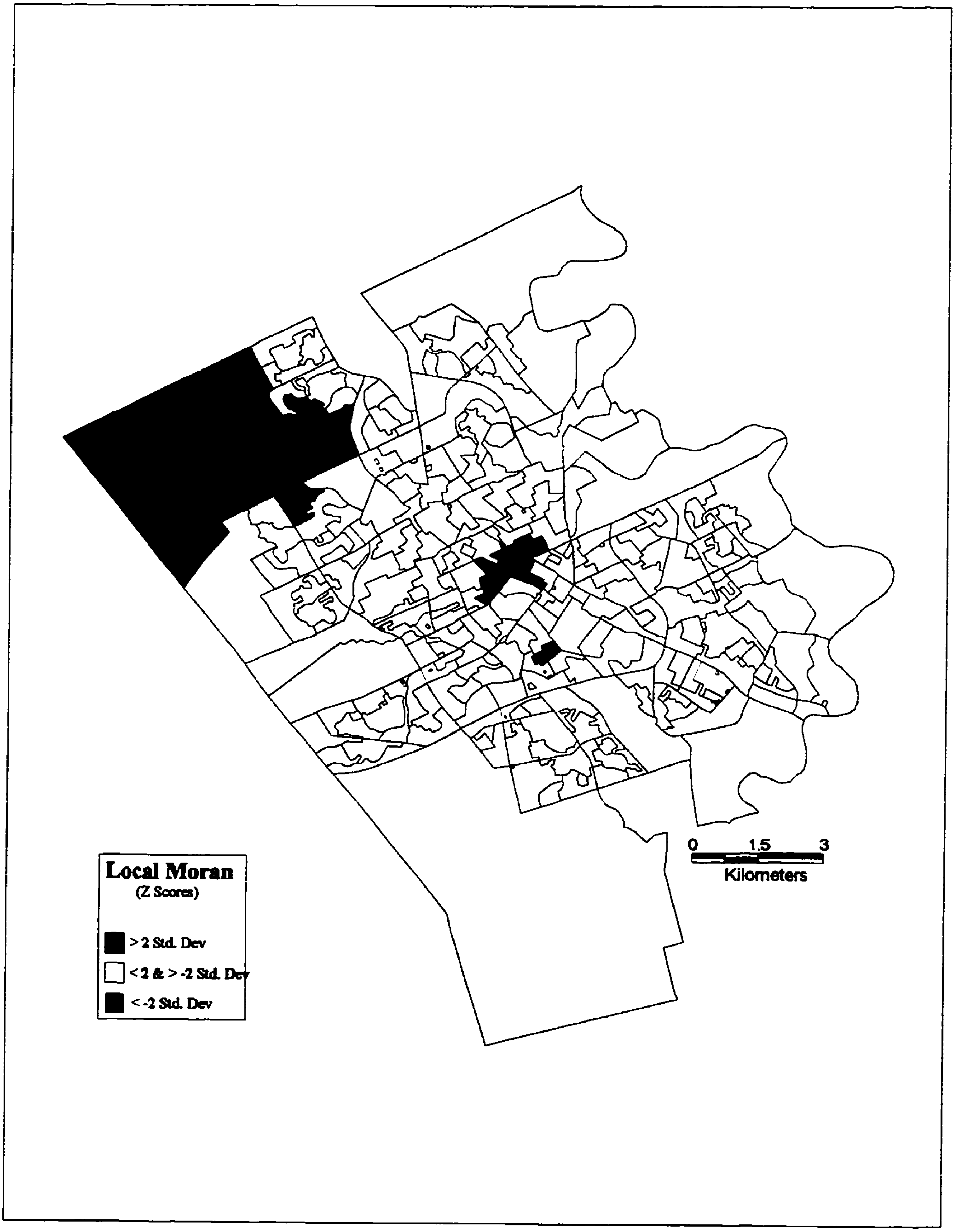
$$I_i = z_i \sum_j w_{ij} z_j \tag{11}$$

where the observations z_i, z_j are in deviations from the mean. Only neighbouring values $j \in J_i$ are included in the calculation of the statistic. The weights matrix used is generally row-standardized to facilitate interpretation. By further calculating the expected value and the variance it is possible to calculate standardized Z values of local association.

The local Moran was calculated using the original burglary rate data. These values were then mapped to visualize the distribution of association across the study area. From Figure 4.6 it is clear that central neighbourhoods are surrounded by areas of similar values, while in the rest of the study area there are neither any extensive areas of similarity or dissimilarity. An important exception is the suburbs in the northwest corner which do exhibit a degree of clustering of similar values.

The EAs comprising the northern edge of downtown Kitchener are very high risk areas, with burglary rates generally double or triple that of the study area's mean. From the window average map we can see that they form the nucleus of a large central area of steadily declining risk. From these results we can infer that the high level of spatial association is that between similar high risk neighbourhoods. This is also the case –although to a lesser degree– in the northewestern corner of the study area.

Figure 4.6: Local Moran of Burglary Rate Map



Conversely, areas of dissimilarity were few in number. In the Centreville district there is some juxtapositioning of high and low risk areas. Also, to the west of the downtown core there are sharp breaks in the degree of association between EAs. However, the exploration of the dependent variable's spatial association did not suggest any extreme spatial structural changes.

When one considers the correlogram results which indicate a three lag radius of spatial autocorrelation, and the lack of pockets of non-stationarity, the evidence suggests that the study area is comprised of generally smooth trends between a series of local peaks, with few areas of rapid change. Thus there are distinct regimes in the data, but no sharp breaks separating them. Instead, the inner and outer high risk areas are separated by an area of average rates which exhibit neither positive or negative autocorrelation. However, it should be noted that the local Moran is unable to effectively measure the level of association between "average" neighbours (in terms of the variable of interest). The next section incorporates the explanatory variables to further explore the spatial complexities of the data set.

Spatial Regimes

The burglary rates' non-linear relationship with distance to the CBD suggested that spatial heterogeneity was present in at least this variable's relationship with burglary. The possibility of a similar shift occurring with the other explanatory variables was therefore explored through the partition of the study area into zones. This exercise indicated the most appropriate course of action for the final confirmatory analysis.

The observations were divided between those EAs located within a four kilometre radius of the CBD and those outside. This particular radius was chosen because the resulting delineation divided the total number of dwellings in the study area roughly in half. Pearson's correlation statistics were then calculated separately for each zone between the transformed burglary rate and all of the explanatory variables. These results were compared with each other, as well as to the corresponding measure for the entire study area (see Table 4.2).

Table 4.2 Significant Correlations by Spatial Regime

VARIABLE	ENTIRE AREA	INNER 4KM	OUTSIDE 4KM
Offender			
hhsz1			-.280
lmavginc		-.178	.286
lt15000	.139	.186	
wloneprt		.177	.213
Opportunity			
avghhsz			.226
bedrooms	.174		.392
industry	.120	.181	
lnewdens	-.397	-.292	-.490
logvalue	.154		.358
rented	-.135		-.271
repairs	.119	.297	
retired	-.137	-.167	
wbedroom		-.210	
wlnewdens	-.141		-.347
wlvalue		-.205	.194
wrepairs	.125	.261	
wyrblt	-.115	-.198	
yearblt	-.216	-.364	
Access			
hwy_exit	.127		.213
km_cbd		-.316	.320

A substantial change occurs between each zone in the significance and even direction of many of the correlation measures. Specifically, we can see that within the inner zone, high risk EAs contain smaller, older, cheaper and more dilapidated housing and tend to border or contain industrial areas. They are home to higher numbers of lower income males and single parent families, and fewer retired people. A significant negative relationship between distance to the CBD and burglary rates is present in this area.

The relationship of burglary to the lagged explanatory variables reveals further patterns. First *lt15000*, which measures the proportion of males over 15 years of age earning less than $15,000 per annum, is significant in the inner zone but the lagged version of this variable is not. Secondly, it appears that affluent central EAs, particulary those within an enclave of similar areas, experience lower burglary rates. Thus there is strong evidence to suggest that within four kilometres of the downtown, burglary is a very localized event with low income people burglarizing their low income neighbours.

Within the outer zone an entirely different pattern emerges. High risk EAs located further than four kilometres from the CBD tend to be affluent areas and/or areas with large houses and large households. Furthermore, a *positive* relationship exists between burglary rates and distance to the CBD within the suburban belt. Unlike the inner zone EAs, offender variables are not significant, suggesting that burglars are travelling from other areas. The spatial pattern which emerges points to the work of the more professional burglars who are highly mobile, selective in their targets, and willing to travel a long distance to their target.

The only common variable which the two zones share is population density, which is a significant deterrent in both areas. Thus it is apparent that trying to model the determinants of burglary without accounting for this shift in relationships will result in a poor model. One solution to this spatial heterogeneity may be to account for the distance from downtown before measuring the relationship between a variable and burglary rates. However, the relationship of *km_cbd* variable itself shifts in direction between the inner to the outer part of the study area.

In addition to this evidence of two different types of criminals at work in the study area, it was felt that many of the offender variables likely take on different meanings between the two zones. For example, mapping the *lt15000* variable mentioned previously revealed that there are two distinct clusters of low income men: in the area of Waterloo near to the two universities and secondly in the central core of Kitchener. Clearly there are two types of low income males in the study area: poor inner city dwellers of Kitchener and university students in suburban Waterloo. It is anticipated that the latter group is somewhat less criminogenic (in terms of residential burglary) and so the explanatory power of this variable would be reduced in the outer ring. Other variables which may take on substantially different meanings between the two zones are: males aged 18 to 24 (as well as its spatial variant), average male income, and % houses that are rented. The binary nature of many of the study area socio-demographics further contributed to the case for running a separate analysis for each regime.

A series of OLS stepwise regression models were run to identify the best zonal partition and in order to test the spatial regimes hypothesis in a multivariate setting. A second objective was to select the variables for the final confirmatory analysis.

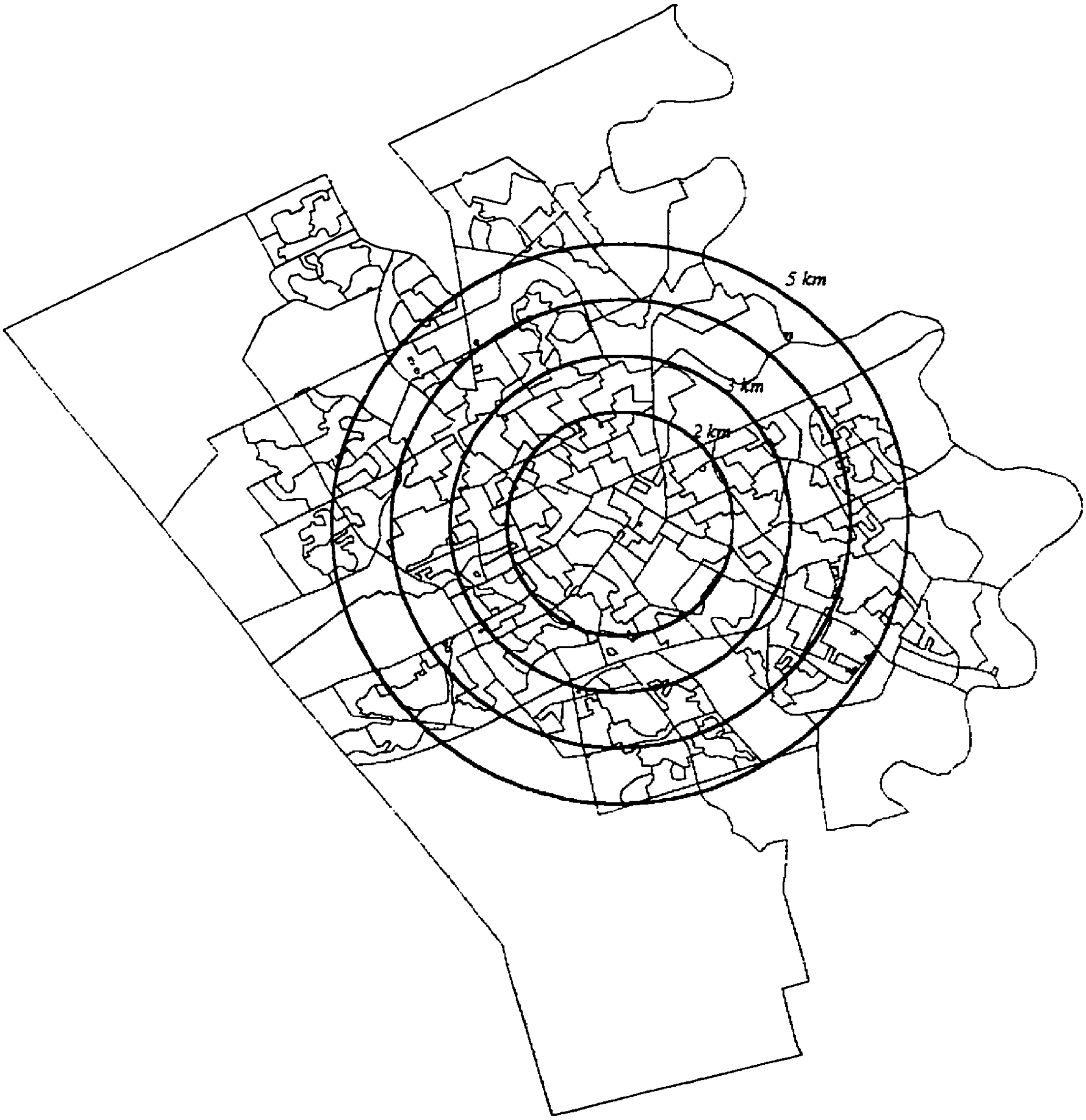
Determining the best partition was done through testing a number of definitions of "inner". The rings were estimated for 2, 3, 4 and 5 kilometres from the intersection of King and Queen streets (see Figure 4.7). The regression results show that the ideal delineation of the two zones appears to be the four kilometre radius originally used for the correlation measures (see Figure 4.8). It is at this point that the OLS models using lagged variables for both sectors are close to convergence and both have a relatively good fit (inner R^2: 0.26; outer R^2: 0.39). The inner core is, however, best fitted when defined as 2 km (R^2 of 0.40), although this figure is based upon only 51 cases. The four kilometre delimitation also has the advantage of fairly evenly splitting the study area, with 49979 dwellings located within the inner four kilometres (56.5%) and 43418 located outside (43.5%).

An exploration into the possibility of a third "transition" zone was made, in part because of the evidence of a smooth transition in burglary rates provided by the spatial statistics outlined previously. However, the results were not dramatically improved. Furthermore, it was felt that this would be an inappropriate course of action because a circular band lying somewhere between the downtown and the suburbs likely does not represent a discrete part of the city. Secondly, this would be a case of fitting the analysis to suit the data.

Upon choosing the four kilometre radius as the final zonal partition, the two zones were compared by their values for the neighbourhood-specific (not spatially adjusted) variables used in this study. A one-way ANOVA was performed to identify significant differences between the inner and outer areas. As is evident in Table 4.3, the inner zone is characterised by many of the characteristics of an inner city –with some important exceptions.

The housing is older, cheaper, smaller and more dilapidated, yet no more densely built than housing in the outer zone. The higher average of the *mjrd_100* variable in the inner zone suggests that there are fewer residential enclaves, a vulnerability opportunity factor anticipated to reduce the burglary rate. Most significantly, the burglary rates for the inner and outer areas show no significant difference.

Figure 4.7: Distance Bands Map



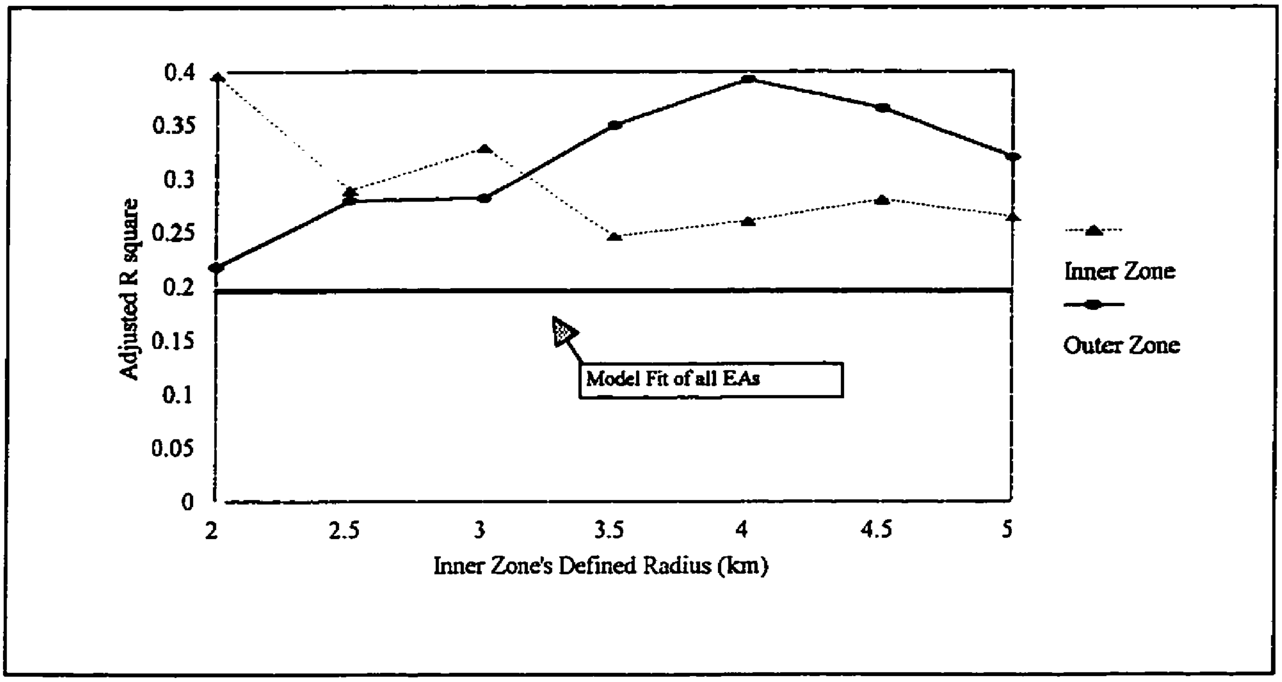


Figure 4.8 OLS Goodness-of-Fit by Zone Definition

The inner zone contains significantly more renters, one person households and older persons; yet no greater a proportion of males aged 18 to 24 than in the outer zone. This is likely due to the presence of two universities located in the northern portion of the study area. There are fewer teenagers and infants in the inner area, yet there are significantly more single parent families, suggesting that the instability factor outlined previously is much more prevalent in the central zone.

The shift in the *lmavginc* variable indicates that males over 15 residing in the inner zone earn significantly less than their outer zone counterparts. Yet the presence of low income men (*lt15000*) is no different between the two zones. Clearly men are less well off closer to downtown, but there is much less variance in income than in the outer area. Again, this may be in part due to the large number of university students who reside in the northern portion of the study area.

Table 4.3 One-way ANOVA Tests of Differences in Zones

Variable	Inner Mean	Outer Mean	F statistic
burgrate	1.86	1.82	-
Offender			
hhszl	27.55	15.50	48.0
lmavginc	4.44	4.52	36.9
loneprnt	14.92	11.43	12.56
lt15000	12.57	12.01	-
m15_17	3.57	4.48	10.24
m18_24	11.64	12.23	-
Opportunity			
avghhsz	2.43	2.93	58.6
bedrooms	2.42	2.91	53.1
infant	6.69	7.60	3.88
lnewdens	3.88	3.87	-
logvalue	5.16	5.27	59.3
rented	48.14	35.67	11.67
repairs	31.00	21.96	30.37
retired	14.09	6.98	33.23
Access			
hi_schl	1.30	2.20	76.8
mjrd_100	31.93	21.97	12.35

Finally, the inner zone EAs are generally much closer to a high school than outer zone EAs. Coupled with the prevalence of retail and recreational attractors in the inner core, these neighbourhoods are much more likely in the awareness space of the entire study area's juvenile delinquents.

Variable Selection

The second objective of the OLS models was to determine the relevant variables for the final confirmatory analysis. The spatial statistical software used in the final analysis does not contain a stepwise approach to model building, therefore it is necessary to determine the relevant variables before running a spatial regression model. Again, it should be emphasized that OLS may not be appropriate for this analysis. However, Haining uses an ordinary least-squares regression to "sift out" variables in order to provide the "best" statistical model from a large, multicollinear data set containing spatial effects. He notes that

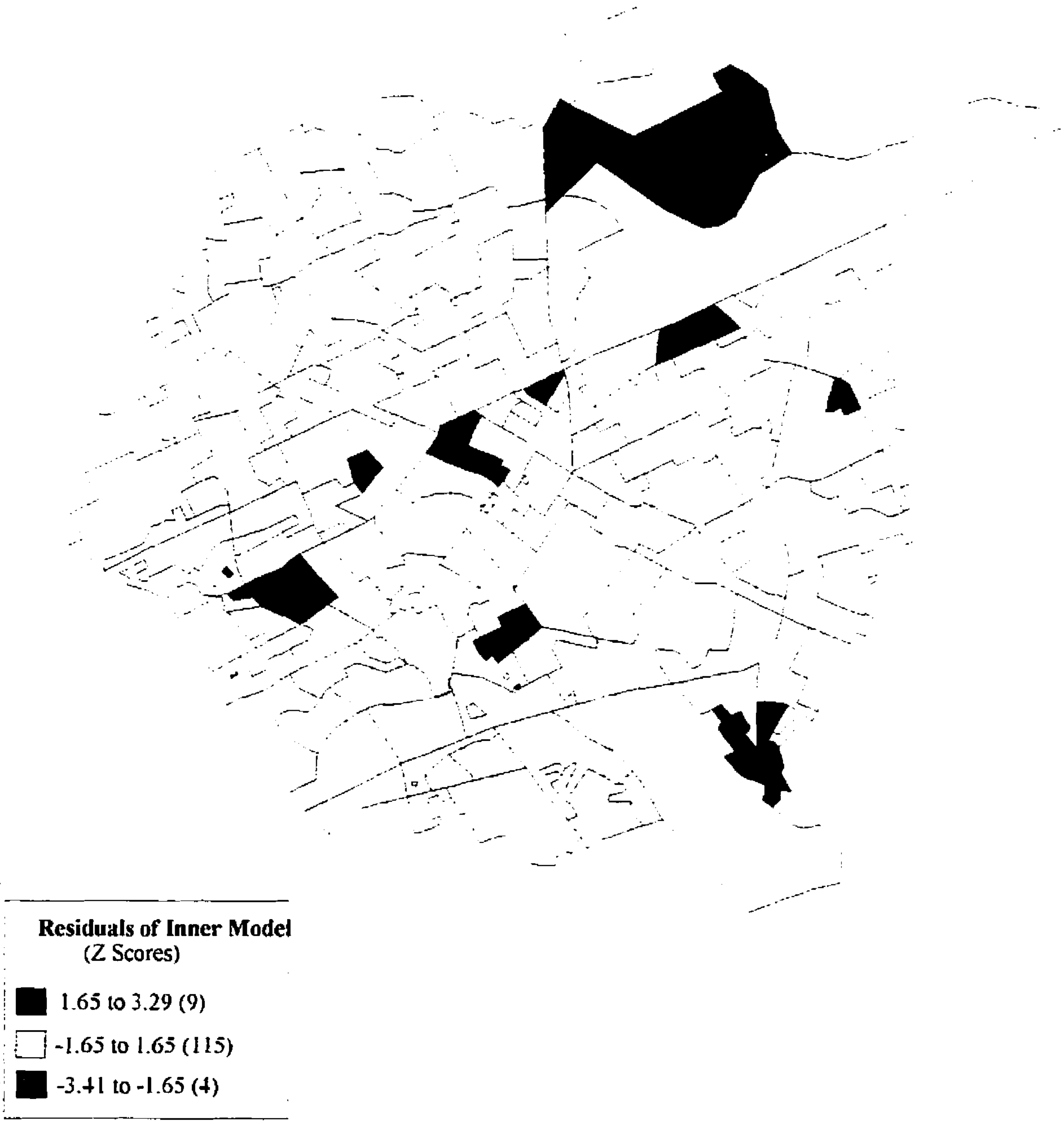
> "Where spatial correlation and/or heteroscedasticity is present...the R^2 measure is, of course, unreliable as a measure of goodness of fit. Thus, although in this paper these problems are taken into account in the final stage of estimation and model assessment, they were not incorporated into the earlier stages of analysis in choosing variables to include." (Haining, 1991: 224)

The inner model, including parameter estimates and significance measures for the t statistics, is as follows:

$$
\begin{aligned}
\text{ft\_burg} = \quad & 5.281 && (0.00) \\
& - 0.092\ \text{lnewdens} && (0.00) \\
& + 0.002\ \text{loneprnt} && (0.02) \\
& + 0.027\ \text{m1517\_1} && (0.02) \\
& + 0.005\ \text{wloneprt} && (0.00) \\
& - 0.002\ \text{year\_blt} && (0.00)
\end{aligned}
\qquad \textbf{(12)}
$$

The model fit is fairly weak (adjusted R^2 of 0.26; F statistic of 12.66) but substantially better than the global model created in section 4.1.1. The results show that the central area's burglaries remain predominantly a function of offender variables. High risk areas tend to be those both containing and bordering a disproportionate number of single parent families. They also are located near clusters of male teenagers. High risk areas also are built at lower densities, and contain older housing.

The residuals of the inner model do not exhibit any serious autocorrelation. The standardized Moran's I score of 1.22 is not significant. Measures more suited to detecting autocorrelation in the residuals, the Lagrange Multiplier diagnostics for error and lag dependency, confirm these results with 1.33 and 0.81 respectively. A map of the inner zone residuals also shows that there is no clear pattern to the unexplained variance of the inner model (see Figure 4.9). This indicates that in this zone the ordinary least squares model is adequate for the analysis.


Residuals of Inner Model
(Z Scores)
1.65 to 3.29 (9)
-1.65 to 1.65 (115)
-3.41 to -1.65 (4)

The outer model's fit is much better (adjusted R^2 of 0.39; F statistic of 17.46). The equation including parameter estimates and significance measures for the t statistics is as follows:

$$
\begin{aligned}
\text{ft\_burg} = \quad & 13.227 & (0.00) \\
& + 0.068 \text{ bedrooms} & (0.00) \\
& + 0.390 \text{ km\_cbd} & (0.00) \\
& - 0.063 \text{ lnewdens} & (0.00) \\
& - 0.011 \text{ winfant} & (0.03) \\
& - 0.007 \text{ wyrblt} & (0.00)
\end{aligned}
\qquad \textbf{(13)}
$$

High risk suburban areas are those which contain newer, larger housing built at lower densities. They tend to be located further from the CBD. They also are inclined to be located near areas with few infants. These are clearly a different set of determinants than found inthe inner zone.

The outer model's residuals do exhibit clear autocorrelation . The standardized Moran's I was calculated as 2.13, the Lagrange Multipliers for error and lag were 3.53 and 2.5, respectively. Thus ordinary least squares regression does not adequately model the data due to the presence of spatial effects. The topic of the next section deals with the array of spatial regression models available to analyze such data.

4.2 Confirmatory Spatial Data Analysis

This section outlines the final stage of the analysis, the construction of the multivariate spatial models used to describe the spatial distribution of burglary rates in the study area. It begins with a review of many of the spatial regression models available to account for spatial effects. This includes the justification of the CSDA model selection. It then presents the results of their application, including a description of the parameter estimates, measures of fit and residual diagnostics.

4.2.1 Spatial Regression Model Selection

In section 2.4 the ordinary least squares regression model was shown to have a number of limitations when applied to spatial data. There are a number of alternatives to the traditional regression model which use maximum likelihood parameter estimation to account for spatial effects. The following section explains the final choice among these models by way of reviewing the more common alternatives. The first section deals with spatial autocorrelation, the second with spatial heterogeneity.

Models for Spatially Autocorrelated Areal Data

The purpose of spatial regression models is to account for the spatial dependency during the parameter estimation procedures. As outlined in section 2.4, spatial autocorrelation has two primary causes. The first relates to the effect of a substantive spatial process whose impact needs to be quantified and incorporated into the model. The second form of spatial dependency is known as nuisance autocorrelation. It results from less direct causes such as the choice of spatial partition. Consequently, there are two different approaches to building mixed regressive–spatial-autoregressive models

The spatial lag model includes a spatially lagged dependent variable as an explanatory variable. This parameter is used to quantify the autocorrelation in the data:

$$y = \rho Wy + X\beta + \varepsilon \qquad (14)$$

where y is the dependent variable, ρ is the spatial autoregressive coefficient, and Wy is the spatially lagged average for the dependent variable. Each coefficient is estimated simultaneously using a maximum likelihood approach.

This form of dependence is referred to as substantive spatial dependence, since it relates to the dependent variable and is an expression of a spatial process, such as diffusion (Anselin, 1992). An example of the application of this model is the analysis of the role of contiguity in the geography of international conflicts, where it has been shown that the potential for one nation to go to war is in great part a product of its neighbours' potential (O'Loughlin, 1986: 64). This effect has also been found with epidemics (Cliff and Ord, 1981).

In this study it was felt that the use of the substantive model was unwarranted. Burglary was not considered to undergo any spatial process such as diffusion. To regress the weighted average of burglary rates on a given burglary rate would add considerable explanatory power in a statistical sense, but not substantively. Furthermore, diagnostics of the OLS residuals suggest the use of an error model instead of a lag model. However, spatial lag models can be used to eliminate dependency in the data –the autoregressive parameter is simply not interpreted to have any explanatory power. Therefore, it was also applied to the data as an alternative model.

The second form of spatial dependence occurs when the error term follows a spatial autoregressive process. When dealing with this kind of data, the objective is to eliminate the spatial autocorrelation in the error terms. This approach can be done either through spatial differencing or the construction of a spatial error model.

Spatial differencing is an uncomplicated way of eliminating spatial dependency:

$$(I - \rho W)\ Y = (I - \rho W)\ X\beta + \varepsilon \tag{15}$$

where ρ is the autoregressive parameter and I the identity matrix.(a matrix of equal dimensions whose diagonal is comprised of 1s).This equation is fitted using ordinary least squares on a range of ρ values. Whichever value of ρ that gives the most acceptable residuals is chosen. Using this approach, it is possible to obtain better estimates than could be using ordinary least squares regression and simply ignoring spatial autocorrelation (Bailey and Gatrell, 1995: 285). This approach was deemed too simplistic for this particular data set and consequently was not used.

The spatial error model is used specifically for dealing with spatial autocorrelation in the residuals:

$$\begin{aligned} y &= X\beta + \varepsilon \\ \varepsilon &= \lambda W\varepsilon + \xi \end{aligned} \tag{16}$$

In this case, the autoregressive parameter, λ, is introduced in the calculation of the errors. It is called a nuisance parameter because it does not provide any explanatory power, but is used simply to remove any autocorrelation from the model. The error term ξ is assumed to be well behaved (uncorrelated with fixed variance).The mean of the error term ϵ is zero, regardless of the value of λ, therefore, the mean of y is not affected by the spatial error dependence. Again, all of the parameters are estimated simultaneously using a maximum likelihood approach.

Models for Spatially Heterogeneous Data

A principal finding of the exploratory analysis is a distinct pattern of heterogeneity in the data which is not simply an artefact of first or second order spatial effects. In order to properly model burglary in this study area, it is necessary to recognize that two very different versions of this crime are occurring. The statistical evidence strongly suggests that within a four kilometre radius of downtown Kitchener burglary is a crime acted out by lower income people on their lower income neighbours. Conversely, in the suburban ring, burglars are more mobile and selective, seeking out affluent suburbs. Conceptually, we can see offender factors dominating in the central core, and opportunity factors the suburbs.

In statistical terms spatial heterogeneity refers to the case where the relationship between the dependent and explanatory variables shifts markedly from one subarea to another. This shift cannot be accounted for by a distance measure (such as *km_cbd*) or an ordinal variable measuring regime membership. What is required is a model which recognizes the distinction between the subareas and incorporates them into the parameter estimation process. There are two formal models available to deal with this spatial effect in SpaceStat: the spatial regimes models and the spatial expansion method.

The spatial regimes model is a recently developed solution to spatial heterogeneity. It involves first implementing an ANOVA test is used to determine if the model parameters vary over space. This is done by regressing individual variables on a binary dependent variable representing membership. If this is the case a regression model is run separately for each zone:

$$\begin{aligned} y1 &= \alpha_1 + X_1\beta_1 + \varepsilon_1 \qquad \textit{for } d = 0 \\ y2 &= \alpha_2 + X_1\beta_1 + \varepsilon_1 \qquad \textit{for } d = 1 \end{aligned} \tag{17}$$

The difficulty with using the spatial regimes model provided in the SpaceStat package is that it is only capable of creating unconstrained models (those which retain insignificant variables). Yet it is evident in this study that a constrained inner zone model would include an almost entirely different set of variables than the outer zone. The parameter estimates for the significant variables would thus be misleading. As a result of this weakness, the spatial regimes model was not used.

A second means to address spatial heterogeneity is the spatial expansion method, which combines a trend surface model with an OLS model to incorporate the parameters' variance over space (Anselin, 1992).

$$\beta_k = \lambda_{0k} + \lambda_{1k}^{z_1} + \lambda_{2k}^{2} + \ldots\lambda_{mk}^{z_m} \tag{18}$$

This method is fairly awkward as it is difficult to extract the spatial component from the aspatial component of the resulting model. While it may be useful as a predictive device it would not provide an adequately interpretable explanatory model for this study. As a result it was not used.

As an alternative to spatial regimes and spatial expansion to account for spatial heterogeneity it was decided to divide the study area into the two zones delineated in section 4.1.3. Subsequent analysis would be done separately for each zone. The inspiration for this approach was provided by an analysis of the voting patterns in Weimar-era Germany. It was found that modelling the support for the Nazi party was best done separately for each of Germany's regions due to their cultural, historical and economic uniqueness (O'Loughlin et al, 1994).

4.2.2 Spatial Regression Analysis

During the final analysis, ordinary least squares, spatial lag and spatial error models were applied to the data. The regression analysis of section 4.1.3 provided a number of suggestions for this phase. First that the data should be analyzed separately in an inner and outer zone defined by a four kilometre radius from the CBD. Second, that OLS is likely to be a sufficient technique for the analysis of the inner zone. Finally, it produced a list of variables that are likely to be significant in each model.

The results of the regression analysis show that in both zones the parameter estimates do not change substantially between model types. Within the inner zone, the goodness-of-fit statistics are very similar (see Table 4.4). Specifically the coefficient of determination ranges only between 0.26 to 0.28. The Akaike Information Criterion (AIC), a more appropriate goodness-of-fit measure for maximum likelihood models, also ranges only between -315.02 and -316.66. More importantly, neither of the autoregressive parameters is significant. Finally, there is no evidence of significant spatial autocorrelation in the least squares model, further suggesting that it is sufficient for constructing a model of the inner zone.

The standardized regression residual plot (against predicted value) reveals a zero correlation between the predicted value and the residual suggesting that no heteroscedasticity exists within the model (see Figure 4.10). A histogram of the residuals also shows a normal distribution which is indicative of a lack of influential points (see Figure 4.11).

Table 4.4 Models of the Inner Zone - Lagged variables included

Variable	OLS	R-SAR	ERROR	Test	OLS	R-SAR	ERROR
w_ft_burg		0.117 (0.26)		adj R^2	0.26	0.28	0.27
Constant	5.281 (0.00)	4.897 (0.00)	4.997 (0.00)	AIC	-315.75	-315.02	-316.66
lnewdens	-0.092 (0.00)	-0.093 (0.00)	-0.095 (0.00)				
loneprnt	0.002 (0.02)	0.002 (0.02)	0.002 (0.01)	Z(I)	1.22 (0.22)		
m1517_1	0.027 (0.02)	0.027 (0.02)	0.025 (0.03)	LM/LR (LAG)	1.33 (0.25)	1.28	0.32
wloneprt	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)	LM/LR (ERR)	0.81 (0.37)	0.17	0.92
year_blt	-0.002 (0.00)	-0.002 (0.00)	-0.002 (0.00)				
lamda			0.114 (0.30)				

Figure 4.10: Standardized Residual Scatterplot (Inner Zone)

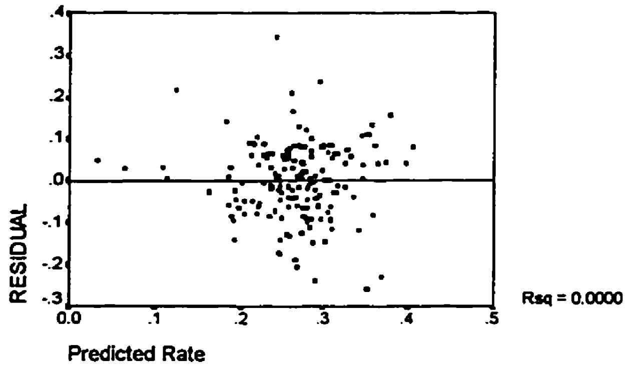
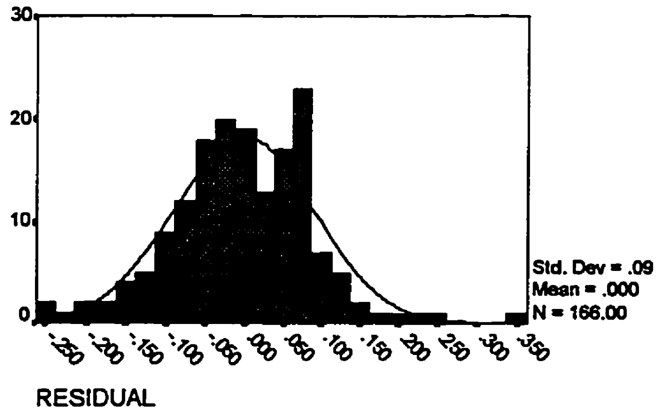


Figure 4.11: Histogram of Residuals (Inner Zone)



Within the outer zone there is conflicting evidence as to the best choice of model (see Table 4.5). The coefficient of determination points to the spatial lag model, although the AIC statistic is marginally lower for the error model (which is indicative of the best fit). However, the lag coefficient is significant while the nuisance parameter is not. Regression diagnostics indicate a lack of spatial autocorrelation in the residuals for the two spatial models, as does their visualization (see Figure 4.12). Again, the differences in the model parameters are only marginal. The choice was made to keep the spatial lag model and simply not assign any substantive value to the autoregressive parameter.

There is some evidence of heteroscedasticity when the predicted values were plotted against the residuals (see Figure 4.13). An improvement to this model may be to use some form of weighted regression to eliminate this trend from the model. The residuals are normally distributed, however, pointing to a lack of influential observations (see Figure 4.14).

Table 4.5 Models of the Outer Zone - Lagged variables included

Variable	OLS	R-SAR	ERROR	Test	OLS	R-SAR	ERROR
wft_burg		0.259 (0.01)		adj R^2	0.42	0.42	0.40
Constant	13.228 (0.00)	12.087 (0.00)	12.952 (0.00)	AIC	-236.30	-237.50	-238.78
bedrooms	0.068 (0.00)	0.065 (0.00)	0.069 (0.00)				
km_cbd	0.039 (0.00)	0.031 (0.00)	0.035 (0.00)	Z(I)	2.13 (0.03)		
lnewdens	-0.063 (0.00)	-0.062 (0.00)	-0.065 (0.00)	LM/LR (LAG)	3.53 (0.06)	3.21 (0.07)	0.59 (0.45)
winfant	-0.011 (0.03)	-0.010 (0.04)	-0.011 (0.03)	LM/LR (ERR)	2.50 (0.11)	0.02 (0.88)	2.48 (0.12)
wyrblt	-0.007 (0.00)	-0.006 (0.00)	-0.006 (0.00)				
lambda			0.194 (0.09)				

Figure 4.12: Outer Zones Residuals Map

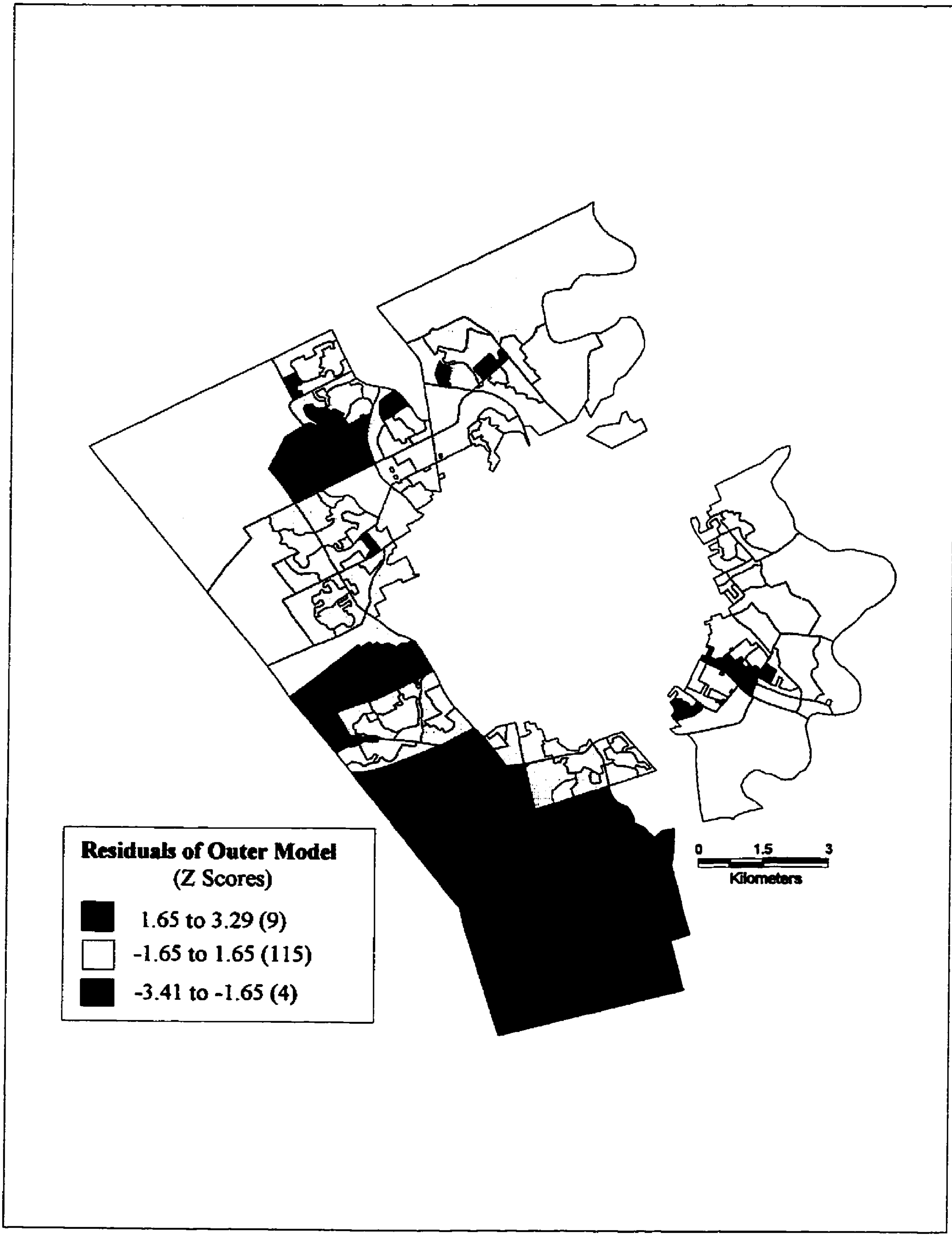


Figure 4.13: Standardized Residual Scatterplot (Outer Zone)

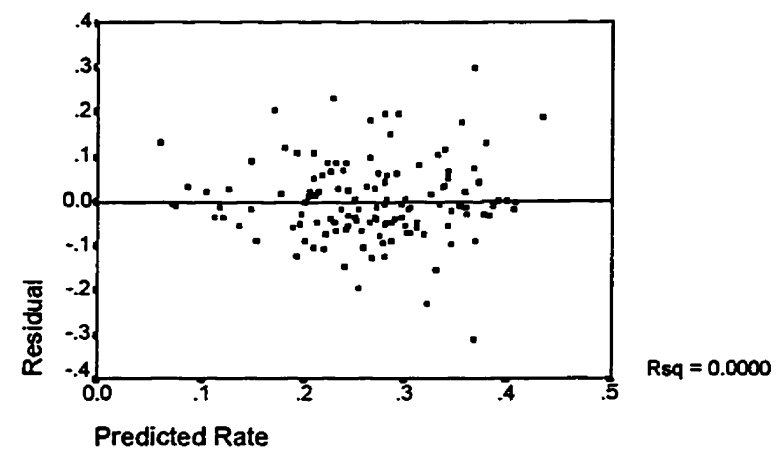
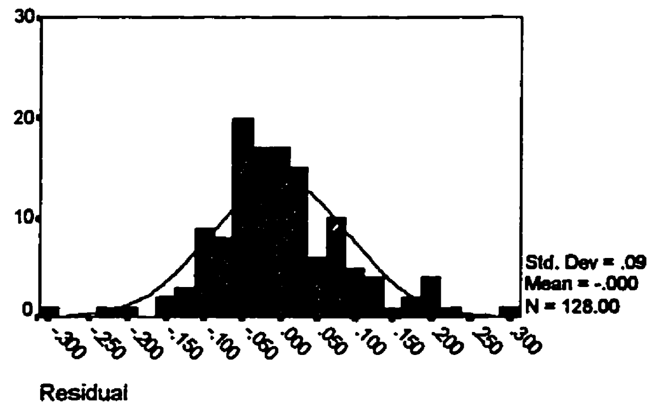


Figure 4.14: Histogram of Residuals (Outer Zone)



4.3 Interpretation

Interpretation was a necessary part of the analysis, so that many of the trends in the data have already been described. The purpose of this section is to combine all of the evidence gained from the entire analysis in order to gain a comprehensive understanding of the relationships at work in the data set.

The *m1517* variable is not significant in either the bivariate or multivariate analysis (discriminant analysis included). However, its spatial variant, *m1517_1*, is an important determinant in the inner zone regression model (see Table 4.4). This suggests that despite evidence of the central area's break-ins being very local affairs, the journey to crime maintains an important role in the burglary distribution in this area.

The *m1824* variable is never relevant in the analysis, but its spatial variant is significant in the discriminant analysis. It fails to remain a significant factor in the regression analysis. This was interpreted to be a function of they types of EAs that each multivariate technique is applied to. As noted in section 4.1.3, there are two clusters of young men in the study area: the central core neighbourhoods and the EAs located near the two universities in the northern suburbs. Only the first of these areas contains many high risk EAs, and neither contains many low risk areas. Consequently, when a discriminant analysis of the ERGs is performed it is looking primarily at neighbourhoods far removed from the sizeable student population in the study area; therefore, the *m1824_15* variable takes on a substantially different meaning. An improvement to this study would be to refine this variable by eliminating from this age-gender cohort those persons who are enrolled in school.

Table 4.6 Significant Explanatory Variables in the Analysis

VARIABLE	Regression (All EAs)	Discriminant ERGs	Regression (Inner)	Regression (Outer)
Offender				
loneprnt			+	
wloneprt	+	+	+	
Opportunity				
bedrooms		+		+
lnewdens	-	-	-	-
wyrblt				-
year_blt		-	-	
winfant				-
Access				
km_cbd				+
m1517_1			+	
m1824_15		+		
mjrd_100		+		

As described previously, the *lmavginc* variable displays a marked difference between the inner and outer zone correlation measures. This indicative of a significant shift in offender travel behaviour between the two sub-areas. Within the inner area the crime was inferred to be a very local event as areas of comparatively affluent men experience lower burglary rates; whereas, in the suburban belt, high average male income is a significant positive correlate of burglary rate. Its absence from the discriminant analysis equation is likely due to the neglect of the spatial heterogeneity associated with this variable. The average male income variable holds a significant correlation with variables that are included in the models: *loneprnt* (-0.56), *bedrooms* (0.60) and *year_blt* (0.39). Therefore, the contribution of the *lmavginc* variable is likely already included with these variables, causing it to be excluded during the stepwise model building stages.

The second measure of male income, *lt15000*, is a more direct measure of poverty among men. It is a significant correlate within the inner zone, although its lagged variant is not. This provides further evidence of the localized nature of central area break-ins. Again, its lack of significance in the discriminant analysis is likely due in part to the neglect of the parameter shift between sub-regions of the study area. And, as with the *lmavginc* variable, its significant correlation with other explanatory variables also is likely to have caused its omission from the final models.

The instability variable *loneprnt* proved to be relevant in the discriminant analysis as well as in the correlation and regression analysis for the inner zone. It appears that clusters of single parent families do tend to produce a disproportionate amount of offenders, although direct inferences to this variable's contribution are confounded by this variable's correlation to measures of poverty. The high proportion of lone parent families residing in the area, but lack of a corresponding proportion of teenagers or infants (see Table 4.3) implies that this dimension of instability is far more prevalent in the central area .

The *hhsz1* variable is negatively correlated with burglary in the outer zone . It was anticipated that this instability variable would, in fact, act to increase burglary rates. Within the outer area it is more likely to be a measure of the presence of elderly persons than unattached young men due to these variables' correlation of 0.67. A future improvement to this study would be to replace *hhsz1* with a variable measuring the proportion of residents who are male aged 18-34 and living alone.

The *rented* variable was used as a substitute measure of social cohesion because the census data for persons who had moved within the last year was not available. This is not an ideal variable, and it too had the opposite relationship with burglary than was expected. Its negative correlation with burglary in the outer zone is likely a reflection of less affluent targets who are less likely to be targeted by mobile offenders.

The *avghhsz* occupancy variable also produced an opposite relationship to what was originally anticipated. EAs with higher levels of larger households were found to be positively correlated with burglary rates in the outer zone. However, this variable is positively correlated with other explanatory variables: logvalue (0.33) and lmavginc (0.46). This suggests that although these areas likely maintain a larger daytime population, they also include more attractive opportunities for mobile offenders. This variable is not relevant in the multivariate models probably due to its correlation with explanatory variables that were included in the final equations.

The *infant* variable was significant only in the outer zone regression model and only in its lagged form. Its negative influence on burglary rates can be attributed to the higher levels of daytime population that neighbourhoods with pre-school age children maintain. The fact that an EA's neighbouring areas' measure of this variable are more significant than the original variable suggests that the variable represents some other kind of deterrent to burglary. It is more likely a reflection of a less criminogenic age structure (young families) of the local residents than of increased surveillance.

The *retired* variable is significantly negatively correlated with burglary in the inner zone but loses its predictive power in a multivariate framework. What was originally anticipated to be a significant deterrent to this crime (due to higher levels of daytime occupancy) is more likely to be an artefact of residential segregation; the retired variable is significantly negatively correlated with both distance to nearest high school (-0.31) and males aged 15 to 17 (-0.39). Therefore, persons over 65 are simply less likely to reside within the routine activity space of young offenders.

The *logvalue* measure of affluence is insignificant as a correlate with burglary in the inner zone but is positively correlated with the dependent variable in the outer area. This provides further evidence of the presence of professional burglars in the outer zone. This variable is not significant in the multivariate models, but is represented by some of the variables included in the final model with which it is highly correlated: *bedrooms* (0.48), *infant* (0.47), *km_cbd* (0.46), *loneprnt* (-0.44) and *lnewdens* (-0.44).

The *bedrooms* variable is important in the correlation and regression measures of the outer zone, as well as in the discriminant analysis. Suburban areas with larger houses (measured by the average number of bedrooms) tend to experience higher break-in rates. Its absence from the inner model further indicates that affluence factors do not play an important role within the central area.

The *lnewdens* variable holds a strong negative association with burglary rates across the entire study area, and throughout the progression from aspatial bivariate to spatial multivariate framework. Therefore, this is the single most important determinant of residential burglary rates at this scale of analysis. This lends support to the argument that density has the net effect of reducing the number of break-ins because of increased surveillability within a neighbourhood. An interesting expansion of this study would be to run the analysis with data recorded for census tracts to determine if this relationship remains true at larger scales.

The results pertaining to the *year_blt* variable and its lagged variant reveal that throughout the study area, EAs containing relatively newer housing experience lower burglary rates than older areas. This is true for both the discriminant analysis and regression modelling. This variable is correlated with income variables: *lmavginc* (0.39) and *lt15000* (-0.30). Therefore, EAs comprised of older dwellings are likely to contain a disproportionate amount of lower income males.

The *repairs* variable is a significant attractor in the inner zone, but it is also significantly correlated with income variables: *lmavginc* (-0.49) and *lt15000* (0.28), as well as the *loneprnt* instability variable (0.30). Therefore, it does not appear in the final model. It was felt that burglars would not be attracted to areas simply because they contain many homes in a state of disrepair. Rather this variable is likely to reflect the lower socioeconomic status of offender (and, therefore, opportunity) areas within the central core.

The *industry* variable is not significant except in the inner zone correlation measure. This was a difficult variable to operationalize, and is likely to be more relevant at a block-level scale of analysis than for the EA level.

The *km_cbd* variable has a negative correlation with burglary rates in the inner zone. It also is correlated with important offender and opportunity variables: *lmavginc* (0.41), *logvalue* (0.46), *rented* (-0.51), and *repairs* (-0.50). This suggests that the gradient hypothesis holds true within the inner zone. Conversely, it is significantly positively correlated with burglary rates in the outer zone, which suggests that the lower densities and higher levels of affluence associated with peripheral EAs act as a significant attractor to professional burglars. This variable is not relevant in the discriminant analysis likely due to the shift in direction of this variable between each zone. It is not included in the inner zone regression model likely because *year_blt*, with which it is highly correlated (0.69), is included.

The *hi_schl* variable does not appear to be relevant in any of the statistical tests. It was only included in an exploratory sense as it was not suggested as relevant in any of the literature. This attempt to quantify the routine activity space of the prime offender age-gender cohort did not provide any meaningful results. The *m1517_1* variable offered much more explanatory power, which suggests that the routine recreational activities of young offenders are more likely to coincide with spontaneous opportunities than any time associated with school.

The *hwy_exit* variable was also included for exploratory purposes. It was not relevant either. This variable was found to be significant in an analysis of the property crime rates of Chicago's suburbs, but the study was conducted at the intercity scale in a metropolitan area housing over eight million people (Brown, 1982). Kitchener-Waterloo is comparatively a very small urban area, and so proximity to a highway exit is unlikely to be a relevant factor to the mobile offender.

Finally, the proximity to a major road variable, *mjrd_100*, was important in the discriminant analysis but not in any of the correlation or regression models. This may be an artefact of the types of neighbourhoods which were included in the discriminant analysis, of which the downtown area was significantly over represented. Downtown EAs have high burglary rates and also are mostly located entirely within the buffer of one or more major roads. Subsequently, there may have resulted a spurious positive association between these two variables. The regression models' incorporation of all of the EAs seems to correct for this imbalance.

CHAPTER 5: SUMMARY

The final chapter of this thesis summarizes the findings of the analysis as well as the limitations of the study. It also offers recommendations to other researchers of crime.

6.1 Results

This study demonstrates that the ecological analysis of crimes such as burglary can be a valid and insightful method, although its success is contingent upon several modifications to the traditional approach.

First and foremost, ecological analyses must recognize that crime rates are to a large extent the product of spatial processes at work in the study area. Offenders travel in order in to commit this crime, which has enormous implications on the relevance of observation unit-specific offender variables. Furthermore, as demonstrated in the urban-suburban shift in significant predictor variables, the range of scales at which the different types of offenders operate must also be incorporated into an analysis of this crime.

Secondly, it must be recognized that there exists a spatial structure within aggregate crime data sets. This is partly an artefact of using arbitrary spatial units to measure a phenomena operating over a variety of scales (both smaller and larger than the size of the units of analysis); partly a result of the substantive spatial processes at work, such as the journey to crime; and partly due to the nature of the city itself, which tends to contain clusters of different types of criminals as well as different types of targets. The incorporation of this spatial structure must be made using the appropriate analytical methods.

Finally, the ecological approach must be expanded by applying the new technologies available to store and query spatial data. These allow the researcher to supplement the traditional use of census data, which alone was never fully capable of capturing many of the dimensions underlying the criminal event. The creative use of GIS to create new variables must only be restricted by the limitations of data availability, which is decreasingly a problem.

GIS clearly provided an important contribution to this study. The most relevant variable, population density, was constructed primarily in ARC/INFO. Other important GIS- constructed or transformed variables include the major roads variable, the spatially transformed age-gender cohorts, as well as the lagged variables. GIS also provided a valuable service by visualizing variables and spatially disaggregated statistics.

This study's relative success at modelling the determinants of this crime is a direct result of incorporating these three modifications. Despite the inclusion of spatially transformed variables, a traditional ecological analysis provided weak results. The explicit exploration and incorporation of spatial effects improved the level of explanation by over 100% in the case of the outer zone (from an R^2 of 0.20 to 0.42). The inner zone model improved 33% (from an R^2 of 0.20 to 0.26). These results are satisfactory considering the number of limitations confronting the study.

6.2 Limitations of the Study

This data set contained most of the problems encountered by spatial data analysts. The dependent variable is a measure of a rare event over a relatively short period of time that has been aggregated to a fairly small observation unit. Despite the measures taken to account for this kind of data (Poisson scores and the Freeman-Tukey transformation) the analysis was unavoidably weakened by the sensitivity of the dependent variable to a relatively small number of cases. More reliable results would likely be gained by using burglary data gathered from three or more years. Furthermore, the heteroscedasticity suggested to be at work by the regression diagnostics, may be reduced if there were an increased number of events with which to construct the dependent variable.

The temporal mismatch between the dependent and explanatory variables also likely had a negative influence on the analysis. This is of particular concern because the extent to which the 1991 data was obsolete is likely to have varied between different types of neighbourhoods. This systematic bias in the analysis can only be eliminated by using the 1996 Census data when it becomes available.

A related limitation is the absence of certain key variables in this analysis (such as the per cent of people who had moved in the last year), as well as the missing data values for variables that were included. These problems were partly solved by using proxy variables and interpolation techniques, respectively. Yet these solutions can never fully replace important data for an analysis.

There is an implicit assumption in this analysis that all of the offenders responsible for the break-ins in Kitchener-Waterloo reside within the study area. This fails to account for out-of-town burglars, such as those journeying from Cambridge or Hamilton. Yet this limitation may be offset by the dominance of opportunity variables in the outer zone, as one would have to assume that burglars from relatively distant areas would have to be motivated by large anticipated rewards.

This study is also limited in its neglect of the temporal dimension in the analysis. An improvement in the study would be to explore the spatio-temporal patterns of breaking and entering to gain further insight into the distribution of this crime. This was not done in this particular analysis because the data could only be analyzed in their aggregate form.

6.3 Recommendations

Recommendations based upon this study's findings are presented for those interested specifically in burglary in Kitchener-Waterloo, as well as researchers interested in the wider topic of the geography of crime.

It is difficult to provide specific recommendations to the police because case-specific evidence likely has much more bearing on a burglary investigation. Experienced officers are much more likely to recognize the work of a professional or the opportunist, and already are likely to be aware of the urban-suburban split in the distribution of each criminal's targets. What may have some preventative value for suburban residents is the distribution of information on (1) where the high risk areas are located, (2) when this type of crime occurs during the day and (3) which specific cues burglars look for. Further information on what professional burglaries look like when they occur may also be beneficial to the community.

Defensible space concepts are evidently valid, as demonstrated by the deterrent effect of population density and residential enclaves on the burglary rate. Future urban planning initiatives should recognize that the way communities are designed has an important influence on burglary rates, and possibly other forms of property crime. Furthermore, day-time occupancy is also important, so that the ideal form of safe community is one which houses a wide assortment of people in terms of their routine activities. Therefore, any new communities should be of mixed use and higher densities, as these neighbourhoods provide for much better policing by residents than the homogeneous suburban areas that are now being constructed on the study area's periphery.

Socio-demographic factors such as age and income are also clearly important determinants of burglary rates. It is therefore important to monitor changes in these variables in order to anticipate the growth or decline of this crime across the city.

The findings of this study are limited in their contribution to the wider field of the geography of crime because this analysis is spatially and temporally static. While the results do provide insights into the dynamics underlying this crime, they cannot be extrapolated to other study areas or other time periods. However, an interesting expansion of this study might be to look at similar cities, such as London or Windsor to see if the parameters are similar, or to look at the change in parameters over time in the Waterloo region.

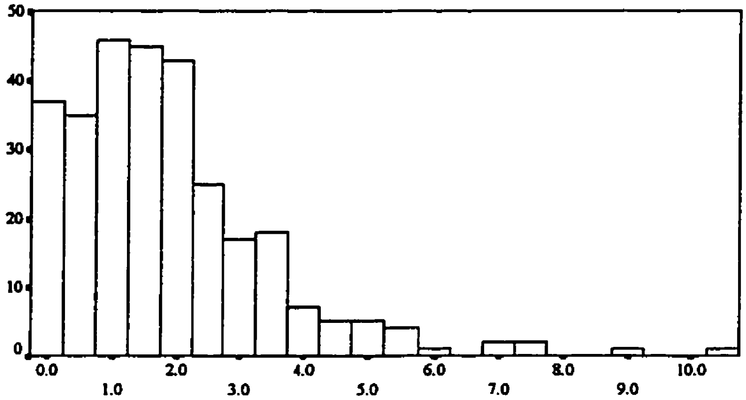
The success of applying GIS and SDA together is clearly a marked improvement on traditional ecological analysis. Yet the full potential of GIS was not reached. For example, spatial windowing, which allows for dynamic viewing of statistics compiled for subareas of the area under study, would be a useful addition to the ESDA stage –particularly in specifying the exact extent of the spatial regimes found within a study area. Another technique would be to use GIS to perform sensitivity analysis on CSDA models to the modifiable area unit problem. These techniques would have greatly facilitated and improved the exploratory component of the analysis. Unfortunately, such new spatial analytical techniques are not yet widely (or cheaply) available for the GIS packages used in this study.

Another problem encountered in this analysis was that of software limitations. SpaceStat is an important advancement in spatial statistical analysis but it is limited. In particular it does not provide stepwise regression making it difficult to rid the model of multicollinearity in a systematic way, and, as such, is only able to produce unconstrained models. This option would be a significant in the software. Furthermore, a great advancement in this program would be a windows interface and rudimentary mapping capabilities, so that it would not be necessary to constantly import and export data between the SDA and GIS packages.

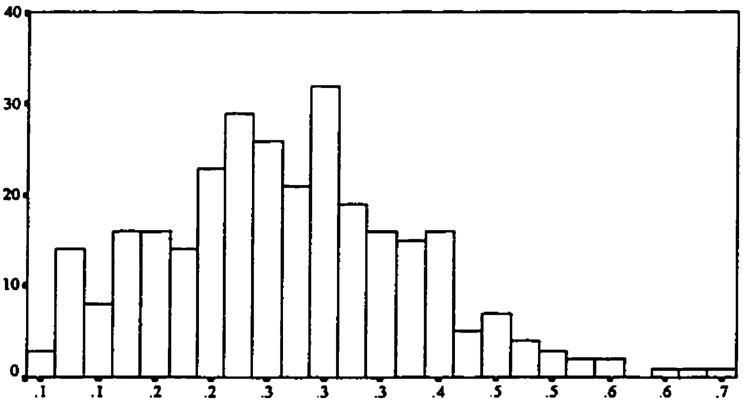
Finally, burglary is ultimately the result of an individual (or small group of people) selecting a single target. The analysis of burglary rates does provide insight on the dynamics of this crime, but can never be as effective as a study of the individual crime sites. What aggregate studies do contribute to is an understanding of the neighbourhood context of this crime. An improved understanding of the distribution of breaking and entering in the city would certainly be gained by a detailed examination of the individuals in Kitchener-Waterloo who commit this crime and the households that are their victims.

APPENDICES

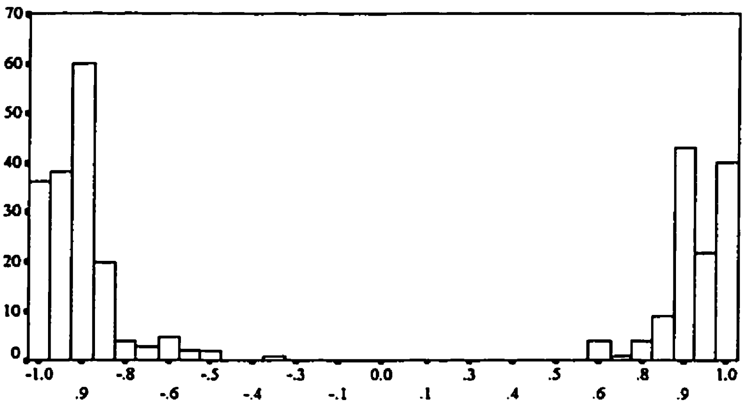
Appendix A: Histograms of the Dependent Variables



Burglary rate (per 100 dwellings)



Freeman-Tukey transformed burglary rate



Poisson transformed burglary rate

Appendix B: Descriptive Statistics

Variable	Mean	Minimum	Maximum	Std. Dev.	Skewness	Kurtosis	Valid Obs.
avghhsz	2.65	1.10	4.00	0.60	0.06	-0.75	294
avgmntrs	1.43	1.10	1.80	0.13	0.31	-0.37	294
avgvalue	171916	80000	649318	62359	3.87	0.29	277
bedrooms	2.64	0.90	4.00	0.62	-0.09	-0.57	294
burgrate	1.84	0.00	10.59	1.59	1.69	4.74	294
density	9130	29	201545	26806	1.69	4.74	294
km_cbd	3.74	0.16	8.93	1.71	0.19	0.28	294
hhsz1	22.26	0.00	91.30	15.94	0.83	0.28	294
industry	0.20	0	1	0.40	1.48	0.18	294
infant	7.09	0.00	28.00	3.95	0.72	2.52	294
loneprnt	13.39	0.00	50.00	2.23	1.27	2.23	294
m_avginc	31565	16820	80277	10150	1.81	4.10	261
near_hs	1.70	0.16	5.26	0.97	0.81	0.49	294
new_dens	18207	257	385442	48404	5.19	28.90	294
mjrd_100	27.56	0.00	100.00	24.58	0.90	0.45	294
p_retire	10.97	0.00	85.00	10.97	2.45	9.67	294
m15_17	3.97	0.00	13.95	2.47	0.61	0.81	294
lt15000	12.28	0.00	29.79	4.37	0.55	1.60	261
m1517	4.08	0.00	9.38	1.51	0.31	0.40	294
p15171	4.09	0.00	7.13	0.97	0.38	1.31	294
p18_24	11.90	0.00	44.44	5.24	1.64	8.00	294
p182415	11.65	6.75	16.94	1.46	0.28	1.07	294
rented	42.67	0.00	100.00	31.63	0.34	-1.13	294
repairs	27.03	0	72.73	14.64	0.11	-0.47	294
yrblt	1966	1924	1993	14.60	-0.67	-0.07	294

Appendix C: Pearson's Correlation Matrix of Study Variables

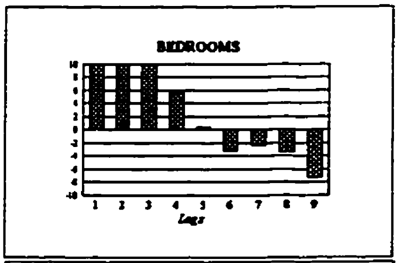
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
avghhsz (1)																							
avgmntrs (2)	.83																						
bedrooms (3)	.88	.78																					
burgrate (4)	.07	-.02	.12																				
ft_burg (5)	.11	.03	.17	.92																			
hhsz1 (6)	-.93	-.81	-.85	-.05	-.10																		
infant (7)	.35	.28	.10	.02	.02	-.30																	
km_cbd (8)	.50	.51	.48	-.04	-.01	-.51	.16																
km_exit (9)	.19	.19	.25	.07	.13	-.13	.04	.25															
km_hschl (10)	.43	.41	.37	.00	.04	-.40	.21	.60	.40														
lmavginc (11)	.46	.43	.60	-.01	.03	-.46	-.06	.41	.35	.39													
loneprnt (12)	-.29	-.41	-.47	.05	.00	.24	.30	-.29	-.18	-.23	-.56												
lnewdens (13)	-.23	-.27	-.40	-.31	-.40	.22	.09	-.10	-.17	-.17	-.23	.39											
logvalue (14)	.33	.30	.48	.06	.12	-.33	.46	.46	.35	.37	.68	-.44	-.44										
lt15000 (15)	-.20	-.24	-.21	.19	.14	.22	-.12	-.15	-.03	-.23	-.53	.28	-.08	-.17									
m1517 (16)	.48	.35	.42	.08	.08	-.41	-.03	.18	.09	.13	.23	-.05	-.10	.15	.02								
m1517_1 (17)	.53	.43	.41	-.05	-.05	-.45	.09	.31	-.05	.17	.25	-.08	.09	.14	-.10	.52							
m1824 (18)	-.25	-.16	-.22	.05	.07	.18	.00	.01	-.01	-.16	-.37	.26	.03	-.11	.51	-.15	-.10						
m1824_15 (19)	-.30	-.26	-.21	.11	.10	.22	-.10	.09	-.12	-.16	-.23	.14	.02	.02	.17	-.17	-.33	.35					
mj_rd100 (20)	-.29	-.27	-.30	.02	.00	.31	-.02	-.27	.02	-.11	-.17	.18	.12	-.17	.11	-.09	-.10	.05	.05				
rented (21)	-.70	-.70	-.84	-.06	-.13	.70	-.51	-.30	-.17	-.30	-.62	.62	.46	-.35	.32	-.29	-.27	.30	.25	.25			
repairs (22)	-.31	-.36	-.35	.14	.12	.30	-.50	-.43	-.23	-.27	-.49	.30	-.05	-.32	.28	-.15	-.23	.17	.18	.12	.31		
retired (23)	-.65	-.55	-.47	-.10	-.14	.67	-.51	-.38	-.08	-.31	-.09	-.12	.12	-.15	-.07	-.35	-.39	-.25	.11	.10	.28	.07	
year_blt (24)	.44	.43.	.29.	-.22	-.22	-.43	.23-	.69	.16.	.44	.39.	-.12	.24	.28	-.30	.21	.45	-.07	-.16	-.16	-.14	-.59	-.30

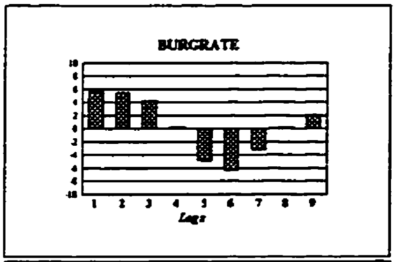
Appendix D: A Comparison of Discriminant Analysis Misclassification and Regression Residuals

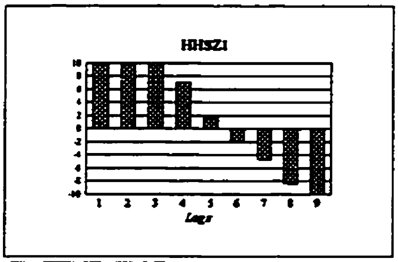
Inner Zone EA	high predicted as low (residual Z score)	low predicted as high (residual Z score)
38156	.530	
38163	.855	
38164	1.817	
38169	3.488	
38106		-2.423
38205		-1.635
38265		-1.491
89007		-1.090
89021		-2.527
89109		-.494
89118		-.748
89165		-1.078
89167		-.820

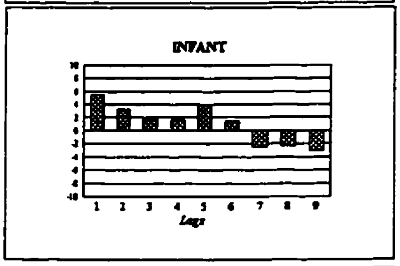
Outer Zone EA	high predicted as low (residual Z score)	low predicted as high (residual Z score)
38001	.590	
89413	.770	
38022	.825	
89121	2.121	
89206		-1.655
89261		-1.019
89208		-.886

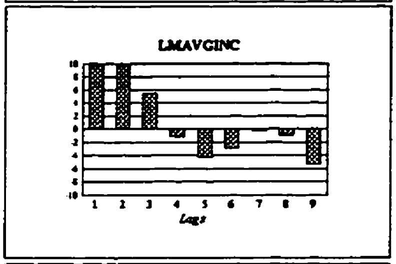
Appendix E: Spatial Correlograms Using Moran's I

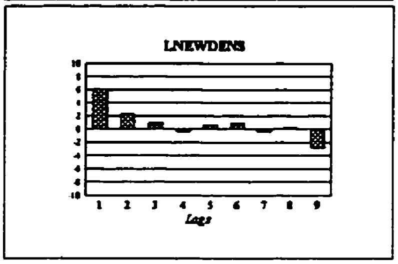


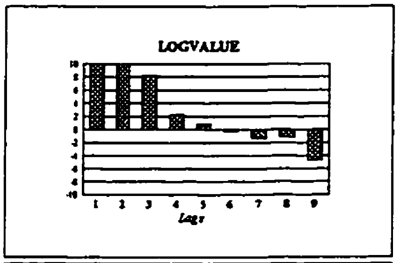


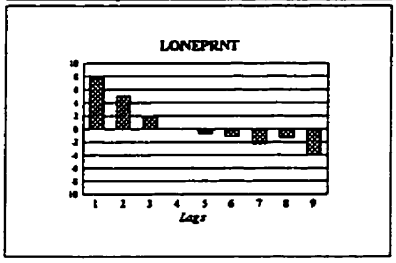


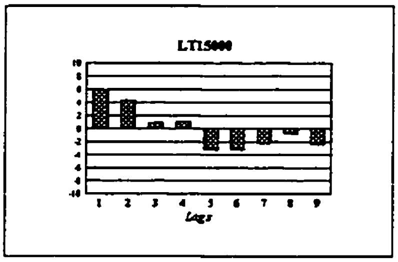


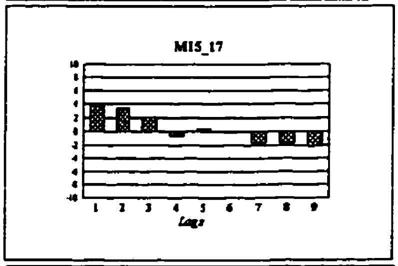


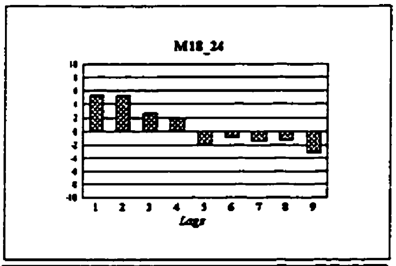


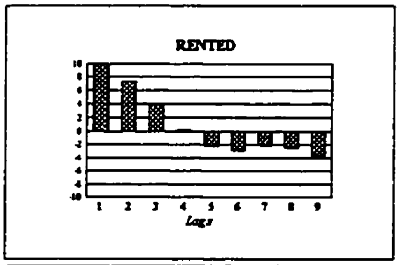


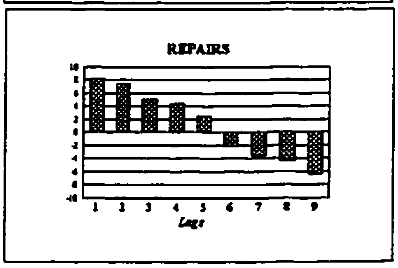


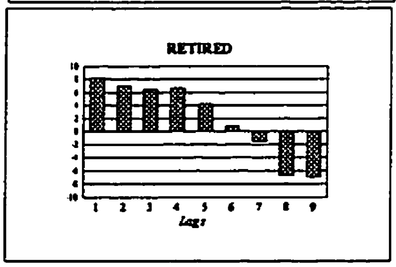


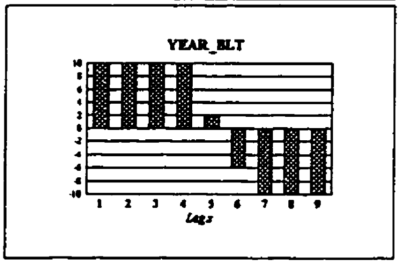












REFERENCES

Anselin, L. (1992): *SpaceStat: A Program for the Analysis of Spatial Data.* National Centre for Geographic Information and Analysis, Santa Barbara.

Anselin, L. and A. Getis (1992): Spatial Statistical Analysis and Geographic Information Systems. *The Annals of Regional Science*, Vol. 26, pp. 19-33.

Anselin, L. (1995): Local Indicators of Spatial Association - LISA. *Geographical Analysis*, Vol.27, pp. 93-115.

Bailey, T.C. & A.C. Gatrell (1995): *Interactive Spatial Data Analysis.* Longman House, Essex.

Baldwin, J. and A.E. Bottoms (1976): *The Urban Criminal.* Tavistock, London.

Bennett, T. (1989): Burglar's Choice of Targets. In *The Geography of Crime*, eds. D.J. Evans & D.T. Herbert, pp. 176-192, Routledge, New York.

Bennett, T. (1990): *Evaluating Neighbourhood Watch.* Gower Publishing Company, Brookfield, U.S.A.

Berry, W. (1993): Crime-pattern Analysis Through the Application of GIS Techniques to Tayside Police Offence Data. In *Crime and the Urban Environment*, ed. H. Jones, pp. 127-139, Gower Publishing Company, Brookfield, U.S.A.

Brown, M.A. (1982): Modelling the Spatial Distribution of Suburban Crime. *Economic Geography*, Vol. 58, pp. 247-261.

Brantingham, P., and P. Brantingham (1981): Introduction: The Dimensions of Crime. In *Environmental Criminology*, eds. P. Brantingham and P. Brantingham, pp. 7-26, Sage, Beverly Hills.

Brantingham, P., and P. Brantingham (1981): Notes on the Geometry of Crime. In *Environmental Criminology*, eds. P. Brantingham and P. Brantingham, pp. 27-54, Sage, Beverly Hills.

Bursik, Jr., Robert J., and Harold G. Grasmick (1993): *Neighbourhoods and Crime: The Dimensions of Effective Community Control.* Maxwell Macmillan, Toronto.

Byrne, J.M. and R.J. Sampson(1986a): Key Issues in the Social Ecology of Crime. In *The Social Ecology of Crime*, eds. J.M. Byrne and R.J. Sampson, pp. 1-22, Springer-Verlage, New York.

Byrne, J.M. and R.J. Sampson(1986b): Neighbourhood Family Structure and the Risk of Personal Victimization. In *The Social Ecology of Crime*, eds. J.M. Byrne and R.J. Sampson, pp. 25-46, Springer-Verlage, New York.

Chard, J. (1995): Breaking and Entering in Canada *Juristat* Vol. 15, No. 13. Catalogue 85-002. Statistics Canada, Ottawa.

Christie, Todd, and Kim Shields (1996): Crime Analysis System Helps Peel Regional Police Battle Crime. *ARC News* Vol 18. No. 2, pp. 28.

Cliff, A.D. & J.K. Ord (1981): *Spatial Processes: Models and Applications*. Pion Limited, London.

Coke, Sir E. (1797): *Institutes of the Laws of England, 4 volumes*. London.

Costanzo, C.M., Halperin, W.C., and N. Gale (1986): Criminal Mobility and the Directional Component in Journeys to Crime. In R. Figlio, S. Hakim and G. Rengert (eds) *Metropolitan Crime Patterns*, pp. 73-96, Criminal Justice Press, Monsey, NY.

Davidson, R.N. (1980): Patterns of Residential Burglary in Christchurch. *New Zealand Geographer* Vol. 36, No. 2, pp. 73-78.

Davidson, R.N. (1981): *Crime and Environment*. Croom Helm, London.

Davidson, R.N. (1993): New Directions in Environmental Criminology. In *Crime and the Urban Environment*, ed. H. Jones, pp. 1-13, Gower Publishing Company, Brookfield, U.S.A.

Dubin, R. (1992): Spatial Autocorrelation and Neighbourhood Quality. *Regional Science & Urban Economics*, Vol.22, pp. 433-452.

Dunn, C.S (1980): Crime Area Research. In *Crime: A Spatial Perspective*, ed. D.E. Georges-Abeyie and K.D. Harries. Columbia University Press, New York.

Evans, D.J. (1989): Geographical Analyses of Residential Burglary. In eds. D.J. Evans and D.T. Herbert *The Geography of Crime*, pp. 86-107, Routledge, New York.

Evans, D.J. and G. Oulds (1984): Geographical Aspects of the Incidence of Residential Burglary in Newcastle-Under-Lyme, U.K. *Tijdschrift voor Econ. en Soc. Geografie*, Vol. 75, No. 5, pp. 344-355.

Felson, Marcus (1986): Predicting Crime Potential at Any Point on the City Map. In eds. R.M. Figlio, S. Hakim and G.F. Rengert *Metropolitan Crime Patterns*, pp. 127-136, Criminal Justice Press, New York.

Getis, A. & K.Ord (1992): The Analysis of Spatial Association By Use of Distance Statistics. *Geographical Analysis*, Vol.24, pp. 189-206.

Getis, A. (1995): Spatial Filtering in a Regression Framework: Experiments on Regional Inequality, Government Expenditures, and Urban Crime. In *New Directions in Spatial Econometrics*, eds L. Anselin and R. Florax, pp. 172-185, North Holland, Amsterdam.

Grescoe, Taras (1996): Murder, he Mapped. *Canadian Geographic* Vol 116, No. 5, pp. 48-52.

Hagerstrand, T. (1970): What About People in Regional Science? *Papers of the Regional Science Association*, Vol. 24, pp. 7-21.

Haining, R. (1990): *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, Cambridge.

Haining, R. (1991): Estimation With Heteroscedastic and Correlated Errors: A Spatial Analysis of Intraurban Mortality Data. *Papers in Regional Science*, Vol. 70, pp. 223-41.

Hakim, Simon (1980): The Attraction of Property Crimes to Suburban Localities: A Revised Economic Model. *Urban Studies*, Vol. 17, pp. 265-276.

Harries, K.D. (1980): *Crime and the Environment*. Charles C Thomas, Publisher, Springfield.

Hendrick, D. (1995): *Canadian Crime Statistics, 1994*. Catalogue 85-002-XPE Annual. Minister of Industry, Ottawa.

Herbert, D.T. (1989): Crime and Place: An Introduction. In *The Geography of Crime*, eds. D.J. Evans and D.T. Herbert, pp. 1-15, Routledge, New York.

Hernandez, A.P. (1990): Artificial Intelligence and Expert Systems in Law Enforcement: Current and Potential Uses. *Computers, Environment and Urban Systems*, Vol. 14, pp. 299-306.

Hirschfield, A., Brown, P. and P. Todd (1995): GIS and the analysis of spatially-referenced crime data: Experiences in Merseyside, U.K. *International Journal of Geographical Information Systems*, Vol. 9, No. 2, pp 191-210.

Hirschfield, A. Brown, P. and J. Marsden (1991): Database Development for Decision Support and Policy Evaluation. In *Spatial Analysis and Spatial Policy using Geographic Information Systems*, ed. Les Worrall, pp. 152-187, Bellhaven Press, New York.

Jacobs, J. (1961): *Death and Life of Great American Cities*, Cape, London.

Jones K. and J. Simmons (1990): *Location, Location, Location*. Nelson Canada, Toronto.

Maguire, M. And T. Bennet (1982): *Burglary in a Dwelling: The Offence, the Offender, and the Victim*. Heinemann, London.

Maltz, M.D., Gordon, A.C. and W. Friedman (1991): *Mapping Crime in Its Community Setting: Event Geography Analysis*. Springer-Verlag, New York.

Miron, J. (1984): Spatial Autocorrelation in Regression Analysis: A Beginner's Guide. In Spatial Statistics and Models, eds. G.L. Gaile and C.J. Willmott, pp. 201-222, D. Reidel Publishing Company, .

Newman, O. (1972): *Defensible Space*. The MacMillan Company, New York.

O'Loughlin, J., Flint, C., and L. Anselin (1994): The Geography of the Nazi Vote: Context, Confession, and Class in the Reichstag Election of 1930. *Annals of the Association of American Geographers*, Vol. 84, No. 3, pp. 351-380.

O'Loughlin, John (1986): Spatial Models of International Conflicts: Extending Current Theories of War Behaviour. *Annals of the Association of American Geographers*, Vol. 76, No. 1, pp. 63-80.

Poyner, Barry (1983): *Design Against Crime: Beyond Defensible Space*. Butterworths, Toronto.

Rengert, G. (1989): Behavioural Geography and Criminal Behaviour. In *The Geography of Crime*, eds. D.J. Evans and D.T. Herbert, pp. 161-175, Routledge, New York.

Repetto, T.V. (1974): *Residential Crime*. Bollinger, Cambridge, Mass.

Rhodes, W.A., and Conly, C. (1981): Crime and Mobility: An Empirical Study. In *Environmental Criminology*, eds. P. Brantingham and P. Brantingham, pp. 167-188, Sage, Beverly Hills.

Sampson, Robert J. (1986): The Effects of Urbanization and Neighbourhood Characteristics on Criminal Victimization. In *Intrametropolitan Crime Patterns*, eds. R. Figlio, S. Hakim and G. Rengert, pp. 3-26, Criminal Justice Press, Monsey, NY.

Scarr, H.A. (1973): *Patterns of Burglary*. LEAA, Washington D.C.

Shannon, Lyle W. (1986): Ecological Evidence of the Hardening of the Inner City. In *Intra-metropolitan Crime Patterns*, eds. R. Figlio, S. Hakim and G. Rengert, pp. 27-54, Criminal Justice Press, Monsey, NY.

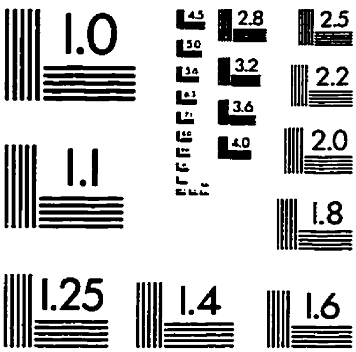
Shevky, E. and W. Bell (1955): *Social Area Analysis: Theory Illustrative Applications, and Computational Procedures*. Stanford University Press, Stanford.

Stahura, John M. and C. Ronald Huff (1986): Crime in Suburbia, 1960-1980. In *Intra metropolitan Crime Patterns*, eds. R. Figlio, S. Hakim and G. Rengert, pp. 55-70, Criminal Justice Press, Monsey, NY.

Waller, Irvin and Norman Okihiro (1978): *Burglary: The Victim and the Public.* University of Toronto Press, Toronto.

The Globe and Mail, Tuesday, December 17, 1996. Editorial: "Preventing Youth Crime"

IMAGE EVALUATION
TEST TARGET (QA-3)



150mm

6″

APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989