1999

Political participation in Kitchener-Waterloo: A spatial analysis of the 1993 federal election (Ontario)

Paul Churcher
Wilfrid Laurier University

POLITICAL PARTICIPATION IN KITCHENER-WATERLOO:

A SPATIAL ANALYSIS OF THE 1993 FEDERAL ELECTION

By

Paul Churcher
Honours Bachelor of Arts, Carleton University, 1992

THESIS
Submitted to the Department of Geography
in partial fulfillment of the requirements
for the Master of Arts degree
Wilfrid Laurier University
1999

0-612-40332-7

Abstract

This study examines political participation in the 1993 federal election in Kitchener-Waterloo. Voter turnout is modeled at the polling division level using statistical analysis with spatial diagnostics. This study is different from other Canadian research since it uses a fine scale spatial unit and it explores the geographical aspect of political participation. The findings confirm results from other studies as well as showing that geography plays a small but significant role in the decision to vote. The results of the statistical analysis show that polling divisions are appropriate spatial units for the examination of political participation. Demographics, political motivation, space, and voter disaffection are shown to be significant determinants of participation at the urban level. GIS and spatial statistics are used extensively and their application to political analysis is discussed.

Acknowledgments

Many people have provided assistance during the course of this research. My fellow students and family have been invaluable in proof reading the manuscript and in recommending changes. Especially worthy of mention are Martin Anstey who provided many useful suggestions, and Hélène LeBlanc of the Wilfrid Laurier University Library who supplied tremendous support for my research. In addition, I would like to acknowledge the assistance of my defense committee: Dr. Bob Sharpe of Wilfrid Laurier University, and Dr. Doug Dudycha and Dr. Peter Woolstencroft of University of Waterloo. Lastly but not leastly my thesis supervisor, Dr. Barry Boots, whose knowledge and assistance kept me on track throughout the whole process.

TABLE OF CONTENTS

Page

LIST OF TABLES AND FIGURES

Chapter One

Introduction

In his most reflective moments even the most experienced politician senses a nagging curiosity about why people vote as they do.

(Key quoted in Landes, 1987)

Individually, the act of voting is a simple procedure where a citizen marks a ballot regarding an issue or candidate. Collectively, the act of voting is a powerful statement that can peacefully overthrow strong governments and bring a dramatic end to political dynasties. Voting is a method of allowing public participation in the political process by anonymously choosing candidates who best reflect the views and beliefs of the general citizenry about how the governing of the nation should be undertaken. If these representatives are perceived as being ineffective, voting can remove them from power. The power of voting was demonstrated during the 1993 federal election when the governing Conservative party was swept from power and reduced to holding only 2 seats of 295.

Voting is perceived as an indicator of support for policies and is used in referendums to decide fundamental issues. There is a need for effective representation in the 1990s as economic pressures force budgets to shrink and tough choices must be made regarding which programs should be kept and which should be terminated. These choices affect all citizens and the electoral process is the only way that the public can make their preferences known. Participation is sometimes perceived as being unimportant by voters who feel that their one vote will not make a difference amongst the millions cast in an electoral event. However, the recent referendum on Quebec separation in October 1995 was defeated by only a few thousand votes and several ridings in the 1993 election were won by less than twenty votes. Participation can clearly make a difference. Failing to participate means that one loses any chance of influencing the outcome of important political events. The number of major electoral events over the past half decade (1992 and 1995 referendums, 1993 and 1997 general elections, and ten federal by-elections in 1994, 1996, and 1998 as well as dozens of provincial events) indicate that elections are a growth industry.

Electoral participation in Canada is voluntary and some electors will choose, for a variety of reasons, to not participate in the process. Elections are the driving machinery of representative democracy and a decline of citizen participation in these events is seen as alienation from the political system and disenchantment towards political issues (Bryan, 1998, p.161). This variation in the electoral turnout has geographic implications as it will not be distributed evenly over space and time. Examination of these turnout patterns can help to better understand the process at work whereby people participate in politics. This geographic variation will also occur at different scales and can be examined at a number of levels from the national level down to spatial variations within a metropolitan area. A temporal analysis at this level of detail would be too complex an undertaking so only one particular electoral event is examined.

This study examines the spatial variation in electoral participation in the 1993 general election in the cities of Kitchener and Waterloo. Turnout is modelled using fine scale spatial units rather than the much larger federal ridings which show spatial variation over larger regions but are not very effective in smaller regions. The results illustrate the appropriateness of spatial analysis to modelling electoral turnout and help in exploring the political participation process in general.

This research is presented in the following manner. Chapter Two sets the stage for exploring participation by introducing the Canadian electoral process and politics, as well as the behaviour of the Canadian voter. The findings relating to voter participation of the 1991 Royal Commission on Voting is also examined as this provides a useful background into national level studies. Chapter Three introduces geographical issues of relevance to voter turnout. Voting as urban spatial behaviour is examined as well as issues of scale and data aggregation, and the importance of using finer scale spatial units to model turnout. Previous research into political phenomena at the polling level and the study region are the subject of Chapter Four. Chapters Five and Six move away from general discussion and introduce the data and techniques used in this research. Chapter Five explains what data were used and why certain variables were chosen. Chapter Six explains the methods and techniques used to model turnout. Particular attention is paid to the spatial techniques that are important and yet often neglected in social science research. Exploratory data analysis, and multivariate analysis and results are the subjects of Chapters Seven and Eight. The research concludes at Chapter Nine which examines how the results relate to previous research and the suitability of fine scale analysis of electoral turnout. Conclusions about the importance of space in determining political participation are presented as well as a discussion regarding the future implications of communication technologies on the spatial dimension of voting, and participation in general.

The motivation and ideas for this research were developed by the researcher while working in the Electoral Geography Division of the Office of the Chief Electoral Officer during a four year period covering the 35th General Election in 1993 as well as numerous by-elections. These experiences and the contacts developed have proven to be a great aid in the execution of this research. The 1993 event was chosen over the more recent 1997 event because of personal familiarity with the 35th Election and because the fine scale electoral data had not yet been released for the 1997 election by the time this study had started.

Chapter Two

Elections in Canada

The Canadian Voter and the Electoral Process

The right to vote in Canada is extended to all Canadian citizens over the age of eighteen. Legal requirements in the past restricted the franchise to individuals based upon race, sex, age and residence but these limitations have been either eliminated or reduced to such a level that very few citizens are disenfranchised in federal elections. Electoral participation in Canada has been quite high and averaged 75% in general elections over the past thirty years (Landes, 1987, p.341). This is a good rate when compared to national rates in other democratic countries such as the United States. The government assumes the responsibility of enumerating all eligible voters through the enumeration process. This is believed to contribute to higher participation rates than in systems where it is the responsibility of citizens to get themselves on the electoral lists (Landes, 1987, p.338). The creation of a permanent list of electors using a geographic information system is expected to increase the participation rates even further (Chief Electoral Officer, 1997b) although there are fears that electoral information will be linked to tax data bases enabling the revenue agencies to track down debtors. This could result in some people who are otherwise eligible to vote not registering through fear of being found by Revenue Canada. There also exists the possibility of administrative disenfranchisement when someone meets the voting criteria but has been missed in the enumeration. Generally, several attempts are made to enumerate voters if they are not at home but this may not always be possible if the distances are great as in rural ridings. In some cases it may even be too dangerous to send enumerators out at a time when potential electors are home. In polling divisions designated as rural, an unenumerated elector may show up on voting day and be allowed to vote with valid identification. This is in response to difficulties of rural enumeration and is not allowed in urban areas.

Factors Affecting Electoral Participation

The Royal Commission on Electoral Reform of 1991 examined the question of voter participation. This research was done at the riding level but helps identify relevant factors in the Canadian context. Much of the research in electoral geography has been done in the US or UK setting and it does appear that Canadian politics and voters are quite different from these countries. The Royal Commission examined many of the administrative differences between countries and their effects on turnout but the first section of the commission's report is of the most interest. It contains an ecological analysis of turnout in Canada and provides both a good starting point for understanding the subject and furnishes a background framework for a more localised analysis. The difference of scale does mean that not all of the factors identified by the commission can be incorporated into a finer scale examination.

The ethnocultural make up of the population is considered important in predicting riding level electoral turnout. The ethnic group identified in the Commission report as being the most significant in this respect are aboriginals. This is a group that has felt excluded by the existing political infrastructure and usually has little incentive to participate in this system (Hall, 1993). Enumerators are often unwelcome on the reserves where aboriginals are concentrated and there exists a strong possibility of administrative disenfranchisement through poor enumeration. Furthermore, there also appear to be cases where the local government (band councils) decree that band members not take part in an electoral event. This was particularly evident in the 1992 constitutional referendum that in part concerned aboriginal issues.

Another factor affecting turnout is the social cohesion of a riding. This can be thought of as the residential stability of the population. People are more likely to know their neighbours, local candidates, and have common concerns if they have been living in the same neighbourhood for a long period of time. These social bonds will not form with an unstable population which is prone to movement. A common measure social cohesion at a riding scale can be the urban/rural composition of an electoral district. Participation rates are generally higher in rural ridings that in urban ones. A number of theories have been put forward to account for this with a popular one being the greater social cohesiveness of small communities that make up rural ridings (Eagles, 1991). Another measure of social cohesion is the residential stability of the population. High growth areas with a large influx of people are associated with low turnout since newcomers are less likely than long term residents to develop community ties and local knowledge necessary to develop high levels of participation. The demographic characteristics can also be associated with the social cohesiveness of the riding since the young population tend to be more mobile resulting in fewer community ties. The commission report identifies this as a factor but felt unable to incorporate this aspect into the analysis. Variations in age groups within a city would be quite evident since there are quite marked differences in the age compositions of the inner city and suburban zones. These intra-city age variations would probably not be very evident at a coarser electoral district level. Incorporation of different factors is an advantage of a finer, polling division level analysis. Another factor linked in a positive manner to participation rates is education and a common variable used in electoral studies is the percentage of the population with a university degree.

Turnout is affected by more than just the socio-economic and demographic characteristics of a riding. Eagles categorised both of these characteristics as "raw material for political mobilisation". The closeness of a race between parties or candidates is an important determinant of participation. The ability of a individual voter to influence the outcome is considered higher in a close race and this will generally increase the turnout. Interestingly this perception can exist also at a regional level. Voting in British Columbia starts because of time zone differences three hours after Ontario and Quebec where the bulk of the population lies. The feeling exists that the issue has already been decided well before the west coast polls close. Clearly, an important factor in electoral participation is that it be simultaneous so that all participants feel that they are able to influence the outcome. The number of independent candidates contesting a riding can be indicative of a tight political contest for that seat. Finally, political campaigning, television ads, television debates etc. are all thought to affect the turnout although this is difficult to analyse ecologically since they occur at a broad national level. However, campaign spending for each riding does show a positive correlation to the participation rates (Eagles, 1991).

Many other factors are also considered to affect participation rates. It is believed that weather patterns can be important in this regard. This has never been researched but seems to be common wisdom (Eagles, 1991, p.7). It is particularly relevant in Canada where the extreme climate is believed to generate a seasonal variation in turnout with both winter and summer being poor times for an election if participation is desired. The winter cold will affect the ability of political parties to canvas and get their messages out, while enumerators will have problems registering electors who will in turn have difficulties reaching polling stations. The summertime is believed to be a poor time electorally due to the large proportion of people who go on holiday during this period. The time of the week is also considered important. All federal events in Canada take place on a Monday. Conventional wisdom would indicate that holding an electoral event at the end of a week or on a weekend would not inspire high turnouts since people would be much more interested in other activities during their time off.

From the national study we see that the education level, ethno-cultural characteristics, and social cohesiveness of a riding are important determinants of participation as well as political considerations such as campaign spending and the tightness of the contest. Some of these characteristics will undoubtedly be important at a finer scale of analysis but some factors such as political spending per riding cannot be incorporated at a finer intra-riding scale. The results from the riding level research do provide a basis to start an analysis and see how these characteristics interact when they are combined with the spatio-structural characteristics that are in operation at the polling division level.

Political factors operating at the local level can also have some bearing upon electoral turnout. During an election campaign, party workers undertake door to door canvassing in order to spread information concerning their candidates ideas and policies. Canvassing may not be spread evenly throughout the riding and the different levels of canvassing may have an influence in turnout. In addition, political campaigns are rarely fair contests between honourable people and much occurs that is unethical and sometimes unlawful in the pursuit of victory. It is difficult to judge the effect of these occurrences on the election since they are not often acknowledged or discussed openly. This section serves as a brief overview of some of the things that can happen at the local level during an election campaign that can have a bearing upon voter participation. Brook (1991) discusses this type of activity much more extensively in his book based upon his own experiences as a campaign organizer. It should be stated that there is no evidence that these have occurred in Kitchener-Waterloo in 1993 but it is a rare electoral event that doesn't involve some form of underhandedness. Many tactics such as sign vandalism and distribution of faked campaign literature are undertaken in order to disrupt an opponent's campaign and may have an impact on how people choose to vote but probably have little effect on whether or not they choose to participate. Other tactics will have a direct impact upon turnout such as misdirection of an opponent's supporters to an incorrect polling station (Brook, 1991, p.217), harassment of electors outside polling stations, and organized telephone harassment of a campaign headquarters. Another tactic is the misdirection of an opponent's fleet of drivers (Brook, 1991, p.216), whose job it is to get supporters to the polls, with fake calls and incorrect addresses. Supporters of one candidate will sometimes pretend to support another in to order free their candidate's fleet of drivers and tie up an opponent's fleet. This was witnessed first hand by the researcher in the 1995 Québec referendum. It must also be remembered that returning officers and their staffs are political appointees and that this control over the staff running the polling stations can make it difficult for people to vote according to their ethnic or demographic background (Brook, 1991, p.217). Most of these tactics, although questionable, do not actually violate electoral or criminal law. Other well known tactics such as enumeration and casting of ballots on behalf of pets and the deceased, as well as the discovery of "lost" ballot boxes are quite rare as the running of elections is undertaken these days by relatively unbiased electoral agencies rather than the political party in power.

Canadian Politics In an Age of Restructuring

Much has been written in the field of political science regarding Canadian politics and attitudes towards politics and politicians. The present era of restructuring has lead to strong upheavals in the economy and has had a tremendous impact upon society. There is a feeling in Canada and in other parts of the Western World also that many of the decisions and policies that are affecting the country are not subject to consultation by the public and that citizens have little or no say in their formulation.

A term applied to Canadian politics is "brokerage politics" and it refers to the idea of using short term issues and promises to get elected. This type of politics is not limited to Canada but its impact upon the electorate does seem to be unique. National election campaigns consist of a wide mix of short term issues, policies, and images (Clarke, 1996, p.93) that can be conveniently disposed of after power is achieved. This was well illustrated by the Liberal government after its 1993 campaign when it rapidly adopted the neo-conservative policies of the previous government and the Reform opposition. This appears to have occurred as a result of cabinet decisions rather than consultation with the public or even with the Liberal caucus (Saul, 1998, p.402). Politicians themselves change parties as the political careers of Lucien Bouchard and Jean Charest demonstrate. As a result Canadian voters are very volatile and have few loyalties to parties and politicians. The socio-economic cleavages or social class divisions that characterise political behaviour in other western democracies are not present in the Canadian context (Clarke, 1996, p.94) and an angry and cynical electorate is a reality in federal politics of the nineties (Clarke, 1996, p.175).

This calls into question the meaning of an election. A great many ideas, images, and policies are presented during an election campaign and all of these influence how, or if, the electorate votes. This is a complex interplay of elements but all that the final results determine is who sits in parliament. With the "first past the post" system of victory, a party can even gain a majority of seats without a majority of the votes cast. The results regarding policy decisions are always subject to interpretation as shown in 1993 when the three major parties all claimed the results gave them a mandate to pursue their policies. The electorate clearly perceives a mandate for the victorious party and this explains much of the anger and frustration expressed when campaign promises are broken or reversed. This also applies to areas that have not been examined in an election campaign such as the constitutional changes that were attempted by the Conservative government in the 1990s.

A significant contributor to the apparent cynicism is evidence that public consultation and participation is not wanted by politicians on important issues. The Free Trade Agreement (FTA) was an election issue only because its implementation had been held up in the Senate and the1992 referendum seems to have been meant as a rubber stamp for a done deal as worked out by the political élites rather than consultation on constitutional change. It was only held when it became apparent that there would be unacceptable political fallout in the upcoming election if there were to be no attempt to consult the public.

This serves to increase the perception that parties will say or do anything to get elected and have little respect for the wishes of the ordinary citizen. Public opinion polls, the Spicer Commission on Constitutional Reform, as well as increases in other forms of participation, most notably in cyberspace, all indicate a growing frustration with the existing forms of representation. There is considerable evidence that the Canadian public is increasingly becoming politicised and wants to have a say in how issues are decided. Recent increase in registered candidates and parties seem to indicate this as well (Figures 2.1 & 2.2). This is in spite of recent electoral laws that increasingly restrict this type of activity. Yet participation rates do not seem to be on the increase and possibly seem to be decreasing. Figure 2.3 shows the figures for ridings that make up the region of study and similar trends can be seen at the national level. These figures would appear to indicate that much of the public has lost faith in the ability of the present system of representation to listen to and act upon the will of the public. Elections are often seen as mere mathematical exercises held once every four or five years. Certainly with the system of party loyalty and the lack of free votes in the House of Commons, MPs do not serve as representatives of

Figure 2.1) Registered Federal Parties

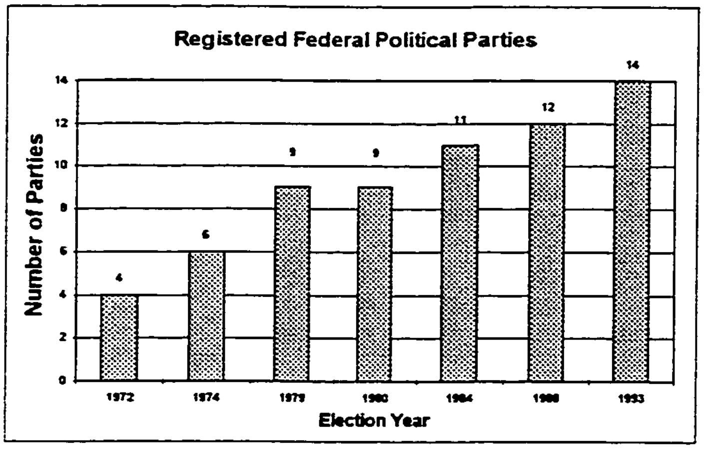


Figure 2.2) Candidates in Federal Elections

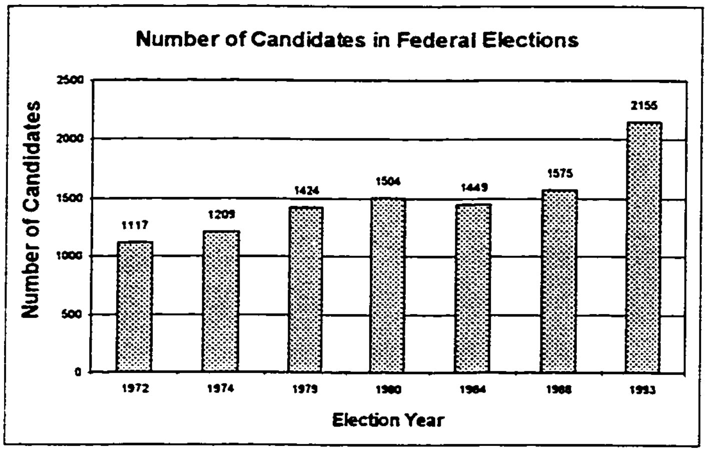
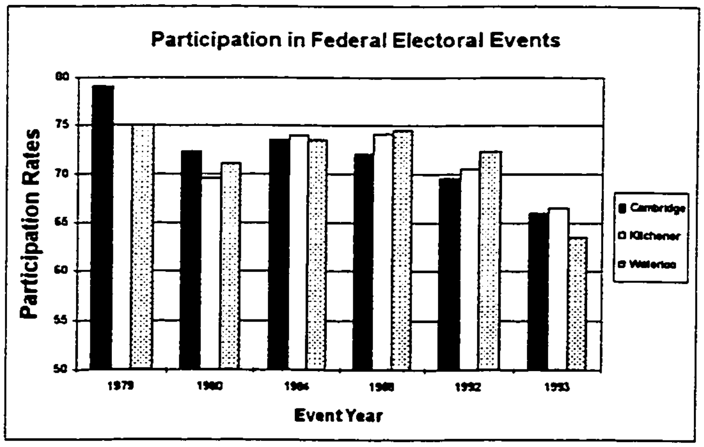


Figure 2.3) Turnout in Kitchener-Waterloo



Source: Chief Electoral Officer; Official Polling Results for the General Elections of 1972, 1974, 1979, 1980, 1984, 1988, 1993; and the Official Polling Results for the Referendum of 1992.

their constituents to Parliament but rather as ambassadors of a party to a riding. With the present practise of brokerage politics and the feeling that MPs do not really represent constituents, participation in the voting process appears to be dropping even if interest in the way we are governed appears to be on the increase. Voter frustration with the current system will manifest itself as variation in turnout which can be examined in a spatial manner.

The 1993 Federal Election

Before examining the election at a local level, it is necessary to set the broader national context of the 1993 election. The event took place on the 25th of October and resulted in a dramatic change in the federal political landscape. The five year term of the government would have been up in November but the election was called a few weeks earlier than necessary in order to make use of the voters list from the 1992 constitutional referendum held on October 26th, 1992 which was legally valid for one year only. Remaining in power for a couple more weeks until the very end of the government's term would have meant that a full enumeration costing millions of dollars would have to be undertaken thus rendering the Conservative government unpopular for causing unnecessary public expense. The final results indicate that this probably made little difference to the outcome: the Liberal opposition took 177 seats giving it the largest majority since the 1950s while the Conservative governing party was swept from power and reduced to holding 2 of the 295 seats in Parliament. The strong regional nature of federal politics manifested itself when the separatist Bloc Québecois became the official opposition party with 54 seats while the western based Reform Party won 52 ridings. The NDP managed to hold 9 seats and one riding elected an independent candidate in Québec.

Several major themes dominated this election campaign. The poor performance of the Canadian economy was a large issue as growth slowed and the economy entered a recession in the early 1990s. The government was held accountable for the country's economic woes due to the inflation policies of the Bank of Canada, the unpopular goods and services tax, and the FTA with the United States (Frizzell, Pammett, and Westell, 1994, p.2). Secondly, the government's failed attempts to solve the constitutional problems with the Meech Lake and Charlottetown accords of 1990 and 1992 respectively also played a part in its downfall. Surveys indicate that the two main issues involved in the rejection of the Charlottetown Accord in 1992 were the distinct society status for Québec and the guarantee of 25% of seats in Parliament for that province. The agreement was widely supported by the Canadian "élite" - the major federal parties, provincial governments, the news media, and the business community (Frizzell, Pammett, and Westell, 1994, p.4). The referendum vote revealed a deep animosity of the public towards policies and attitudes of the élite in general and of Prime Minister Mulroney personally. Another notable issue was the acquisition of expensive military helicopters at a time of economic cutbacks. The 1993 election was seen as a chance to vote out a government and a prime minister who was perceived to have become condescending, arrogant and out of touch with the Canadian people. The last minute replacement of Brian Mulroney with Kim Campbell as Conservative party leader did little to change this.

In spite of the referendum vote and its rejection of the élite, a new style of political campaigning did not emerge for the 1993 election. Issues of importance were never clearly debated and policies shifted in response to public opinion polls. Overall, the impression left upon the public was that politicians of all parties would say anything to get elected. The election campaign also became quite nasty when the Conservative party ran ads which were interpreted as attacking the facial deformities of Liberal leader Jean Chrétien. These clearly contradicted previous statements that Campbell would be a different type of politician (Woolstencroft, 1994, p.21). The Conservative campaign was generally considered to be inept and often seemed beset with internal turmoil since Campbell does not seem to have had time to leave her own impression upon the party. This is in strong contrast to the Liberals who had a very well organised campaign that was able to respond rapidly to events and polls and able to use the media to great effect (Woolstencroft, 1994, p.23; Clarkson, 1994 p.32)

The overall result of the election would indicate that, as in the referendum, there was a very large protest vote with the big winners really being the two fringe parties, Reform and the Bloc Québecois, who were able to achieve large numbers of seats in Parliament. Support for the Conservatives was relatively evenly spread throughout the country but the present electoral system clearly favours geographic concentration of party allegiances. This was demonstrated as the Liberals were able to pick up tremendous electoral gains in Ontario and the Maritimes. Geographic grievances are evident from the concentrations of support for the fringe parties in Western Canada and Québec. The overall result was that the Liberals were able to gain many seats based upon a split vote between the Conservatives and the fringe parties. This situation was to repeat itself in the 1997 general election.

1993 Electoral Turnout

Most major electoral events in Canada seem to achieve a turnout near the mid seventies and the 1993 event at 69.6% did not come close to this. The lower than usual turnout could have been expected since this vote did come within a year of another major electoral event, the 1992 referendum. Interest in elections tends to drop off if there is only a short time between major events. This phenomenon can be seen with the decreased turnout for the 1980 election which came less than a year after the 1979 event. This event was held in the winter which can also be expected to have an impact upon the willingness of electors to vote. Furthermore, the 1993 campaign does not seem to have been a close race. A tight contest between parties and candidates generally yields a higher interest and consequently higher participation rate (Eagles, 1991; Landes, 1987). The opinion polls generally indicated that the Liberals had very strong popularity stretching years before the beginning of the campaign. The Reform and Bloc Québecois were perceived as geographic oddities and not viewed as serious contenders to form a government while the NDP suffered from being associated with the extremely unpopular NDP government in Ontario. When the Tory campaign began to rapidly self destruct, a Liberal victory seemed assured. Overall, these conditions did not favour the impression of a tight race and the turnout was subsequently lower than most general elections.

Chapter Three

Electoral Geography and Scale Issues

Voter Participation as Spatial Behaviour

Voting is a form of political behaviour in that it allows citizens to signal their interests to government and controls those who govern (Bakvis, 1991, p xvii). Many studies have shown that electoral participation is related to the population characteristics in an area of study (Eagles, 1991; Landes, 1987). Clearly, who we are is important in examining political participation. However, voting is also a spatial phenomenon since it involves a trip to a specific location during a specific time period in order to cast a ballot.

This spatial dimension to the participation process has been rarely examined in electoral geography. Furthermore, to effectively incorporate spatial effects, the analysis must be done at the polling division level which is a level of analysis infrequently examined in Canada or elsewhere. In Canada, there is no choice over where an elector must go in order to cast a ballot. There is some degree of choice in the time an elector can vote since polling stations are open from 8am to 8pm on polling day. Furthermore, electors that are going to be out of town on polling day are able to vote in advanced polls that open several days before the event. Even here, there is no spatial choice involved since the polling division in which an elector is enumerated will determine which of the advanced polls can be used. Some efforts have been made recently to ease problems caused by this spatial dimension to voting. The 1992 referendum saw the introduction of mobile polls which are designed to ease voting for the sick and the elderly. Mobile polls are set up at hospitals and homes for the elderly and essentially go to the voter instead of the other way around. These polls are called mobile since they will change their location (eg. floor) within the institution several times during voting day in order to facilitate elector access. Voting at the advanced polls as well as the mobile polls was ignored for this research since it is impossible to determine a participation rate for the advanced polls since there is no fixed total of voters who would use the advanced polls.

Combining the decision to vote which is an act of choice with the trip to vote which involves no choice results in a unique spatial process which has received very little attention. Establishing polling stations to enable the public to vote can be thought of as the provision of a service. When use of a service is voluntary there is usually a choice in where to go to participate in this service. We have a choice regarding whether or not we wish to see a movie, open a bank account, buy furniture, renew a driving license etc. and there is always a matter of choice about where to go to do these activities. Other activities are not voluntary such as children going to school or adults undertaking jury duty and there is very little choice about where this activity is done. The spatially restricted participation process seems to be limited to voting. It should be noted that not all countries restrict where people go to vote. Indeed, much of the electoral research done at the city/polling district level was done in local elections in New Zealand where there is no restriction regarding which polling station voters use (Johnston, 1977b, p.11; Forrest, 1977, p.35).

This spatially restricted process also determines how the problem is examined. Systems that have no spatial restrictions can be examined by using spatial interaction such as the gravity model or, more recently, self-organising systems models (Batty, 1997, p.19). A different approach is required when the element of spatial choice is restricted. When both the home location of voters and their polling station is known, distance can be measured and it can be included as a variable in an ecological model. Brunk writes that distance is not appropriate as a variable in standard regression analysis (Brunk, 1985, p.55) since error variance is not constant over distance and will violate the assumption of regression analysis. However, this was in regard to distances measured from numerous geographical units to a single destination (a one to many spatial relationship). This is not the case with research using distance to polling stations since there are many different polling stations involved.

It would seem that strong party loyalties would lead to an electorate that is more willing to traverse the distance and barriers to a voting booth. An electorate with few loyalties would probably not be as likely to do this. The weak political cleavages in the Canadian electorate (Clarke, 1996, p.94) would appear to indicate that the spatio-structural characteristics of a riding would be more significant in Canada than in other countries. It is possible that the effects of space upon participation vary with the type of event examined. Spatial effects may manifest more clearly in a local political event where the political atmosphere is less charged than in a federal election. However, this is difficult to confirm since there is so little of this type of research performed in any country.

Scale Issues

The issue of scale is of significance in the study of electoral events. In research involving socio-economic data the scale of study can vary from the national level to the individual and choice of scale can greatly affect the success of studying the phenomena. Choice of too fine a scale will swamp the patterns of interest in background noise of individual variance while too coarse a unit will aggregate a pattern to non-existence. The level of aggregation can even change the manner in which a pattern behaves and in which variables interact. Problems with the aggregation of data over various scales have long been a concern of geographers. This has been referred to as the modifiable areal unit problem and it is known that the interaction among variables can be modified or changed significantly when data are aggregated to different scales. Wong (1996, p.93) shows how the parameter estimates of a regression equation change as data is aggregated from individual block groups into town and county level information. Significance of variables was affected and in some cases even the direction of the relationships (i.e. positive or negative) changed. This has particular relevance in electoral studies where data originally recorded at a very fine scale is examined at a much coarser scale.

Most electoral studies have been conducted at regional or riding levels using data that has been condensed from the much smaller polling divisions that serve as collection units for electoral data. Table 3.1 shows how much aggregation must occur when electoral and socio-economic data is analysed at these larger scales. It can be seen that a lot of aggregation occurs as tens of thousands of base collection units are aggregated into larger scale units. Considerable variation in the data is lost and processes that may be occurring in areas too small to be represented by a single riding cannot be studied. This includes electoral participation when examined as a form of urban spatial behaviour. Few studies have examined electoral issues at the polling division level in Canada. It is felt that this is a valid level of study since this geographic unit is the original unit used to capture the voting data and is relatively free of scale effect when compared to larger composite areal units. The scale at which a phenomenon is examined can determine what variables come into play with regard to the problem. A spatial analysis conducted at riding or regional levels will have to incorporate phenomena that occur at these scales. A electoral example of this would be campaign spending per riding or the rural/urban composition of the riding or region. In the Canadian context, a regional factor would probably have to be incorporated to account for the highly regional nature of Canadian politics at the national level. At a lower level of analysis these factors can be thought of almost as constants. This is not to say that they are of no importance but that their effect simply does not come into play at the lower scale. A polling division level analysis will need to take into account the spatial structural composition of the study area and of the electoral and census units. Factors that will come into play at this scale are actual and perceived proximity of voters to polling stations, the spread of political information throughout the study area by means of signs and canvassing, proximity to major roads, local impacts of broader issues of interest during the electoral event, and local effects such as the neighbourhood effect which examines the electoral impact of a candidate upon a surrounding area.

Table 3.1) Data Aggregation in Canadian Electoral Studies.

	ELECTORAL UNITS	CENSUS UNITS
Finer scale data collection units	N/A*	763636 (Block Faces)
	51770 (Polling Divisions)	45749 (Enumeration Areas)
Larger and highly aggregated spatial units	295 Federal Electoral Districts**	
	12 Provinces and territories	
	1 Country	

* Electoral geography does not have an equivalent of the block face used in the Census
**1993 figures. The number of Federal Electoral Districts increased in 1996 to 301.

Source: Chief Electoral Officer (1993b) p.xxvi, and Statistics Canada (1991b) p.173.

The polling division and census enumeration area are much better at capturing the scale at which people are accustomed and to which they can relate their spatial behaviour within the city. During the researcher's time in the government, the electoral maps that were always of the greatest interest to visitors were the polling division maps. Ridings were abstractions that often had little relevance but the polling divisions, by showing peoples' neighbourhoods, places of work, roads, and schools, revealed the structures that help shape peoples' day to day lives (Lynch in Jordan and Rowntree, 1986, p.383). These fine spatial units were units that have meaning and relevance to the electorate. This is not to say that these units are perfect in capturing the characteristics of neighbourhoods but they are certainly much better able to do this than the coarser units more commonly used in electoral studies.

Chapter Four

Previous Research and Study Area

Previous Research

There has been considerable interest in the study of elections in both the fields of political science and geography. This is because elections determine who has political power in society and voting behaviour provides insight into how people judge issues. Politicians who have misjudged the public's opinion and values will often pay a considerable electoral price. Elections provide a rich source of quantitative data which can be linked with other social data to gain insights into voting behaviour. Elections are of interest to geographers since they are spatial in nature with people organised into ridings and polling divisions. This process can be thought of as one of the last vestiges of the feudal systems whereby people were effectively tied to a parcel of land.

The study of elections has focused primarily upon voting results and not so much upon the participation factor. In the field of political science, studies are non-spatial in nature and focus more upon how rates vary between different political systems (eg. mandatory vs. non-mandatory, effect of different systems of enumeration, and also between levels of government. Elections at the national level tend to have much larger participation rates than those at the local scale. This is quite ironic since the powers and policies at the local level would appear to have more impact upon the day to day lives of the electorate than do national policies. The bulk of research in election studies focuses upon the United States and Great Britain. The Canadian research would seem to indicate that Canadian politics, and the electorate, behave in a different manner than in these other countries. Part of this seems to be the diminished role of class voting in Canadian politics. Gildengil (Gildengil, 1992) provides a good summary of research trends in Canadian Electoral Studies dating back to the 1960s. Studies examining the role of social class in voting are quite common as are regional studies. Furthermore, the role of party identification has also been well examined (Gildengil, 1992, p.231). She identified several large gaps in electoral studies and calls for further investigations regarding psychological and personality influences upon political behaviour, as well as gender research. There is no mention of the role of geography and space in small scale political analysis. Analysis of the impact of spatial factors at the intra-urban level in Canadian elections appears to be lacking.

Geographical studies of elections have also taken place in the US/UK context with focus upon the results of elections. Other topics commonly examined the manipulation of boundaries and redistricting as well as in the feedback process whereby voting results in a riding affect policy decisions regarding that riding. Canada seems to be a nation dominated by regionalism and the way in which different regions behave politically is a very common theme in Canadian electoral studies. With the continued threat of Québec separation this will probably continue to be the case in the future. A complete history of electoral geography is certainly beyond the scope of this paper but good summaries can be found in Walmsley (1984) and Johnston (1990).

Little work was done in election participation in Canada until the Royal Commission on Electoral Reform examined the issue in 1991. Volume 15 of the final report addressed the issues of participation in a riding level ecological study as well as a comparison between different countries and possible reforms to the Canadian system. The results of the ecological analysis have already been examined in Chapter Two and this study can be regarded as being the definitive research in Canadian participation at that particular scale. However, there are ways in which it could have been improved. Interestingly, there is no examination of residuals in this study. Residual analysis is an important part of a regression result and examination of why some parts fit a model while others don't is one of the more interesting parts of a study. Furthermore it is useful in examining how well the parameters fit and patterns in residuals can yield clues regarding any missing variables.

At the polling division level, much work has been done on examination of the neighbourhood effect (Reynolds, 1969; Johnston, 1977b; Walmsley, 1984, p.148), whereby local knowledge of a candidate will affect voting behaviour but voter participation studies are few and far between. In a Canadian context research was done to examine links between electoral turnout and questionnaire response in Saskatoon (Smith, 1967, p.115). The theory behind this research was that questionnaire response was being influenced by socio-economic factors and that this could be confirmed by examining the questionnaire responses to polling division turnout in a provincial election. This research used a form of areal interpolation in order to match polling division results to census tracts. It should be pointed out that there is a potential for scale mismatches since the socio-economic data was recorded at the census tract level which is a larger spatial unit than the polling division. Some conclusions were reached regarding the positive relationship between education, wealth, and turnout but this was not the prime focus of the research. Furthermore the effect of the structural characteristics of the data such as size and distance was not incorporated.

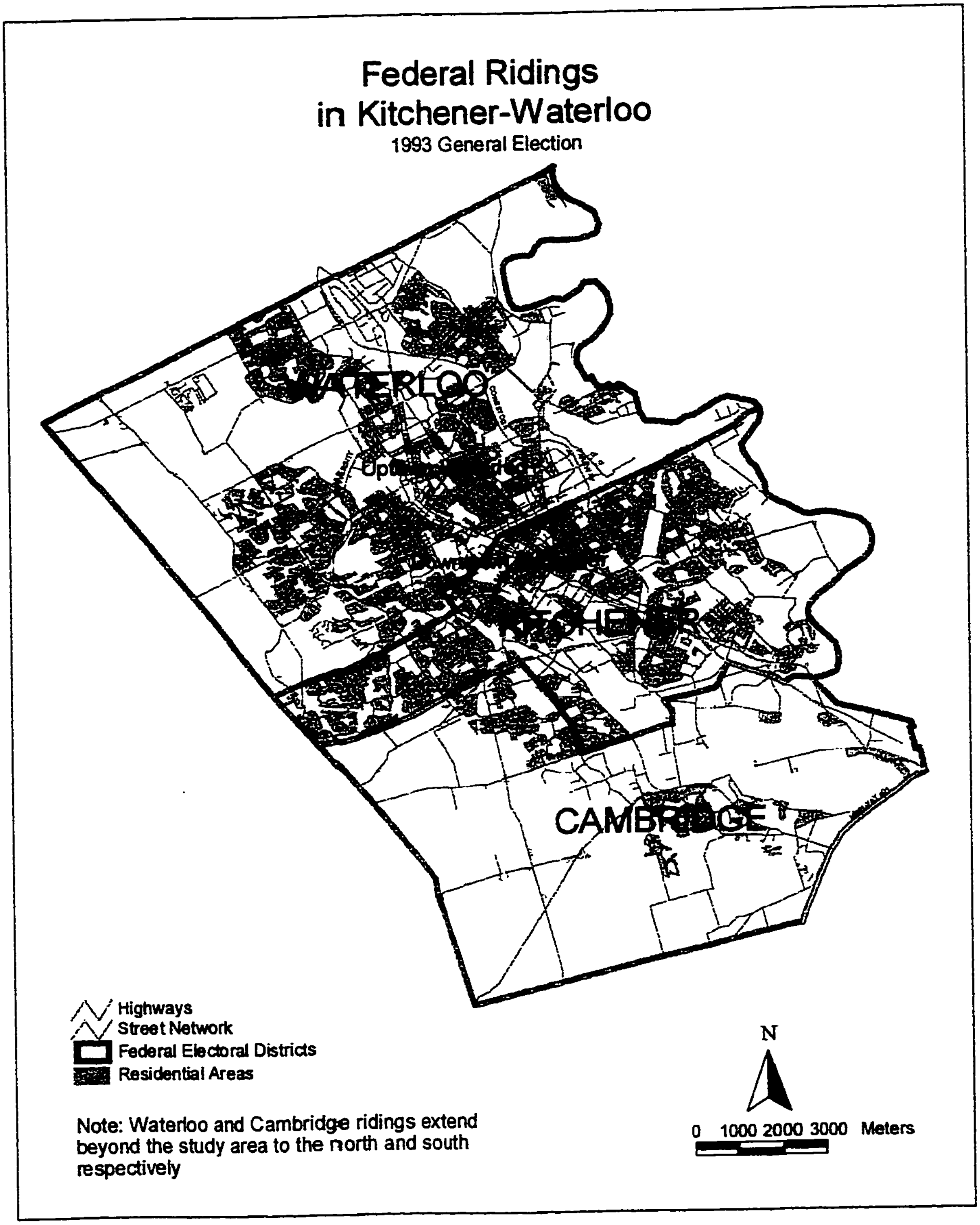
The distance to polling stations was incorporated in Taylor's examination of electoral turnout in Swansea, Wales (Taylor, 1973). An inverse relationship was found between participation and distance travelled when measured by time but a better predictor was found to be the perceived distance travelled. This was measured in terms of the number of road intersections or barriers that had to be traversed in order to reach the polling station. This study is very localised and looks at three polling stations and census tracts. An advantage of a detailed study at this level is that the structure of the area can be examined and the effects of gradients and hills was incorporated into the study. Overall this research examines many relevant issues; certainly a lot more than one would expect from an article four pages long. However, turnout relationship with socio-economic indicators is not examined in a statistical sense and it is not certain if the observed relationships would hold in other areas of the city due to the very small study area.

Overall, research has focussed on voting results rather than participation rates. Moreover, the bulk of research is dominated by American and British studies. Research also tends to be focused at the coarser regional or riding level with little interest in smaller unit analysis. It has been postulated that one reason for this lack of attention results from the problem of areal mismatches between the electoral and socio-economic units that must be used for this level of analysis (Reynolds, 1969, p.122). Use of geographic information systems (GIS) allows these areal mismatches to be overcome and it is possible that more interest will develop as GIS becomes more common in geographic and political research. For the moment however, Canadian research that examines small scale political participation that incorporates geographic elements at the polling division level is lacking and this research will help plug this gap.

Study Area

The study area chosen for this research is the cities of Kitchener-Waterloo (Figure 4.1). This region is located in the farm belt of south-western Ontario, roughly halfway along Highway 401 between London and Toronto. These cities are fairly small in population and have clear geographic limits which is important in minimising any contagion effects of surrounding regions. A similar study in a larger, more centrally located city such as the City of Toronto might become too large of a project due to extensive amounts of data as well as the presence and effects of the surrounding communities. The cities of Kitchener and Waterloo must be examined together since they are closely intertwined historically and geographically, and the smaller Waterloo can be thought as an extension to the larger, centrally located Kitchener. Examining all of the region allows a broad, diverse population to be researched. The region has a tremendous variety of industries ranging from a traditional manufacturing base associated with rust belt economies to a booming high technology sector. The presence of two universities also

Figure 4.1) Federal Ridings in Kitchener-Waterloo



brings a concentration of skilled intellectuals to the area. The city of Cambridge is also located close by but this was not incorporated into the study region since it is located south of Highway 401 which serves as an effective barrier severing Cambridge from the twin cities. The 401 also serves as a barrier in terms of civic boundaries and the finer electoral polling division units are also defined by the highway. Similarly, the rural regions in the north of Waterloo are not included even though they are part of the Waterloo federal riding since they have very different structural characteristics, large areas void of population which causes significant errors to an area weighted interpolation, as well as having more relaxed voting rules regarding enumeration and registering to vote. Excluding this region is regrettable since it would have been interesting to see how the large Mennonite population of the area affected participation rates. The study region of Kitchener-Waterloo includes parts of three federal ridings: Kitchener (in its entirety), Waterloo (185 polling divisions), and Cambridge (35 polling divisions). These ridings were established in 1987 and were in effect until April 27th, 1997 when they were replaced by new boundaries.

Historically, the Kitchener-Waterloo region has relied upon the manufacturing industry for its prosperity. Brewing, meat packing, automotive related manufacturing, and the rubber industry are all activities that have fashioned the physical and social geography of the region. Recently, economic restructuring has lead to downturns in the traditional manufacturing industries and increases in the financial sector, management services, and high technology industries. The spatial distribution of the older, land intensive manufacturing industries has lead to the development of an urban region that is highly dispersed and characterised by landscapes of often decaying factories, railways, and warehouses. Historically, the social strata in Kitchener-Waterloo was highly mixed (Bunting, 1984, p.52) and the region never developed low income ghettos or even exclusive enclaves for the rich, unlike its neighbours. This has been attributed to the influence of the German and Eastern Europeans whose cities in Europe did not have the same social differentiations characteristic of western Europe (Bunting, 1984, p.52). The large German population has been gradually supplemented since the end of the First World War by a population of British origin and the combined groups have dominated Kitchener-Waterloo to the extent that virtually no distinctive minority neighbourhoods have evolved (Bunting, 1984, p.50). Recently, the social fabric has begun to change with the influx of more Poles, Chinese, West Indians, and Italians (English, 1996, p.199). Also, the tertiary managerial and high technology sectors have clustered in Waterloo, while Kitchener has remained associated with secondary sector manufacturing and consequently the effects of recent economic restructuring have probably not been felt evenly throughout the region. However, the socio-economic characteristics of the area seem to remain quite mixed geographically.

The three federal electoral districts in the study area serve to show how ridings are poor spatial units for the analysis of political phenomena. The riding of Kitchener includes parts of the inner city as well as outer suburbs. The riding of Waterloo includes the northern rural area of the Regional Municipality of Waterloo which is combined with the city of Waterloo and parts of suburban and inner city Kitchener. These areas have very different social characteristics and political needs. The riding of Cambridge also combines many different regions since the city of Cambridge consists of a number of towns amalgamated in the 1970s. In addition, the riding of Cambridge includes a sparsely populated zone south of Kitchener as well as parts of suburban Kitchener. Futhermore, the entire riding is split by Highway 401 which effectively cuts the riding into two highly fragmented sections. As discussed in Chapter Three, a much finer level of analysis is needed to examine the political behaviour of urban populations.

Politically, the Liberals achieved strong victories in the region as shown by the vote totals for the candidates contesting the 1993 event (Table 4.2). Historically the Liberals have had considerable success throughout the years apart from the Mulroney era in the 1980s. English attributes this to several factors. Firstly, voters of German and eastern European extraction, who we have seen make up a large proportion of the population, have tended to back the Liberals. This pattern has been repeated in German communities throughout industrial Ontario. Secondly, Kitchener-Waterloo differs from other industrial communities in that the NDP, the traditional Canadian workers' party, has been unable to mobilise the large number of unionised workers in the industrial and manufacturing sectors. Also, the Liberal party seems to have been quite well organised in the region (English, 1996, p.180). The region seems to behave in a similar political manner to the rest of the country and, with the exception of Waterloo and Cambridge ridings in 1980, the region has voted for the winning party in every federal election over the past two decades. An examination of the participation rates shows another repetition of national trends with rates generally occurring in the early to mid seventies, apart from a drop in the winter contest of 1980, followed by a steady decline in turnout for electoral events in the nineties.

Table 4.2) Candidates in the Federal Electoral Districts of Cambridge, Kitchener, and Waterloo.

Riding	Candidate	Party	Occupation	Votes Received
Cambridge	Janko Peric	Lib.	Welder	22121
	Reg Petersen	Ref.	Businessman	18890
	Pat Sobeski	P.C.	Member of Parliament	9773
	Bill McBain	N.D.P.	Briefing coordinator	2980
	Ron Cooper	Nat.	Facility manager	1802
	Michael Picard	C.H.P.	Supervisor	407
	Thomas Mitchell	N.L.P.	Teacher	372
Kitchener	John English	Lib.	University professor	26616
	Reg Gosse	Ref.	Business owner	12214
	John Reimer	P.C.	Member of Parliament	10413
	Ian MacFarlane	N.D.P.	Housing coordinator	2373
	Pat Schiebel	C.H.P.	Housewife	475
	Katherine Finlay	N.L.P.	Teacher	438
	Joel Tarbuck	Libert.	Engineering technologist	165
Waterloo	Andrew Telegdi	Lib.	Not listed	26269
	Mike Connolly	Ref.	Semi-retired	15916
	Lynne Woolstencroft	P.C.	College professor	15109
	Scott Piatowski	N.D.P.	Housing coordinator	2822
	Ted Kryn	C.H.P.	Medical doctor	942
	Rita Huschka Sprague	Libert.	Not listed	493
	Blaine P. Watson	N.L.P	Teacher	449
	Don Phillip Faithful	Ind.	University student	332

Christian Heritage Party (C.H.P.)
Independent (Ind.)
Liberal (Lib.)
Libertarian (Libert.)
National Party (Nat.)
Natural Law Party (N.L.P.)
New Democratic Party (N.D.P.)
Progressive Conservative (P.C.)
Reform (Ref.)

Chapter Five

Data

Data Requirements

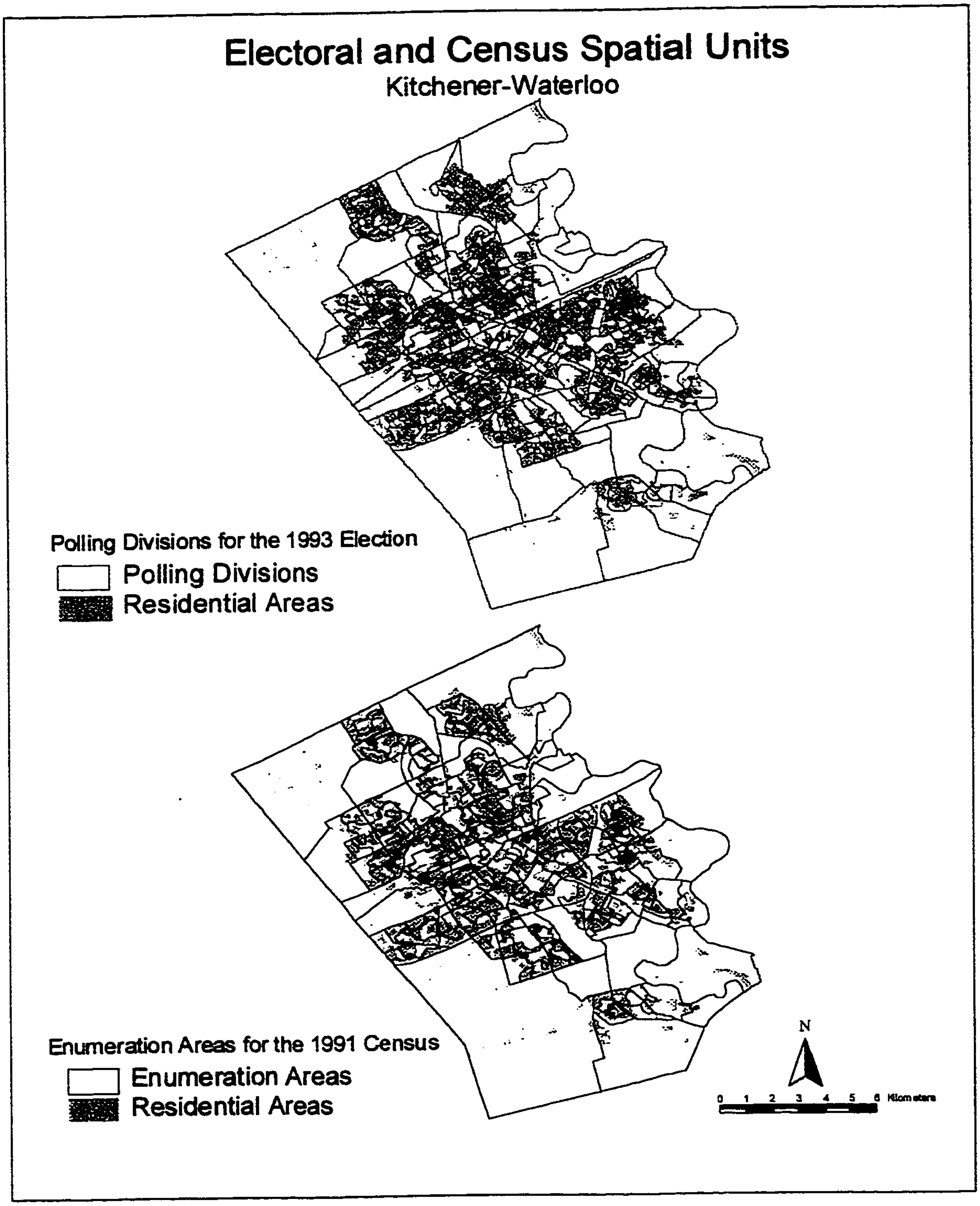
Analysis of voter participation required electoral and social information from a number of sources. Some of this was available directly from the federal government via the Data Liberation Initiative which facilitates the acquisition of usually expensive government data for educational purposes. Other sources of information were traditional paper electoral maps and databases and had to be digitized for integration into a geographic information system (GIS). The data in this stage can be divided into spatial information and attribute information. The spatial information consisted of the 1993 polling divisions (PDs) for the electoral information, and enumeration areas (EAs) for the social and economic indicators of the 1991census. Attribute data refers to the unique characteristics of the spatial units of the census and election. The complete 1991 Census was available on CD-Rom. Social data availability was not a problem because of this although reducing the database to a small number of social indicators was a lengthy process that required examination of previous research. Maps for the road network and residential areas also had to be acquired for this research.

SPATIAL INFORMATION

This research required fine scale spatial units to capture data variation throughout the Kitchener-Waterloo urban region. Digital enumeration area maps and road networks were obtained from the government but the polling divisions and residential areas had to be digitized from paper maps. This required a lot of time since great efforts were made in order to achieve a high quality of database. Great care had to be taken during the creation to ensure that boundaries were precise. A boundary error of only 20 metres may not be significant in a riding or census tract level analysis but could make quite a difference at the much finer level of spatial resolution used in this work. Figure 6.1 shows the enumeration areas, polling divisions, and residential areas in Kitchener-Waterloo. They will be discussed in detail further on in the chapter.

GIS was very useful for the creation and management of the spatial database required for this research. GIS was required to create new variables, organize large amounts of data, tabulate spatial relationships, prepare data for use in spatial statistical modelling packages, and interpolate poorly matched data to new boundaries. The application of interpolation techniques to electoral data is not unique since Smith (1967) performed something similar in his analysis of turnout and questionnaire response in Saskatoon. GIS does allow this process to be performed more quickly, and knowledge of land use patterns can be incorporated in order to get a more appropriate distribution of data over space.

Figure 5.1) Electoral and Census Spatial Units



Digitizing

MapInfo was used for all digitizing since considerable time was needed to capture the polling divisions and residential areas, and an easy, simple to use package was essential in order to speed up this process. Spatial information was digitized from schematic maps rather than directly from a paper map. This avoided the registration problem inherent in using paper maps of varying sizes, projections, and levels of accuracy with a digitizing tablet. The boundaries were typically digitized on screen with the Statistics Canada digital street network file of the region used as a base map for all other layers. When digitizing, lines and polygons can be anchored to existing features and layers. This ensured that boundaries were consistent when digitizing different types of information.

MapInfo was good for digitizing but had few capabilities for more advanced spatial modelling and GIS overlay operations needed for attribute interpolation. These were available in *Arc/Info* GIS and export of the files was done using the *ArcLink* utility for *MapInfo* to create *Arc/Info* export files. The polling station and street network files caused no transfer problems as they were point or line files which are structurally simple files to export between applications. The enumeration areas, polling divisions, and residential areas were much more difficult since they consisted of polygons. *Arc/Info* GIS organises spatial data based upon its topology. All polygons have shared boundaries and topological information is stored regarding connectivity and neighbouring polygons. *MapInfo* does not support topology and all of its spatial entities exist independently with no information about which entities have common boundaries. Overlapping polygons are not permitted in *Arc/Info* but can occur in *MapInfo*. Due to the nature of digitizing it is impossible to avoid overlapping polygons in a contiguous polygon coverage unless the digitizing is done without a single error. The digital database had to be purged considerably in order to eliminate duplicate points and nodes, overlapping polygons and spikes caused by clipping of improperly matched polygons. Several programs obtained via the Internet aided in tracking down these problems and correcting them before *Arc/Info* could read the polygon data. This is a difficult package to learn and was used only because other available programs could not perform the operations required. All of the geographic databases were exported to *ArcView* after the overlay operations were completed since this program is simple to use and offers the data management capabilities needed for the statistical analysis. Ironically, after the *Arc/Info* phase of the research was finished, a free extension for *ArcView* was obtained from the Forestry Department of the state of Oregon that allowed all of the complex *Arc/Info* operations to be done easily in *ArcView*.

Streets

The road network files for Kitchener-Waterloo were important since they were used as geographical references for the digitizing of other layers of information. Furthermore, roads serve as the most common boundary used in the partitioning of space into polling divisions and enumeration areas. The street network was also useful for orientation when examining the database. The street network files consisted of street names, road classification types, and the address ranges of each block face. This was especially useful in verifying the location of polling stations and other locations of interest. Railways, water features, municipal boundaries, and powerlines are also included in the street network file resulting in an extremely large data file. The major roads and highways of the region were extracted into a separate file once the digitizing of polling divisions and residential areas was completed since display of this large file slowed down operations on the computer.

Enumeration Areas

The census geography at the enumeration area level was obtained in digital form from Statistics Canada. Enumeration areas for the cities of Kitchener and Waterloo consist of 350 contiguous polygons. The spatial resolution is sufficiently fine that apartment buildings are often assigned their own individual EA. When the boundaries of these spatial units are determined, an attempt is made to follow features in the urban and natural landscape. Boundaries generally follow roads, rivers, railways, and powerlines. Municipal limits are always respected and EAs are never be split by such a boundary. All parts of the region are assigned to an EA and census units often contain large amounts of open land, industrial zones, parks etc. as well as residential areas. These non-populated regions cause problems when density calculations are required. Furthermore, unpopulated areas within an EA will introduce errors during geographic operations that require a good knowledge of the population distribution within the polygon. Such a case is an area weighted interpolation of data from one set of boundaries to another. A further problem was encountered when it was discovered that not all enumeration areas even contain residential areas or a population for census purposes. These are known as abandoned polygons and examples of these are hospitals. They exist as separate enumeration areas in a geographic sense but contain no resident population. Another example that is somewhat more perplexing is an area of scrubland near an interchange on Conestoga Parkway in Waterloo that corresponds to a graveyard. No population figures are listed for this area. These enumeration areas were removed during the area weighted interpolation discussed below.

Polling Divisions

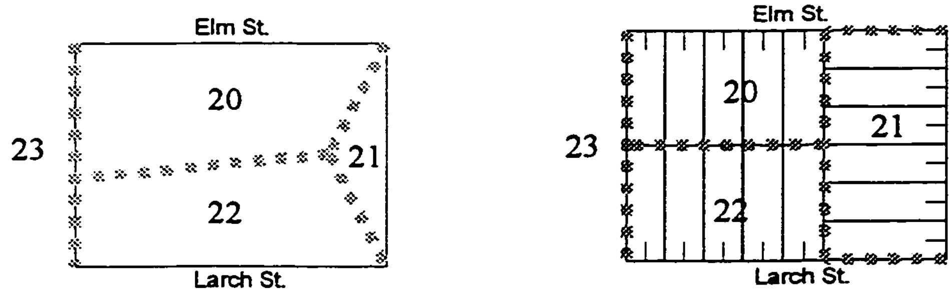
Geographic information regarding the polling divisions was obtained from Elections Canada. The organization has digital copies of these boundaries but they were not available for use outside the agency. These files were also used for organizing data tables based solely on linear street information rather than polling division area and probably would have had to be heavily modified had they been made available. It took considerable time to digitize a good electoral coverage but obtaining the Elections Canada digital data would probably not have sped up the process significantly. Instead, schematic paper maps were obtained and digitized on screen using the street network as a spatial reference.

The polling divisions for Kitchener-Waterloo consist of 424 contiguous areas that completely cover all parts of the region. Like the enumeration areas, attempts are made to follow clear geographic features such as streets. The number of electors will vary but an attempt is made during the delineation of polling divisions to achieve a target population of 350 electors per polling division. This process occurs during the redistribution of electoral ridings that occurs by law after every census. The electoral boundaries in use for the 1993 event were drawn up in 1987 but updated according to electoral enumerations undertaken for the 1988 election and the 1992 referendum. As a result of these updates, polling divisions can be combined if it is felt by the returning officer that there are too few electors to warrant a separate polling division. Alternatively polling divisions can also be split if the returning officer decides that there are too many electors. This manner of this split can vary according to the whim of the individual returning officer and can be performed geographically or according to some arbitrary means such as the street name or the elector's surname. Polling divisions that were split geographically (eg. 170 becomes 170-0, 170-1, and 170-2) were treated as individual polling divisions for this study but non-geographical splits that occurred based upon the surnames of electors (eg. PD 170 becomes 170A and 170B) were simply aggregated.

Polling division boundaries, like census units, respect municipal boundaries. Beyond this, polling division boundaries will follow any convenient feature in the landscape in order to partition space. The returning officers also have different styles of delimiting space in their ridings. The Waterloo riding returning officer had a preference for running boundaries along streets resulting in facing neighbours being assigned different polling divisions. The Kitchener returning officer generally preferred to run boundaries along lots between houses and keep facing neighbours in the same polling division. Polling divisions are supposed to contain approximately 350 electors when they are established but, this can vary considerably depending on population growth and how often the boundaries were revised.

The digital street network file served as a base for on screen digitizing of the polling divisions. When digitizing, boundaries can be attached to features in the street network file to ensure matching boundaries. In some cases boundaries would follow lots borders rather than streets, and other means would be required in order to accurately place the boundary. Digitizing the polling divisions proved difficult since the Elections Canada maps are intended as an indicator to enumerators into which polling division a street block

Figure 5.2) Digitizing Polling Divisions as Areal Units



A) Example of the composition of an orginal polling division map from Election Canada used in the enumeration.

B) Example of how this block was digitized to get an area based polygon coverage.

Polling division maps available from Elections Canada are intended to show enumerators which streets are in which polling division and are not in themselves good indicators of polling division areas. These maps must be interpreted and digitized as areas by examining the lot configurations. Map A tells an enumerator that all addresses on Elm St. are in PD 20 while all addresses on Birch St. are in PD 21. Map B shows where the boundaries actually run according to the lot configuration.

falls rather than as a true depiction of the precise area of the polling division. Figure 5.2 graphically illustrates the problems encountered when digitizing the electoral limits. A digital map of the civic lots for Kitchener and Waterloo would have been useful since much of this information was not available directly from the Elections Canada paper maps and had to be derived from other sources. Maps from local government were required to verify lot configurations and aerial photo surveys were consulted since they also reveal much about the lot configuration. Overall, digitizing the polling division boundaries required more time and effort than anticipated.

Polling Stations

The addresses and postal codes of the polling stations used in the 1993 event were obtained in paper form from Elections Canada. Polling station locations are not finalized until the election is called and can vary between events depending upon availability of location. In practise, during the chaos of an electoral event, the same locations tend to be used since it is more convenient, and both elections officials and longtime residents are familiar with the location. Typically, a polling station will serve several polling divisions although there is a considerable variance with as few as one or as many as seven.

A fairly precise location for each polling station is needed for distance calculations to be valid. Individual apartment buildings and block faces within a large urban area all have a unique postal code and this information provided an easy way to get a rapid estimate of where these locations occurred. A basic location file was derived by performing address matching of this data with the Statistics Canada postal code conversion file for Ontario. This file records the geographic coordinates in metres of the centre of the blockface for the specified postal code. The map produced by this process was verified against roadmaps that typically show the precise locations of the schools, churches, and government offices that are used as polling stations. The results from the postal code file were surprisingly accurate with only about 25% of the points requiring a shift to bring them to a more correct position. The postal code conversion file works well in dense urban areas where postal codes correspond to single block faces. The results would probably have been poorer had the rural regions north of Kitchener-Waterloo been examined in this manner. The street directory of the region also proved to be useful in verifying some of the addresses that were difficult to find. This reference source was also used to find the home location of some of the candidates. The political effects of the proximity to candidates was going to be examined in the manner of Johnston's work in Christchurch (Johnston, 1977b) but this idea was eventually discarded.

Residential Areas

A map of the residential areas in Kitchener-Waterloo was needed in order to get a good representation of population distribution for density calculations as well as the area weighted interpolation. A basic digital map of Kitchener-Waterloo was available via the Grand River Basin project based at Wilfrid Laurier University but this had to be modified considerably before it was suitable since the base map had been digitized from 1:250000 scale maps that do not have sufficient spatial resolution for use with fine spatial units. There was also no differentiation between residential areas and other forms of built up regions. At the EA/PD level of resolution it is necessary to remove all schoolyards, parks, industrial areas, rail and power rights of way, etc. to get a good representation of the residential areas. Information for this was obtained from a number of sources including aerial photographs and land use maps from the planning departments of local and regional government. The best single source of current land use information was a Kitchener Transit map of bus routes. Compilation of residential data required considerable time but was necessary considering the critical role this information played in the areal interpolation.

GIS USED FOR THE CREATION OF SPATIAL VARIABLES

GIS was an essential tool in building the spatial database and it was also used to calculate new spatial variables with which to examine the participation process. Population densities based upon the residential areas only rather than the entire collection unit area would have been difficult without GIS functions. The GIS also enabled the calculation of distances from the poll stations as well as the creation of an indicator for perceived distance by counting road intersections within each polling division.

Perceived Distance

One of the findings of Taylor's examination of participation in Swansea found that perceived distance was an important measure of participation. Much research has been performed in the field of behavioural psychology regarding factors that affect distance perception in an urban environment. Factors found to affect cognitive distance are quite varied and include visibility of the destination (Nasar, 1983, p.279; Nasar, 1985, p.627), the choice of alternate routes as determined by intersections (Sadalla, 1980, p.172), as well as greater land use density which seems to exaggerate distance perception (Walmsley, 1984, p.69). Overall, it seems that close distances are overestimated while distance to further destinations are underestimated (Walmsley, 1984, p.68). Most researchers seem to agree that the number of barriers that must be traversed in order to reach a goal will influence the perceived distance to that goal but disagree on the best ways of measuring these barriers. In an urban environment, perceived barrier would be road intersections, storefronts, parking lots, etc. Taylor measured perceived distance by how many road intersections were traversed. Taylor was examining the participation question in very small study area and was able to trace individual routes. This research area is much larger and more complex, and many possible routes can exist to reach a polling station. Instead the number of road intersections occurring within a polling division is used as a measure of the complexity of the spatial structure of the road network in that area and therefore the number of barriers that must be traversed to reach a polling station. Road intersections can also provide a crude measure of the alternative routes available to the elector. The number of alternative routes to a location was found by Sadalla to be a factor affecting distance perception.

Extracting intersection points required an *Arc/Info* command called NODEPOINT that converted all of the junctions in a street network file into a point file representing road intersections. This was overlayed onto the polling division polygons, which enabled the number of intersections per polygon to be calculated. The number of intersections within 10 metres of each polling division was also included as well as those falling within the boundaries. Intersections falling on the edge of two polling divisions act as barriers to both. Also, the boundaries of polling divisions whose edges should follow street patterns may have shifted slightly during the cleaning and dissolving of sliver polygons. The sliver polygons encountered during this research seemed to have areas of about 100 m^2 and the square root of this figure was used to provide a linear indication of how far a boundary could have shifted. Considering that most sliver polygons are very long and extremely thin the figure of 10 metres is probably a gross overestimation of how far boundaries may have shifted during the clean up of sliver polygons, but it did serve to ensure that all points falling close to the polling division boundaries were counted.

The number of nodes per division is not a perfect measure of cognitive distance since it does not take into account direction of travel. Also, some of the line endpoints were caused by phenomena other than road intersections. Street lines in the street network file are broken when they traverse railways or powerlines and this creates a node point. These were left in the database partly because it can be argued that they also represent a form of barrier that must be traversed but also because it would have been extremely difficult to purify the database since there were thousands of points that would have had to be examined visually. The results of other variable interactions were already being examined when the decision to add the nodes was made and socioeconomic factors were already explaining a big part of the variance. It was decided that it was not worth spending a large amount of time and effort on a variable that would probably not add greatly to the explanatory power of the model.

Distance to Polling Station, Residential Area

Other geographic variables were not as complex to measure. Physical proximity to polling stations was easily calculated in *ArcView* using the UTM coordinates of the polygon centroids and the coordinates of the stations. The coordinates were obtained during digitizing in *MapInfo*. Other geographic information depends upon polygon area and perimeter information. This is automatically calculated in *Arc/Info* coverages and is updated after all overlay operations. The areas of the polling divisions and the residential areas of the polling divisions were used as variables without further modifications while population density was calculated using the interpolated population figures and the residential area. This yields a more correct figure than using the original polling division area. However, this figure is still subject to error since the residential coverage contains errors. Batty (1994) proposes a solution by using the density of the residential street network as a measure of population density when errors are present in land use information. However, for simplicity, the more traditional approach was used.

Proximity to Major Roads

The proximity of the polling station to a major road is an indication of the accessibility to a polling station from the rest of the surrounding area. This geographic variable was easily derived from a classified road network by measuring the distance from a polling station to the nearest point on the closest major road. All road segments in the street network file had classification types and segments representing highways, boulevards and avenues were extracted to a file representing major roads. The NEAREST command in *Arc/Info* calculates distance between a polling station and the closest point of approach of the nearest road. It was hypothesized that a polling station that was easy to access by automobile would be more likely to be frequented than one that was not. A landmark close to a major road would also be highly familiar to people who frequently pass by. In this manner, the measure of road proximity may act as a substitute for the visibility measure found to be an important factor in perceived distance by Nasar.

Eccentricity

Phenomena such as voting and the socio-economic characteristics measured in the census vary over space but do not necessarily vary according to the spatial units of electoral and census information. This variability means that it is possible that the shape of these units can affect the measurement of phenomena and important trends can be exaggerated or even missed. This leads us to the dark side of politics and the process of gerrymandering. Gerrymandering occurs where the structural make up of electoral units can be manipulated to concentrate or disperse the voting power of groups of special interest. Analysis of this spatial phenomena is quite common in the study of electoral geography. Generally, deliberate manipulation of political boundaries would occur at a riding level. There is no reason to suppose that the polling units in the study region have been manipulated to this effect. The possible presence of structural effects is not by design. Area and perimeter information by themselves say little about the shape of a spatial unit but their ratios serve as a measure of the geometric complexity. Compact spatial units will have low values while more complex objects with longer perimeters will have higher values. A constant of 1000 was applied to this variable to convert the small numbers resulting from this process to something more meaningful. This was done to ease interpretation since the relationship between values of 0.0002 and 0.0007 is more difficult to visualize than that of 2 and 7.

GIS proved very useful in managing the large database. Occasionally it was necessary to examine geographic subsets of the database and GIS was able to rapidly extract the region of interest for analysis in a statistical package as well as create a new spatial connectivity matrix which is necessary for the calculation of spatial statistics.

ATTRIBUTE INFORMATION

Attribute data for this analysis came primarily from two sources: the 1991 Census on CD-Rom which provided the socioeconomic characteristics of the region at the enumeration area level and the voting results of the 35th Election, available in paper form from Elections Canada, which provided the polling division level information. In many cases categories had to be aggregated since the numbers would otherwise be too small to have meaning by themselves. An example of this is the votes cast for fringe parties. Most of these parties could only muster a couple of votes in each polling division and had to be combined in order to get an effective measure of the fringe vote. All of this attribute data was organized into records for each polling division or enumeration area. This spatial link served to associate the tables of non spatial data with the digitized boundaries of the spatial units in the GIS.

CENSUS DATA

The complete census of Canada is available as count data at the enumeration level on a CD Rom. This enormous amount of data needed to be processed and aggregated in order to be intelligible. What was needed was some indicator variables that enabled a socioeconomic profile to be built of each polling division. Previous research into the voting process (Eagles, 1991) has identified that the following population characteristics are important in affecting the voting process: age, occupation, employment, income, population mobility, immigration, and religion (Meisel, 1967; Gidengil, 1992). The indicators selected from the census would have to reflect the variability of the population with regard to these characteristics. In all cases, the raw count figures were interpolated to the polling division boundaries after which they were rated by percentage to enable meaningful comparisons to be drawn between spatial units. In order to rate the data, the denominator for the different categories also had to be recorded and transposed to the new boundaries. Counts for the total population, total labour force, and number of households were required even they were not directly incorporated into any analysis.

Databases are rarely free of error and the census database is no exception. Ideally, the population data to use would have been collected on the day of the vote but the closest census in a temporal sense was June 4th, 1991. Other factors affect the quality of the census data in addition to the two year time lag. There are two types of sampling errors that can occur: over-coverage and under-coverage. Over-coverage occurs when people are counted twice. This happens if people are enumerated at several locations due to travelling, etc. Undercounting occurs when people are excluded from the census. This type of sampling error is associated more with segments of the population that are highly mobile. Population groups susceptible to under-coverage are the young and immigrants. A sample of dwellings that were empty on the day of the census are revisited at a later date in order to verify that they were genuinely vacant rather than the occupants simply being gone for a short period. Not all households received the same census form. One in every five received a longer questionnaire with a greater number of more detailed questions. The 20% sample questionnaire had more detailed questions regarding income, ethnic origin, religion, language, and housing condition than the 100% sample. The 20% sample data was weighted by Statistics Canada following the census to reflect the results had they been obtained by a 100% sample.

Occupation

The 20% sample census of Canada contains detailed information regarding the occupational characteristics of the labour force. These had to be aggregated in a logical manner for meaningful trends to be identified. The original data encompassed 18 count type variables that contained too much variation to be meaningful if examined separately. These categories are designed to capture labour force variations in all parts of Canada and not all categories are appropriate to the study area. For example the labour force engaged in primary industries such as fishing, trapping, and forestry is not very high in urban areas like Kitchener-Waterloo while secondary employment such as manufacturing and transport is quite important. All urban areas contain employment in the tertiary sectors such as financing, real estate and education. The data was aggregated into a four labour indicators as shown by Table 5.3. These aggregate categories were then rated by expressing them as a percentage of the total labour force for that area. The rating was undertaken after the interpolation to the polling division boundaries. These categories are certainly not perfect and it is highly possible that other researchers could arrive at very different ways of combining this information.

Table 5.3) Indicators of Occupation

Indicator Variable	Statistics Canada Census Labour Category
Primary Sector Employment	• Agricultural and related service industries • Fishing and trapping industries • Mining (including milling), quarrying & oil well industries • Manufacturing industries • Construction industries • Transport and storage industries • Communication and other utility industries
Financial Sector Employment	• Whole sale trade industries • Retail trade industries • Finance and insurances industries • Real estate operator and insurance agent industries • Business service industries
Public Sector Employment	• Government service industries • Educational service industries • Health and Social Service Industry
Service Sector Employment	• Accommodation, food and beverage service industries • Other service industries

Note: These categories were rated by the total labour force for each spatial unit.

Age

Age was identified as one of the most important factors in voter turnout by the Royal Commission on Electoral turnout (Eagles, 1991). The Census of Canada (20% sample) gives age information for each sex based upon five year cohorts. These were all combined in order to arrive at a more compact measure of age. Firstly, sex differences were not of interest and the equivalent age cohorts of males and females were aggregated. The ratio between male and females at the polling division level showed very little variation. This was not pursued any further since it was believed that this information would act as a constant rather than as a variable. This is not to imply that the gender is not a factor in the participation process. The voting information recorded by Elections Canada does not record the sex of the voters who cast ballots. Even if it existed, release of this type of information would probably severely compromise the sanctity of the anonymous process of voting. The role of gender in the participatory process is best examined using different methods that focus on the individual such as a surveys rather than an ecological examination such as that undertaken here. The relative strengths and weaknesses of these approaches will be discussed elsewhere.

The population was aggregated into categories that reflected the following groups: young adults, early middle age, later middle age, and the elderly. The ages of the population are important since people will have different concerns, objectives, and needs depending upon their age. This will reflect upon how they participate in the electoral process which is one of the few ways in which people can make their opinions and concerns known to those who hold or hope to hold power. Many ways of aggregating the data could be undertaken in order to achieve these categories. The youth category was created by merging data from the 15-19 and 20-24 age cohorts. This category does include some residents that are below the legal voting age at the time of the 1991 census although the two year time lag between the census and the election means that some of these underage residents will have been eligible to vote in 1993. This age group corresponds closely to the university aged group that are very prominent in the study region. This measure is probably the most appropriate way of looking at the influence of this group since the student population is quite diverse, from many different social backgrounds and will originate from many different areas, some of them a long distance away. The early middle age group was formed by merging the cohorts of 25-29, 30-34,and 35-39, while the later middle age group aggregates the cohorts corresponding to ages 40 through 64. These latter two groups were distinguished since they would appear to have different needs and demands on society and its political system. The earlier middle aged group are still relatively new in the workforce and would be more likely to consist of younger families. The later middle age group would be more firmly established in employment as well as being more likely to have children who have left home. These two groups would likely have different concerns on many issues such as national education or daycare programs. Those aged 65 and over make up the final age grouping comprising the elderly. The age of 65 is often the year of retirement, although many people now seem to be retiring earlier if possible, therefore this was used to define the age group of the elderly.

Residential Mobility

Residential mobility has been identified as a important factor in electoral participation (Eagles, 1991) and the census provides information regarding a residents' mobility. The number of residents in an enumeration area who have moved in the past one year or five years previous to the questionnaire is known. This does provide a crude measure of mobility but unfortunately give no clues as to the distance moved. A resident moving within a community would probably maintain strong community ties and remain in touch with the norm of community attitudes and feelings while people moving in from outside the community or city would not have these ties.

Household Income and Education

Canadian political behaviour is quite different from other western societies since the social cleavages that are important determinants of political behaviour in other countries do not seem to function in the same manner in the Canadian context (Clarke, 1996, p.96). Factors such as average income and education level do not show up as being highly significant at a large scale analysis as performed by the Royal Commission on Electoral Reform (Eagles, 1991). However, these factors should be examined as a variable at a finer scale of analysis since they can be associated with different types of neighbourhoods. Neighbourhood political behaviour will not be present in ecological analysis at a riding level but may well be an important determinant of political participation when examined at a city level. Education data extracted from the census looked at the highest educational achievement of the population. This was measured as the number of people with less than Grade 9, number of people with high school education and also the number with university education. Average household income is incorporated into this research since it serves as a corelate for other factors for which there is no measurable data available. Average income would appear to be a good correlate for such factors as car ownership and personal mobility which conventional wisdom would indicate are important in the decision to make a trip to a polling station but for which no data is available at an enumeration/polling division level.

Religious Affiliation

Religion has been identified by some researchers (Meisel, 1967) as being an important indicator of political behaviour. Certainly there appears to be a persistent religious cleavage in Canadian political behaviour although there is no clear evidence of religion being a direct determinant of this behaviour. It appears that religious behaviour acts as a surrogate for values that maybe be shared in a community and passed throughout the generations. Meisel credits the community aspects of religious behaviour as an important political factor rather than the matters of a church policy or interests (Meisel, 1967, p.160). Gidengil advances the idea of religion as a social indicator of people with similar interests and lifestyles. Beyond sharing a common religious experience they are likely to share the same political traits. Several religious variables were derived from the census to explore the impact of religion upon Kitchener-Waterloo participation rates. The percentage of Catholic and Protestants were included since they are the largest religious groups. Also, the percentage of non-believers was also included as a variable.

ELECTORAL INFORMATION

The voting information for each polling division was recorded and released to the public in the official polling results (Chief Electoral Officer, 1993b). This data had to be computerized but presented no major problems other than being time consuming. This data could be used without modification since the spatial units of study are the polling divisions. However, as explained in the section dealing with the areal interpolation, abandoned polygons in the census coverage occasionally meant that a portion of the residential area of a polling division may not have socio-economic information and would have to be removed from the electoral database. Since electoral information is considered to be evenly distributed across the residential area of the spatial unit this removal affected the absolute values of information but had no effect once the data had been rated. This problem occurred in only a few observations.

Participation was defined as the total ballots cast as a percentage of the number of registered voters. The ballots cast includes valid votes as well as those that have been classified as being spoiled. Another political variable examined was the percentage of support for each of the major parties. Fringe party support has been postulated to be an indicator of voter discontent (Eagles, 1991) and is incorporated in this research. It is important to consider the fringe vote since votes are cast for candidates and parties that have very little chance of winning a seat, and can very much be thought of as a vote of protest against the larger established parties. Some parties that contest an election are clearly not meant to be taken seriously and a well known example of this is the Monster Raving Looney Party in the UK. The only party in the 1993 event whose campaign appeared to be humourous was the Natural Law Party. Votes cast for these parties can be considered protest votes. However, problems are encountered at the polling levels since support for each individual fringe party is too small to be meaningful in a statistical sense. Instead, the fringe parties have been aggregated to obtain a fringe vote. However, caution must be used when interpreting this fringe information since a number of groups outside of the political mainstream with very different political agendas have been merged. As an example, two groups classified and combined into the fringe category are the highly traditional Christian Heritage Party which emphasizes family values and the Natural Law Party which advocated meditation and levitation as a solution to the federal spending problems. Finally the plurality was calculated for each polling division. This is a measure of the closeness of the political race in that area. This is simply the difference in voter support between the two parties achieving the highest support. All electoral data was rated in order to allow comparisons to be made.

Table 5.4 shows the final list of variables used in the undertaking of this research although many of these variables were discarded during the analysis. The compilation of this data resulted in a socio-economic dataset based upon the enumeration areas and an electoral database based upon the polling divisions. Analysis of the participation data requires a further process to match one dataset to the boundaries. The enumeration areas were interpolated to the polling divisions since there were a number of geographic variables of interest that were directly related to the structure of the polling divisions. Among these were actual and perceived distance to polling station measures. GIS was useful in manipulating and managing the data as well as in the creation of uniquely geographic variables. It also has an important part to play in matching the enumeration areas to the polling divisions by means of an area weighted interpolation.

Table 5.4) Listing and description of variables

Geographic Variables	Description	Units
CLIPAREA	Residential area	$Meters^2$
AREA_LOG	Residential area (logarithm)	Log ($meters^2$)
DISTANCE	Distance to polling station	Meters
DISTANCE_M	Distance to polling station (weighted)	Meters
DENSITY	Population density	$People/km^2$
DENS_LOG	Population density (logarithm)	Log ($People/km^2$)
ECENTRY	Eccentricity	Area / Perimeter
INTERSECTI	Road intersections	# per polling division
PROXROAD	Proximity to major roads	Meters

Social and Economic Variables	Description	Units
EDU_9	Highest educational achievement: less than grade 9	% of Population
EDU_HS	Highest educational achievement: high school	% of Population
EDU_UNIV	Highest educational achievement: university	% of Population
HS_SIZEA	Average number of people per household	Mean size of household
IMMIGRANT_	Immigrant population	% of Population
INC_HOUSE	Average household income	Dollars
INC_LOG	Average household income (logarithm)	Log (Dollars)
JOB_FINA	Employment: financial sector	% of Population
JOB_MAN	Employment: primary sector	% of Population
JOB_PUB	Employment: public sector	% of Population
JOB_SERV	Employment: service sector	% of Population
MOVE_1	Population that move within past year	% of Population
MOVE_5	Population that moved within the past five years	% of Population
POP18_24	Population aged 18 to 24	% of Population
POP25_39	Population aged 25 to 39	% of Population
POP40_64	Population aged 40 to 64	% of Population
POP65OVR	Population aged 65 and over	% of Population
RELI_CATH	Religious affiliation: catholic	% of Population
RELI_PROT	Religious affiliation: protestant	% of Population
RELI_NONE	Religious affiliation: none	% of Population
PARENT	Single parents	% of Population
RENTED	Population renting accommodation	% of Population
SINGLE	Unmarried population	% of Population
UNEMPLOY	Unemployed population	% of Population

Political Variables	Description	Units
LIBERAL	Liberal party support	% of Ballots Cast
MAJORITY	Plurality	% of Ballots Cast
NDP	NDP party support	% of Ballots Cast
OTHER	Fringe party support	% of Ballots Cast
PARTICIP	Electoral participation	% of Electors
PC	PC party support	% of Ballots Cast
REFORM	Reform party support	% of Ballots Cast

Chapter Six

Methods

Elections are rich sources of numeric data that are well suited to the use of quantitative methods. Regression analysis is a common technique used in geographical analysis and allows the characteristics of a set of independent variables to predict a variable of interest. This section will examine regression analysis as well as the need to couple this technique with spatial diagnostics. The role of space is often ignored in regression analysis but it can have some serious effects as discussed below. However, for regression analysis to work, all variables must be spatially referenced to common spatial units. This is a problem since the base collection units for the election and census do not match. One dataset had to be interpolated to the other boundaries in order to perform any analysis. An area weighted interpolation was used to transpose the census enumeration information to the polling division boundaries. The interpolation was used for building the spatially referenced database rather than modelling of participation but is included in the methods chapter rather than the data chapter since the interpolation was quite complex and computationally intensive.

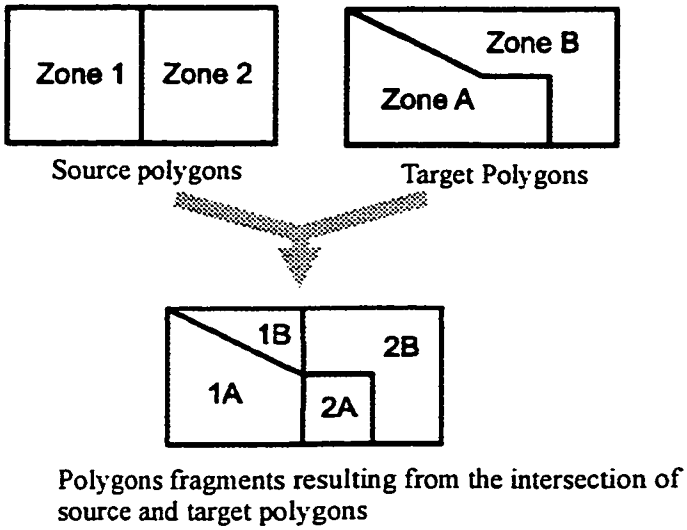
Area Weighted Interpolation

Census data at the enumeration area level and electoral information at the polling division level need to be combined in order to determine the social characteristics of voter participation. This cannot be done directly since the boundaries of the geographic units do not follow the same borders. The complication of dealing with mismatched boundaries is probably one reason why fine scale analysis is uncommon in electoral geography and political science (Reynolds, 1969, p.122). Both types of spatial units respect city limits but beyond that different criteria are used to specify limits. Enumeration areas must aggregate into higher units such as census tracts, census divisions, etc. while polling division boundaries are left up to the judgement of the individual returning officer. Social analysis at the fine scale cannot proceed until the electoral and census databases are integrated into one database using a common spatial unit. This process of interpolation can be complex and computationally intensive. The introduction of GIS has greatly improved the ability of researchers to perform fine scale electoral analysis although the work of Smith (1967) shows that it was not impossible to perform interpolation of fine scale electoral and census units before the wide spread introduction of GIS software.

The problems involved in matching a geographic database to another set of boundaries received little attention until the 1980s when introduction of GIS allowed non-coincident geographic data to be integrated. Lam (1983) provides a comprehensive review of interpolation procedures for both point and area data. Since the two datasets of interest in this research are area based, this examination will be restricted to areal methods only. Two sets of spatial units are involved with the source polygons representing the original database and the target polygons being the spatial units to which the original data will be fitted. Area interpolations can be one of two forms: volumetric and non-volumetric. Non-volumetric techniques use a grid of points to sample the source polygons and perform point interpolation such as spline-fitting, kriging, and distance-weighted least squares. The target polygons are assigned data values based upon averages of these point values. These techniques are considered poor practise (Lam, 1983, pg.139) since they are highly dependent upon the spatial arrangement and density of sample points, and also do not preserve the total value of the data. The process can not be reversed to recreate the value in the source polygon.

Volumetric interpolation preserves the value of the data and also uses the shape and size of the zone as a unit of operation instead of an arbitrary sampling pattern. This technique is intuitively quite simple in that it involves overlaying source and target zones and looking at the areas of intersection between the two sets of zones. These zones can be thought of as "fragments" of the original source polygons which can be assigned data weighted according to the proportion of the source polygon area occupied. These fragments can be reaggregated according to the spatial configuration of the target boundaries. Figure 6.1 portrays this process visually. The overlay technique was the method chosen for this research since the value of the original data is preserved, its polygon overlay technique is well suited to GIS, and it is conceptually easy to comprehend.

Figure 6.1) Union Polygon Overlay



A UNION polygon overlay operation

In an area weighted interpolation new values for target areas are computed based upon how much of the source area is contained by the target area. The mathematical equation for the area weighted interpolation of count data is as follows:

$$V_t = \sum_{s} V_s * A_{ts} / A_s \qquad (1)$$

where V_t = value of the target polygon, V_s = value of the source polygon, A_{ts} = area of the intersection of the target and source polygons and A_s = area of the source polygon

Assuming that a population of 400 is spread evenly throughout an area, then half of that area will yield a population of 200. However not all data can be handled in this manner. Interpolation of rate data (e.g. averages) is more complex since this figure cannot be partitioned according to space. Using the above example, if the average income for the 400 people is $26000, then the average income for half the area remains $26000. The solution to this problem is to convert the rate data into count data by multiplying it by the number of observations (Lam, 1984, p.140) e.g. number of people with income in the case of average income, and area weighting this figure to the new boundaries. The number of observations must also be interpolated before a new rate figure can be calculated. Incorporating the observations into equation (1) to cater to rated data results in the following equation:

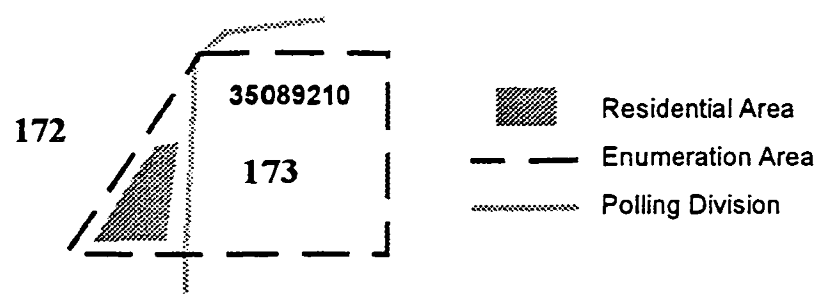
$$V_t = \left(\left(\sum_s V_s * Obs_{st} \right) * A_{ts} / A_s \right) * Obs_s \tag{2}$$

where V_t = value of the target polygon, V_s = value of the source polygon, A_{ts} = area of the intersection of the target and source polygons, A_s = area of the target polygon, Obs_{st} = number of observations in the area of the intersection of the target and source polygons, and Obs_s = number of observations in the source polygon.

In some cases the number of observations for a rated variable may not be known and Lam suggests using the area of the source polygon as a substitute. Many of the variables used in this research are rated but this was done after the count figures had been calculated according to (1). Only two rated variables had to be interpolated for this research (average income and average household income) and in both cases the number of observations was known.

This method makes the important assumption that the data for the source and target polygons is evenly distributed over the entire polygon. In reality, this is a difficult assumption to meet when considering how population is distributed throughout an urban area. Our cities are composed of parks, shopping centres, business parks, industrial areas, lakes, railway yards etc. as well as places of residence. In addition the roads, freeways and parking lots take up large amounts of space. All of these are large areal features which are allocated to a census or electoral unit. At higher levels of aggregation the presence of non residential areal features will probably not make much difference to an area weighted interpolation but it was felt that these features could cause considerable problems at the polling division scale. Some urban polling divisions are quite small in area and the presence of, say, a retail strip mall within its boundaries could have a considerable effect on the distribution of the population throughout that geographic unit. Conversely, in the outer areas of the study area, polling divisions span considerable distances encompassing fields, farmlands, and industrial parks. To make the assumption of even distribution in these cases would also add considerable error and noise to the database. A spatial unit that is almost completely occupied by non-residential space is of concern since the population may well be completely clustered in one location and should be assigned in its entirety to only one target zone instead of being partitioned according to the overlapping areas of the spatial units (Figure 6.2).

Figure 6.2) Residential Areas in Spatial Units



An area weighted interpolation can be inaccurate without knowledge about the location of residential areas within the spatial units. The population attributes of EA 35089210 should clearly be assigned entirely to PD 172 instead of being distributed proportionally to the area of the intersections of the polling divisions with the enumeration area..

Several ways have been proposed to account for this problem. Techniques that allow for normal distribution and binomial distribution of data over a geographic area have been proposed by Flowerdew and Green (1994). These techniques try to incorporate other variables to get a sense of where the population is located. The example demonstrated by Flowerdew and Green uses the number of individual post codes per postal sector as an indicator of population location. These techniques proposed by Flowerdew and Green also examine the effects of differing assumptions about distribution of data (normal, binomial, and Poisson) to combat the problem of missing data values. However, these methods have been developed for a more highly aggregated level of data than the spatial units of interest and might not be appropriate to the smaller units used in this research. Instead, a traditional area weighted interpolation was performed using only the residential areas of the database. The technique of removing areas that are of no interest in a spatial database is referred to as dasymetric mapping (Martin, 1991, p.147). For this method to be effective, a good landuse map must be available. Martin makes the point that dasymetric mapping still has its problems in that the landuse data may not have been recorded at a scale suitable for the analysis for which it is being used. Furthermore, once non-residential areas are removed, the spatial database grows considerably since often simple spatial units will be broken into larger numbers of more complex shapes. The first concern was answered by use of a very highly detailed land use map that captures the residential distribution at the neighbourhood level. The second concern may still be applicable to a very large study but the computing power and storage available to even students has evolved considerably from the time Martin was writing in the early 1990s and presented no problem in this research. Removal of the large regions of non-residential space means that the assumption of even distribution throughout the polygons, essential to an area weighted interpolation, can be met. Moreover, the fine resolution of the spatial units now strengthens this assumption since changes in population density and distribution will be occurring at a neighbourhood level and can be captured by the spatial units. Apartment complexes where the population density will differ drastically from the surrounding area present few problems since they are, for the most part, separate enumeration areas and polling divisions.

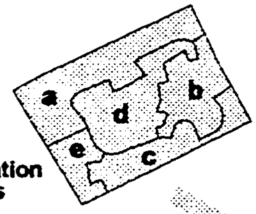
Area Weighted Interpolation using *Arc/Info*

The *Arc/Info* GIS was used to perform the dasymetric mapping as well as the areal interpolation. A layer of geographic information is referred to in *Arc/Info* terminology as a coverage. The area weighted interpolation requires several stages and a series of commands. Figure 6.3 provides a visual representation of the flow of operation required. Polygon overlays consist of spatially joining layers of data. One layer of polygons is overlaid onto another to produce a new polygon coverage. The attribute tables of the original two coverages are combined in the resulting coverage. This establishes a spatial relationship between features of the original input coverages. The spatial features retained in the new polygon coverage depend upon the type of overlay performed: an INTERSECT will act in the manner of a cookie cutter and will remove all features of the first layer that do not correspond spatially with the second coverage while a UNION will create an output coverage that consists of the zones of intersection between all features of the input coverages. This study required an INTERSECT overlay using residential coverage and enumeration areas to remove the non-residential areas from the census units.

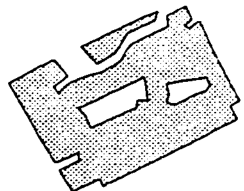
Figure 6.3) Area Weighted Interpolation using *Arc/Info*

AREA WEIGHTED INTERPOLATION USING *ARC/INFO* GIS

Population by Enumeration Area Interpolated to Polling Divisions
Northlake Neighbourhood, Waterloo



Enumeration Areas

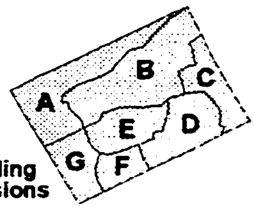


Residential Areas

Data Layers Required

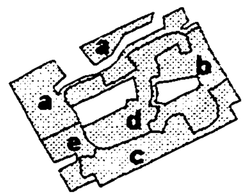
- Enumeration Areas
- Polling Division
- Residential Areas

An area-weighted interpolation requires the **INTERSECT** and **UNION** polygon overlay commands. In addition the **CALCULATE** and **STATISTICS** commands are needed to recalculate attribute data after the polygon overlays.



Polling Divisions

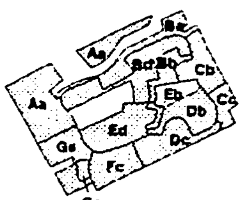
INTERSECT **Enumeration Areas** using **Residential Areas** to remove non-residential zones.



Enumeration Area	Residential Area (m²)	Population
a	375769	812
b	174810	1283
c	284241	1385
d	424906	1225
e	121809	565

Area figures calculated after the INTERSECT polygon overlay

UNION **Polling Divisions** and **Enumeration Areas**

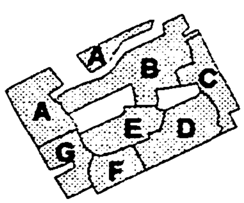


Intersection Zones	Area (m²)	Percentage of Original EA area	Population
Aa	269591	71.7436	582.6
Ba	106179	28.2564	229.4
Bb	57612	32.9569	422.8
Bd	171761	40.4233	495.2
Cb	102925	58.8782	755.4
Cc	48625	17.1070	236.9
Dc	115245	40.5448	561.5
Dd	107257	25.2425	309.2
Eb	14273	8.1649	104.8
Ed	145888	34.3342	420.6
Fc	104953	36.9239	511.4
Gc	15418	5.4243	75.1
Ge	121809	100.0000	565.0

Area figures calculated after UNION polygon overlay

Calculate new populations for intersection zones resulting from UNION based upon area percentage ratios between original polygon and its fragments.

Sum population of intersection zones according to their equivalent polling division. The population figures can be attached to the polling division coverage.



Polling Division	Interpolated Population
A	583
B	1147
C	992
D	871
E	525
F	511
G	640

At this point a further complication arises since spatial units that used to be single polygons are often fragmented into a multitude of polygons. *Arc/Info* deals with this by copying the attribute data into a record for each one of these and recalculates area and perimeter information for each. The areas of all these polygons must be added in order to get the correct residential area for the spatial unit. The RESELECT and STATISTICS commands can be used to calculate the areas of census units occurring in residential zones.

In order to identify zones of intersection between enumeration areas and polling divisions, it is necessary to combine the electoral and census spatial units into one composite coverage that retains all of the boundaries from the two original files. A UNION overlay operation in *Arc/Info* will combine coverages in this manner. The result is a coverage of zones of intersections based upon the intersection of the boundaries of the two input coverages. Throughout these operations the polling division and enumeration area identities are retained and each intersection zone in the composite coverage will have a link that identifies its parent enumeration area and polling division. This link enables the percentage of the source polygon area contained in the intersection zone to be calculated and used to weight the source attribute data. The process of recalculating the values for attribute data can be done within the GIS environment using the CALCULATE and RESELECT commands but the researcher found the *Info* database manager that is part of *Arc/Info* was archaic and difficult to use compared to the simplicity of Windows based database managers. Import and export of attribute data from *Arc/Info* is quite easy and the retabulation of data was done in *MS Access*.

Another problem encountered during the interpolation involved residential areas where electoral data existed but for which the census recorded no population data. An example of this occurs with hospitals, hotels, and sometimes old age homes which are designated as separate enumeration areas. No population data is recorded and these areas are known as abandoned polygons. The electoral information for the section of the polling division corresponding to the abandoned polygon had to be purged from the database. This was done by dividing the electoral information into the target polygons of the composite coverage. All residential polygons that formed part of an unpopulated census unit were deleted from the database. The polling division identifiers were used to reaggregate the electoral information. The overlay based areal interpolation is value preserving and the reaggregated electoral information is identical to the original with the exception of electoral units that contained abandoned enumeration areas.

Sliver Polygons

Arc/Info polygon overlay operations such as INTERSECT and UNION used in the areal interpolation result in large number of "sliver" polygons. These are created when boundaries between two coverages are similar but not exact. Sliver polygons result in a highly fragmented and unnecessarily large database, and must be eliminated. The polygon cannot simply be deleted since that would result in a blank area between two geographic units that are meant to share a common boundary. Instead the sliver polygons must be integrated into one of its neighbours with a consequent adjustment in the boundaries of the polygon. Absorption of sliver polygons is a subject of research in GIS regarding the best method of achieving this. Some techniques apply artificial intelligence techniques to look at the similarity in attributes between sliver polygons and their neighbours in order to choose in which polygon the sliver will be absorbed. Others will use geometric techniques to split a sliver polygon between its neighbours. A much more common technique which was used in this research was to simply absorb the sliver polygon into the neighbour with which it shared the longest boundary. The ELIMINATE command in *Arc/Info* uses this method to absorb specified polygons into their neighbours. *Arc/Info* GIS does not have any routines to automatically deal with sliver polygons and the operator must intervene in order to identify and absorb the sliver polygons. This must be done after each polygon overlay operation but for the sake of simplicity these steps are not shown in Figure 6.3. The lack of automatic polygon sliver identification method in *Arc/Info* means that researchers must devise their own criteria and methods for identification and elimination of sliver polygons.

Locating sliver polygons is important since absorbing non sliver polygons will introduce considerable error by shifting boundaries and possibly even eliminating entire polling divisions and enumeration areas. Generally, slivers are easy to find since they tend to be very small in area, as well as being very thin in shape. They were identified by displaying the smaller polygons in a coverage to examine their shapes. Furthermore, a scree plot (Figures 6.4) of the ranked area of all polygons in the coverage was produced. Scree plots will reveal where groupings of polygons are substantially different in size from the rest of the database. Identification of slivers is an appropriate application of scree plots since their long, thin shapes result in very small areas that should be quite different from the rest of the legitimate polygons. A logarithmic scale was required because of the extreme differences in areas between the largest and smallest polygons. The scree plots show a clear break in the size distribution of the polygons at about 1000 m^2. The smallest polygons represented by the "drooping tail" of the scree plots are much

Figure 6.4) Sliver Polygons After Union Overlay

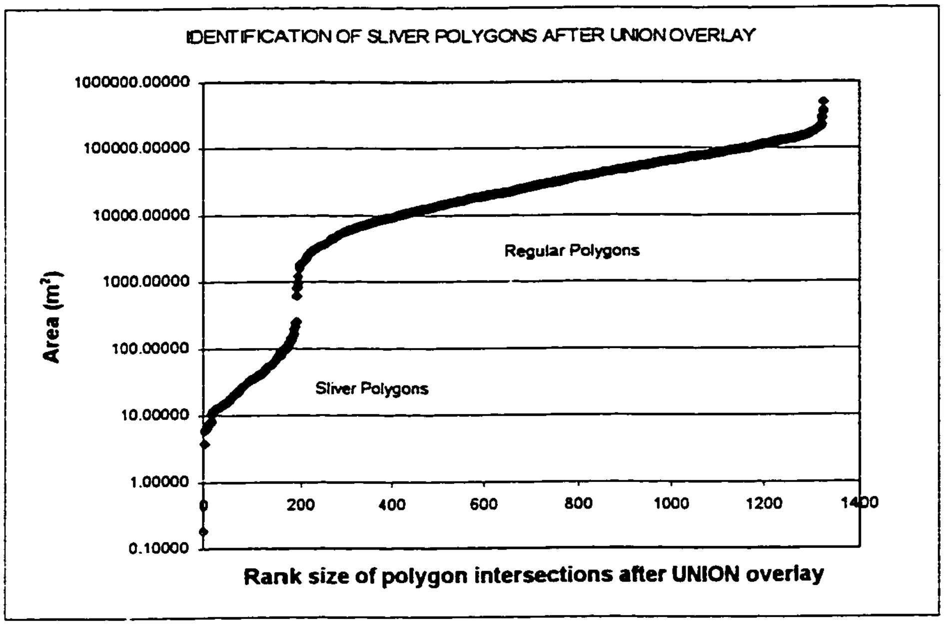
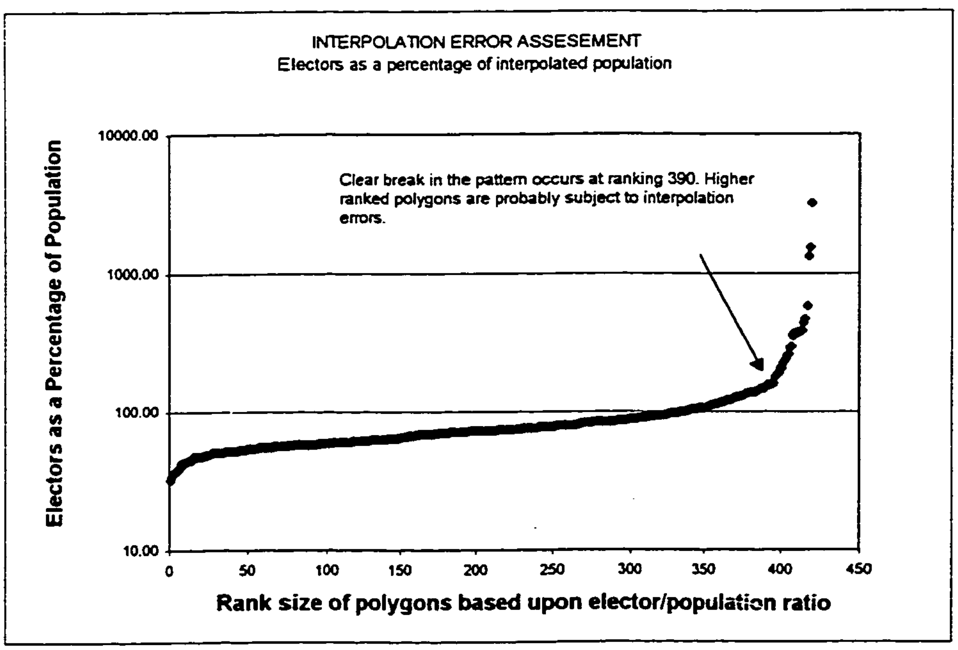


Figure 6.5) Interpolation Error Assessment



smaller than all of the rest of those in the distribution. All of these polygons were declared slivers and eliminated after a visual verification.

Interpolation Error Assessment

Any process of interpolation is subject to error. An attempt was made to examine how sound the interpolated database was since this information would form the basis of the participation modelling. An area weighted interpolation will be very sensitive to errors in areal coverage and the biggest source of error was reckoned to be the residential area coverage. Considerable time and effort was spent trying to get an accurate representation of the residential regions appropriate to the scale of analysis (see Chapter 5 - Digitizing) but inaccuracies will still inevitably occur and will manifest themselves in the database.

Sources of Error

Error in the spatial data come from the following sources:

a) Digitizing of enumeration areas boundaries (performed by Statistics Canada)

b) Digitizing of polling division boundaries

c) Digitizing of residential area boundaries

d) Non-uniform distribution of the population across the residential sections of an enumeration area.

Errors from digitizing will result in boundary errors which will yield sliver polygons during polygon overlay operations. Efforts have been made to identify and absorb sliver polygons into adjacent polygons but the boundary errors generated will still be present. However, the areas of these sliver polygons are minuscule in comparison to the polling divisions and enumeration areas and their effects during the area weighted interpolation are probably negligible. Errors resulting from non-uniform distribution are probably more significant but the assumption of uniform population distribution across a polygon is realistic at this scale once the enumeration area is corrected for residential zones.

One method of verifying how well the interpolation worked is to examine the ratio between the interpolated population for each polling division and the number of electors registered to that area. Although not all residents are registered electors, conventional wisdom would indicate that there should be a fairly high correlation between these two. Errors caused by improper residential coverage would manifest themselves as being fairly extreme in terms of population/elector ratio. The most obvious indication of error would occur if there were more electors in a polling division than population since the electors are a subset of the total population. These ratios were mapped and examined for extreme values. Having more electors than population is not necessarily indicative of an error since there was more than a two year time lag between the census data collection and the finalization of the electoral lists so there is a possibility for population growth and development of new suburban regions. Alternatively the high student population in Waterloo can lead to a large disparity between the electors and the enumerated population. The census of 1991 was conducted in the summer when many students had left town whereas the election was conducted in the middle of the fall term. The rank size of the ratios were plotted using a scree plot (Figure 6.5) to observe values that were substantially higher than the rest of the dataset. Approximately 80% of the polling divisions had, as expected, fewer electors than population while most of the remainder were not much higher than unity. Figure 6.5 shows a break in the pattern of rankings at about 390. At this point the number of electors begins to exceed the interpolated population by large amounts indicating that problems exist with these polling divisions.

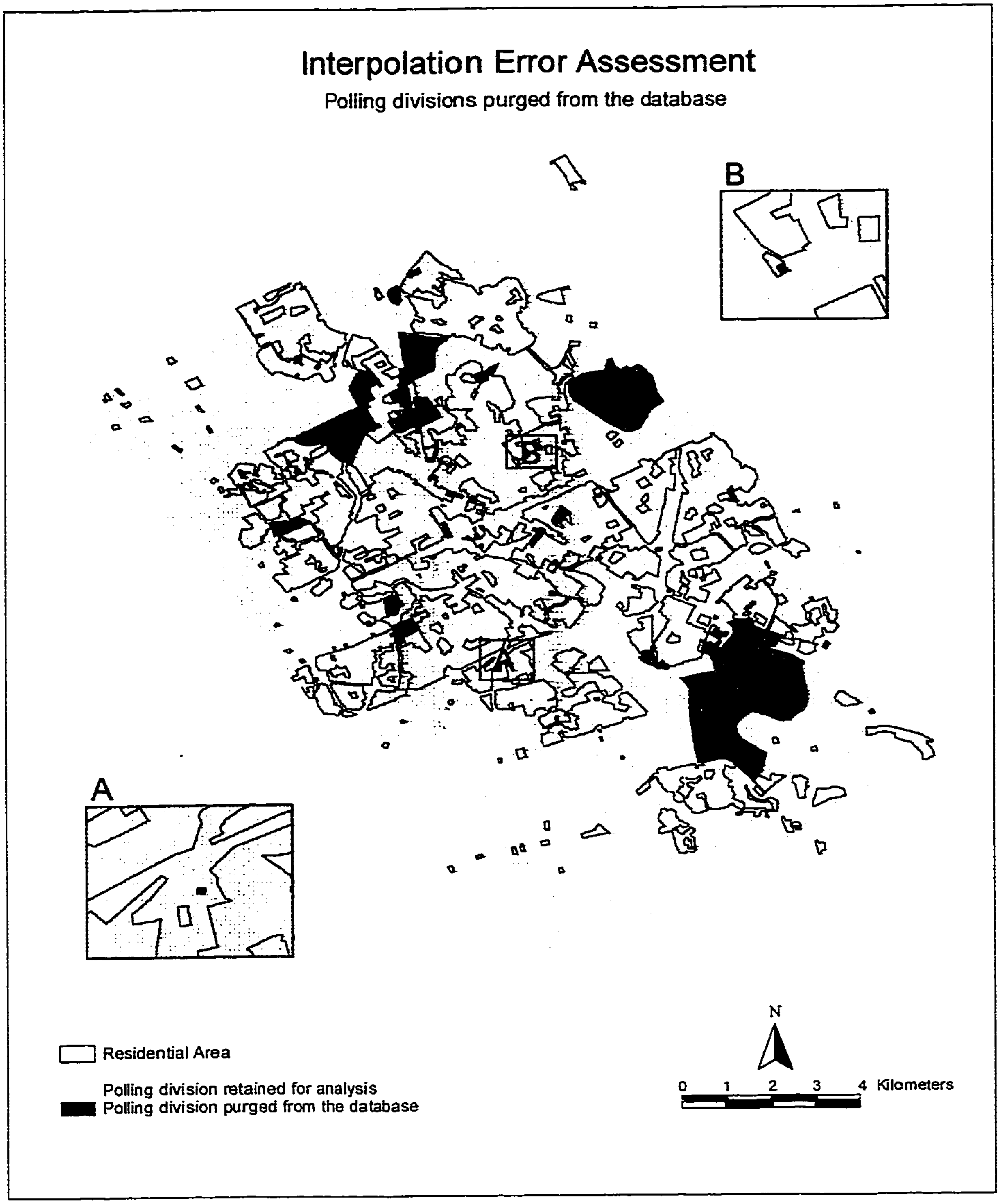
The distribution of higher than unity values were mapped in order to see if they corresponded with areas close to the universities that are expected to have a high student population or regions that have seen substantial suburban development. Some areas of recent development are Doon, Forest Hill Estates, Keats Way etc. It was found that many of the higher than unity polling divisions could be explained by their location in recently developed areas as well as fluctuations in the student population. However, the areas with the highest discrepancies (in some cases the number of electors was found to be 1500% higher than the interpolated population) were found to correspond closely with small polling divisions that are single building units. It seems that an area weighted interpolation works best in regions that are fairly large with a relatively smooth distribution of population. The density peaks caused by the concentration of people in apartment buildings that are not both a separate enumeration area and polling division will lead to strange results with the small areas and relatively high populations involved. A small error in the residential area map will have much more impact on an apartment complex than in a sprawling, homogenous suburban region. It is interesting that problems also seem to exist with apartment blocks which are separate enumeration areas and polling divisions and for which no interpolation is required. Problems with these particular polling divisions can be attributed to errors in the enumeration during either the census and/or election rather than problems during the interpolation process.

All polling divisions that had a ranking of greater than 390 as shown in Figure 6.5 were purged from the database. In addition to these several polling divisions that corresponded to old age homes or hospitals that had not been coded by Statistics Canada as abandoned enumeration areas were also removed. A total of 40 polling divisions of the original 424 were purged from the database. Figure 6.6 shows in red the deleted polling divisions. Most of these consist of polling divisions that have only a very small residential zone. Such polling divisions are probably caused by errors in the residential area coverage. In addition, many of the deleted polling divisions correspond to apartment blocks, some of which are quite small and difficult to see in the regional map. The area weighted interpolation seems to have worked well in the remaining 384 polling divisions and this provides a sufficiently large number of observations to allow statistical analysis to be performed with confidence. Errors are almost certainly still present in this database but there does not seem to be any indication of concentration or clustering of errors and the assumption of random error that is an important condition of spatial modelling is met.

Spatial Modelling

The methods used in this research consist of an exploratory data analysis of the dataset which will aid in building and interpreting a regression model of participation. The interaction of the different population variables helps explain the spatial variations in the participation process. A standard method of predictive modelling in the social sciences has been regression modelling where the interactions of a series of independent variables are related with a single dependant variable. Spatial scientists are also concerned with the distributions of values over space and the basic regression models must often be modified to account for the effects of these spatial distributions. The objective of the spatial modelling process is to produce a stable model that explains a high proportion of the

Figure 6.6) Polling Division Error Assessment



observed variance in participation. The incorporated variables should all be statistically significant and all spatial influences upon the variables should be identified and accounted for. These processes and terms will be explained below.

Spatial scientists differ from other researchers in that they are interested in the role of space and the spatial configuration of the dataset. This spatial configuration can have an important influence in statistical analysis and ignoring the spatial dimension will often negate the assumptions of normality and independence that are common in statistical analysis. Political analysis is no exception to this neglect of the role of space and a recent example is the participation research undertaken by the Royal Commission on Electoral Studies in 1991 which completely ignores any analysis of spatial effects. Before discussing the methods of factor analysis and regression modelling, it is necessary to examine the importance of the spatial dimension, how it can be examined and incorporated into a numerical analysis.

Spatial Dependency

The influences on a spatial unit within a spatial model can come from two sources: its own absolute characteristics and neighbourhood influences from the characteristics of surrounding spatial units. A spatial unit's unique characteristics are quite straightforward. These are the measured values for the variable associated with that spatial unit and these serve as the numeric inputs to a regression model. The residential area, the distance to a polling station, percentage of the population living in rental accommodation are all examples of the unique characteristics of a spatial unit. More complicated are the neighbourhood effects from surrounding spatial units. Firstly, they may not be present;

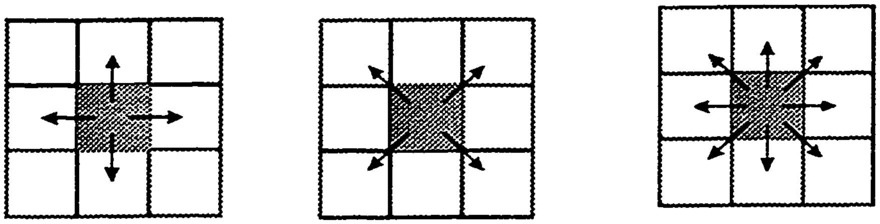
secondly, if present, their origin may be difficult to identify and accounting for the neighbourhood effects may be difficult to implement.

The presence of neighbourhood effects means that values in a spatial unit are not independent of their neighbours and are spatially dependent. This spatial autocorrelation can occur in two ways. Attraction of similar values among entities that results in the clustering of a phenomenon is known as positive spatial autocorrelation. This is characterised by the clustering of high values of a variable in a region while another region will have a clustering of low values. A much rarer form of spatial autocorrelation occurs when repulsion occurs among entities. The spatial pattern of negative spatial dependency is characterised by clustering of unlike values. A classic example of negative spatial autocorrelation is the spatial distribution of white and black squares over a chess board.

Connectivity

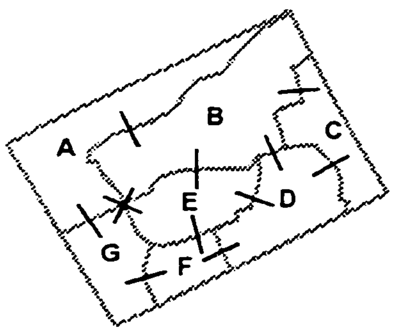
Spatial statistics differ from conventional statistical analysis in that the locations of observations are incorporated into the analysis. A spatial weights matrix must be defined which explicitly states the spatial relationships between the observations. Measures of connectivity of observations can be defined in several ways including distance between observations, lengths of common boundaries, as well as adjacency. The most common method in spatial statistics seems to be adjacency of observations but even here different methods exist for specifying spatial neighbours (Figure 6.7). A rook structure is based upon shared edges, while a bishop connectivity is based upon shared corners. A queen structure combines the rook and bishop methods. Choice of an appropriate spatial weights matrix specification is often done in a haphazard manner based primarily on convenience

Figure 6.7) Types of Polygon Connectivity



a b c

Forms of connectivity based upon adjacent polygons: a) rook; b) bishop; c) queen



Connectivity Information

Polygon	Number of Neighbours	Identity of Neighbours
A	3	B, E, G
B	5	A, C, D, E,G
C	2	B, D
D	4	B, C, E, F
E	5	A, B, D, F, G
F	3	D, E, F
F	4	A, B, E, F

Example of irregular polygon connectivity using a queen type adjacency

(Griffith, 1996, p.65). In addition connectivity can be extended beyond first order neighbours up to the *n*th neighbour. This is useful when creating correlograms that show trends in spatial dependency. The type of phenomena being investigated as well as the regularity of the tessellation will play an important part in what type of connectivity is appropriate.

Several types of connectivity structures were considered for this research. Both rook and queens structures were examined as well as a distance based connectivity using a radius of 1000 metres. Connectivity based upon shared boundary length did not seem appropriate since this would have weighted large rural polygons more heavily than central city polygons. Examination of spatial dependency in some variables showed little difference between weights specifications. A queen structure based upon first order neighbours was chosen for this research. Simple adjacency seems to work well for participation rates for which data exists in all polygons but a distance based connectivity might be more appropriate when looking at rare event data such as crime statistics where not every polygon might contain an observation. The queen structure was chosen because phenomena involving population can be thought of as being continuous and all points of contact between adjacent spatial units should be considered as neighbours. The spatial units used in this research are an irregular tessellation of polygons. Most neighbours share edges and there are few areas where the only point of contact between neighbours is a corner. For this reason an irregular spatial arrangement will not be sensitive to mispecification between a rook's or a queen's structure. This would not be the case in a regular tessellation as researched by Griffith where mispecification could considerably increase or decrease the number of neighbours. Griffith also suggests that low order spatial weights should be given preference over higher order and for this reason only first order neighbours were considered for purposes of connectivity (Griffith, 1996, p.80). Creation of first order rook and queen structures is easily done in the *ArcView* GIS while distance based connectivity established upon polygon centroids was calculated in *SpaceStat*. The ability of GIS to quickly calculate connectivity for selected subsets of the study area meant that spatial models could be fitted quite quickly to subregions of interest without the long process of having to manually determine the connectivity structure of that region.

Regression Models

Regression analysis is a mathematical method of evaluating the form and direction of the relationship between variables. A dependent variable is considered to be controlled by the variation in a number of independent variables. The bivariate regression is the most simple type of regression with only one independent variable. However, voter participation is a function of many factors and any model of this form of spatial behaviour will need to have a series of independent variables whose interactions will control participation. The basic multiple regression equation is as follows:

$$Y = a + b_1X_1 + b_2X_2 + ...b_nX_n \pm e \qquad (3)$$

where: Y is the dependant variable
$X_1...X_n$ are independent variables
n equals number of variables
$b_1...b_n$ are partial regression coefficients
a is a constant
e is an error term

Equation from Shaw and Wheeler (1985), p.230.

The independent variables must meet certain criteria in order for the regression model to be valid. The variables must be measured on the interval/ratio scale, approximate a normal multivariate distribution, and have observations that are independent of each other. Violations of these assumptions will result in unreliable parameter estimates which will compromise the results of the analysis. The distribution over space of the values may also have a role in violating the conditions for a valid regression model.

The most commonly used form of regression is the ordinary least squares (OLS) regression *(equation 3)* which has the criterion of least squares. In a bivariate OLS regression, this involves fitting a line to a scatter of points to understand the relationship between observations. The OLS method seeks to minimize the sum of the squares of the deviations of the points about the fitted line. The least squares method applies also to multiple regression but it cannot be visualized since it involves finding the best fit amongst the scatter of points in n-dimensional space. The fit of the model is revealed by the coefficient of multiple determination (R^2). This is determined by dividing the variance in the dependent variable explained by the independent with the total variation of the actual dependent values. Simply put, the R^2 is the proportion of variation explained by the model. Partial regression coefficients show how much a change in an independent variable will change the dependant variable when other independents are held constant. The partial regression coefficients are in the units of their independent variable and cannot be directly compared. These coefficients can be transformed into standardized beta values β which are comparable. Changes in variables with large β values will have more effect than changes in independent variables with small βs. Partial regression coefficients must also be evaluated using a t-test to determine whether their contributions to the model are statistically different from zero. This is done to examine the spread of the variance for each independent variable to see whether the spread about the regression fit is concentrated sufficiently to be considered statistically significant. A variable may very well add some explanatory power to the regression but have such a wide dispersion as to be statistically meaningless.

Choosing the independent variables for inclusion in a regression is no easy task. It is tempting to simply throw in as many explanatory variables as possible in order to get a high R^2 value and have a large percentage of the variance explained. This approach can be dangerous since one must be on the look out for spurious correlations. These occur when there is a observed relationship between two variables that does not have any real causal basis. For example, riding level examinations of voter behaviour often show religion to be an influence on political behaviour although elections are rarely about religious issues and surveys show individual political behaviour to be rarely influenced by religious considerations (Meisel, 1967; Gidengil, 1992). Religion has been associated with other aspects of social lifestyle and upbringing that do have an effect upon voting behaviour (Meisel, 1967). Aside from spurious correlation, the geographer must also be aware that large numbers of causal variables do not necessarily result in a stable model. Independent variables are often correlated among themselves and will add little explanatory power to a model while violating the assumptions of independence implicit in regression. This problem is referred to as multicollinearity. Presence of multicollinearity will not cause the regression to break down but will lead to poorly estimated parameters. Partial regression coefficients may be found to be statistically insignificant even though the R^2 may be high (Anselin, 1995, p.26-7). The model may also be sensitive with very different results caused by the dropping or addition of a single variable.

Most statistical packages support a method of variable inclusion called the stepwise method. This is an iterative process where successive variables are examined for their explanatory power. After each variable is added, all other variables are re-examined to test their statistical behaviour in the presence of the other variables. This process is repeated until the presence of more variables does not add a statistically significant increase in the variance explained. An analysis of variance F test is used to check if the increase in the R^2 is statistically significant. Stepwise methods help to lessen the effect of multicollinearity since highly correlated variables will not be included together. However, although a stepwise method of variable inclusion will facilitate the construction of statistically sound models, it may also create models that have no theoretical basis. It is important not to follow blindly statistical procedures with little understanding of the phenomenon of interest. Sometimes a model that has a lower explanatory power with few variables may be a better model that one that it statistically perfect with a multitude of hard to interpret and theoretically unsound variables.

Analysis of the residuals will yield much important information about how well specified the regression model is and will yield clues as to how to improve a mispecified model. The residuals are the difference between the predicted values and the actual dependent values. Most regression diagnostics make the assumption that errors in the dataset follow a normal distribution (Anselin, 1995, p.26-8). It is difficult to know how well this assumption is met since the errors are invisible and tests for non-normal error distribution must be performed on the residuals. It is common to standardize residuals by their deviations from the mean. These z-scores, assuming a normal distribution of residuals, will reveal the extreme values where the model has significantly over or under predicted the dependant value. These areas should be examined carefully for clues as to why these places are different. Knowledge of the study area is critical in the evaluation of these areas. Alternatively, measurement errors are also a possibility. These extreme values will have a large influence or leverage in a least squares regression and these points can be removed and the model run again (Shaw and Wheeler, 1985, p.245).

The residuals of a well specified regression equation will be homoskedastic, follow a normal distribution, and have no spatial dependency. Statistical tests exist to check for heteroskedasticity, normal distribution and spatial dependence but histograms and mapping of the residuals are also very handy in detecting mispecified regression models.

Heteroscedasticity is a condition where the error variance in the residuals is not constant over all the values. A model that has heteroskedastic residuals will have misleading diagnostic statistics and R^2 values. Furthermore, it is an indication that spatial effects may be present in the dataset which are not accounted for by the included explanatory variables. The presence of spatial dependency in the residuals is particularly important since it means that spatial effects will have to be explicitly incorporated into the model (see below). Addressing spatial effects is an important part of modelling that is rarely examined in the social sciences and sometimes even in the field of geography itself.

Spatial Dependency in Regression Analysis

Neighbourhood effects can result if the spatial units of the analysis are not appropriate for capturing the variance in the variables for an event. A regional trend examined at a fine resolution would show spatial dependency in the spatial units. The presence of spatial dependency is a serious concern to the spatial scientist because it indicates that the regression model is mispecified and that some of the assumptions for a valid regression model are not being met. Generally spatially autocorrelated residuals indicate that at least one of several problems is occurring: the scale of study may be inappropriate; the variables are not normally distributed and may require a transformation; or that at least one important explanatory variable is being missed. Solutions to these problems can vary. The spatial distribution of regression residuals can be examined for possible clues regarding missing variables. Additional information can be added to the regression in the form of dummy variables that signal the presence or absence of a characteristic e.g. inner city =1, not inner city = 0. Dummy variables, being nominal in nature, technically violate the assumptions of a regression equation although their use is advocated by many spatial scientists (Anselin, 1995, p.8-3). Also, spatially lagged variables can be used in the regression model. In this type of variable, the value for a spatial unit is determined by its neighbours. This is generally the mean values of the first order neighbours but the lag could be extended out to the *n*th order neighbouring spatial units. Spatially lagged variables are especially useful if the spatial characteristics of two phenomena seem to occur at differing scales. Lagging a variable will serve to smooth the distribution of a variable by suppressing the peaks and troughs of individual variation in favour of the more generalised regional trend (Haining, 1990, p.340). A variable representing a fine scale phenomenon could be lagged to fit another that was more regional in nature. Another solution to spatial autocorrelation would be to use a spatial autoregressive regression (SAR) in place of a OLS regression. A SAR model will apply a spatial coefficient for a variable in addition to the regular beta partial regression coefficient for the variable. In addition, regression models can be fitted with spatial error terms. Finally, the study area can be partitioned and a spatial regimes model can be applied using different regression parameters in each region of study. The presence of spatial autocorrelation greatly complicates modelling in the social sciences since it can be caused by a number of different factors. Solutions require much thought about the interactions with the database, as well as experimentation and ingenuity.

Neighbourhood effects are detected through the use of spatial autocorrelation statistics like Moran's I and the Geary C statistic. The Moran statistic seems to be the most commonly used by spatial scientists and is the statistic used in this research for the detection of spatial dependency. The calculation of Moran's I requires the specification of a connectivity matrix (see above) in addition to the numeric values of the variables or residuals in question. The Moran's I statistic is calculated as follows:

$$I = \frac{n\sum_i \sum_j \sigma_{ij}(x_i - x_{avg})(x_j - x_{avg})}{S_o \sum (x_i - x_{avg})^2} \tag{4}$$

where: n = number of spatial units
x = value of variable at spatial unit i or j
x_{avg} is the mean value of variable
S_o represents the number of connections between spatial units defined by $\Sigma_i \Sigma_j \sigma_{ij}$
σ_{ij} represent the spatial relationship between spatial units i and j in terms of adjacency; σ_{ij} will equal 1 if the spatial units are adjacent; otherwise σ_{ij} equals 0.

Equation from Chou (1997), p.275.

Equation (4) tests the spatial distribution of values against a normal distribution and yields a value ranging approximately from 1 to -1. Absence of spatial autocorrelation will yield a statistically insignificant I value that will be close to 0. Significant positive I values signify positive spatial autocorrelation while significant negative numbers reveal negative autocorrelation in the dataset. The actual value of Moran's I will vary considerably according to the number of observations and can also be influenced by the spatial structure of the dataset as demonstrated by Tiefelsdorf (1997). The probability of significance and sign of the I statistic rather than its value are the important considerations when checking for spatial autocorrelation.

Finally, the researcher must be aware of the limitations of regression modelling. Predictions using regression must be regarded with caution since the variables involved are often unique to the specific event examined. The results of a spatial model of participation in the 1993 election in Kitchener-Waterloo might not be applicable to other regions or even to other federal electoral events in Kitchener-Waterloo. In the course of this research, the spatial distribution of the participation rates of the 1992 constitutional referendum were examined and they appeared to be the result of different processes than the 1993 event. Spatial effects and socio-economic cleavages seemed to be much more evident in this event that occurred less than a year before the event studied. This shows that a spatial model of electoral participation is the product of a specific place and time; applying similar techniques to similar data elsewhere may yield very different results.

We have seen that spatial modelling is a highly complex process that is iterative in nature. A regression model is never simply run and then accepted as a final result. Regression diagnostics and residuals must be examined, mapped and potential problems identified, and solutions attempted. In social analysis, a spatial model will never achieve a perfect explanation of all of the observed variance in the phenomena examined since much of human behaviour is often random in nature. Electoral behaviour is particularly susceptible to this (Prescott, 1972, p.87). The best that a spatial scientist can hope for in these circumstances is to achieve a statistically sound model that makes theoretical sense. A spatial regression model is never truly finished since the researcher will probably always have ideas regarding new variables, new definitions of connectivity, or new statistical models to try to explain the elusive unaccounted variance. Some success may be possible but there is also the risk of wasting precious time and resources on creating new variables and models that are not really any better than the originals. Knowing where to stop is just as important as knowing where to start.

Chapter Seven

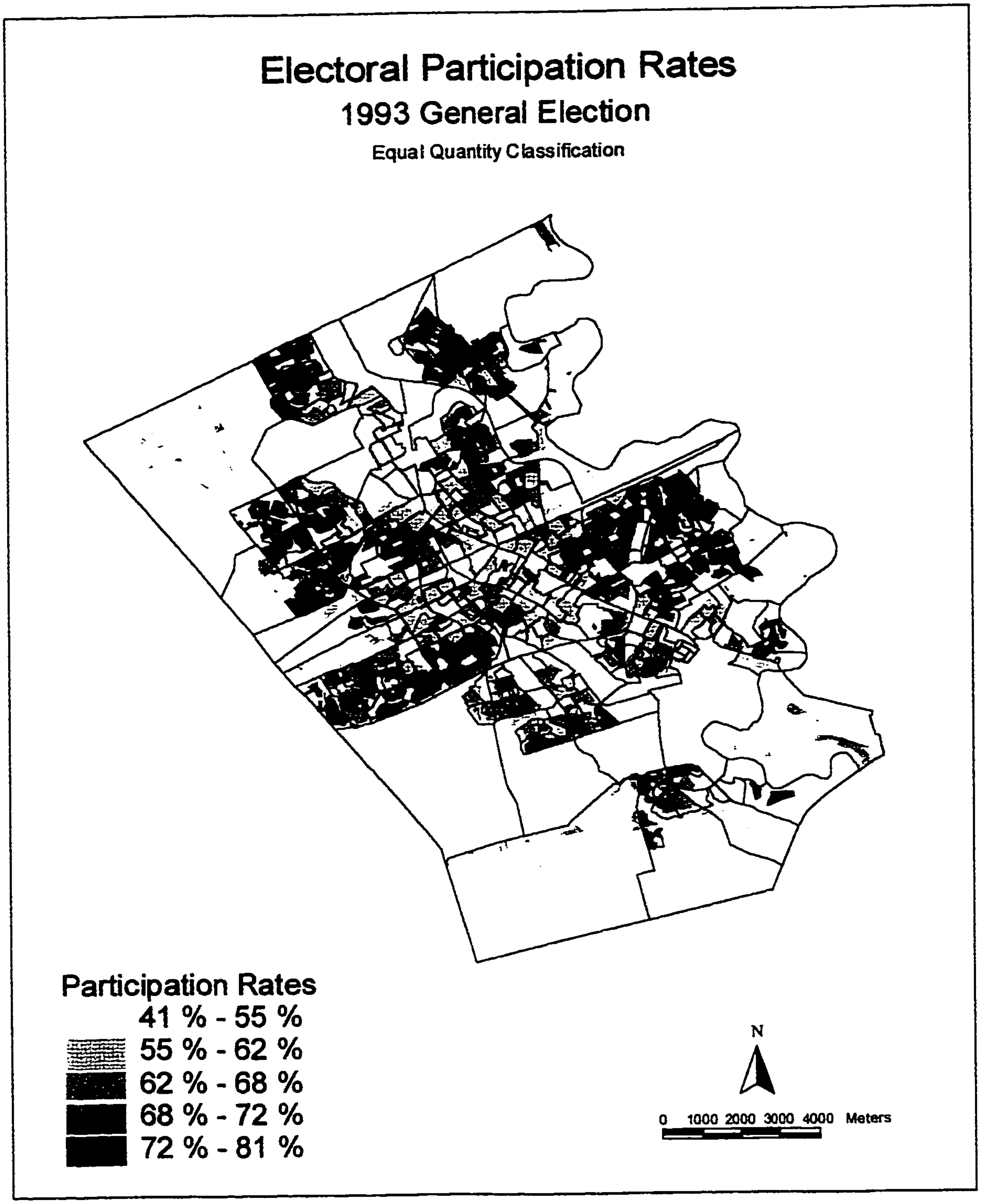
Exploratory Data Analysis

Spatial Patterns of Participation

An important part of spatial modeling involves exploring the database before starting more complex spatial analysis. This is useful since it allows a researcher to gain insights into the social and spatial composition of the region, and anticipate results as well as potential problems. The purging of outlier observations and identification of potential statistical problems regarding highly interrelated variables is also undertaken during exploratory data analysis. Many tools are available to undertake these tasks and include scatter plots, descriptive statistics, bivariate regressions, and spatial correlograms. Another important technique is the mapping of data to allow a quick visual appraisal of the situation which may reveal trends not apparent in the statistical figures.

The spatial distribution of turnout (Figure 7.1) throughout the study area needs to be examined before attempting to model the participation process. The overall pattern

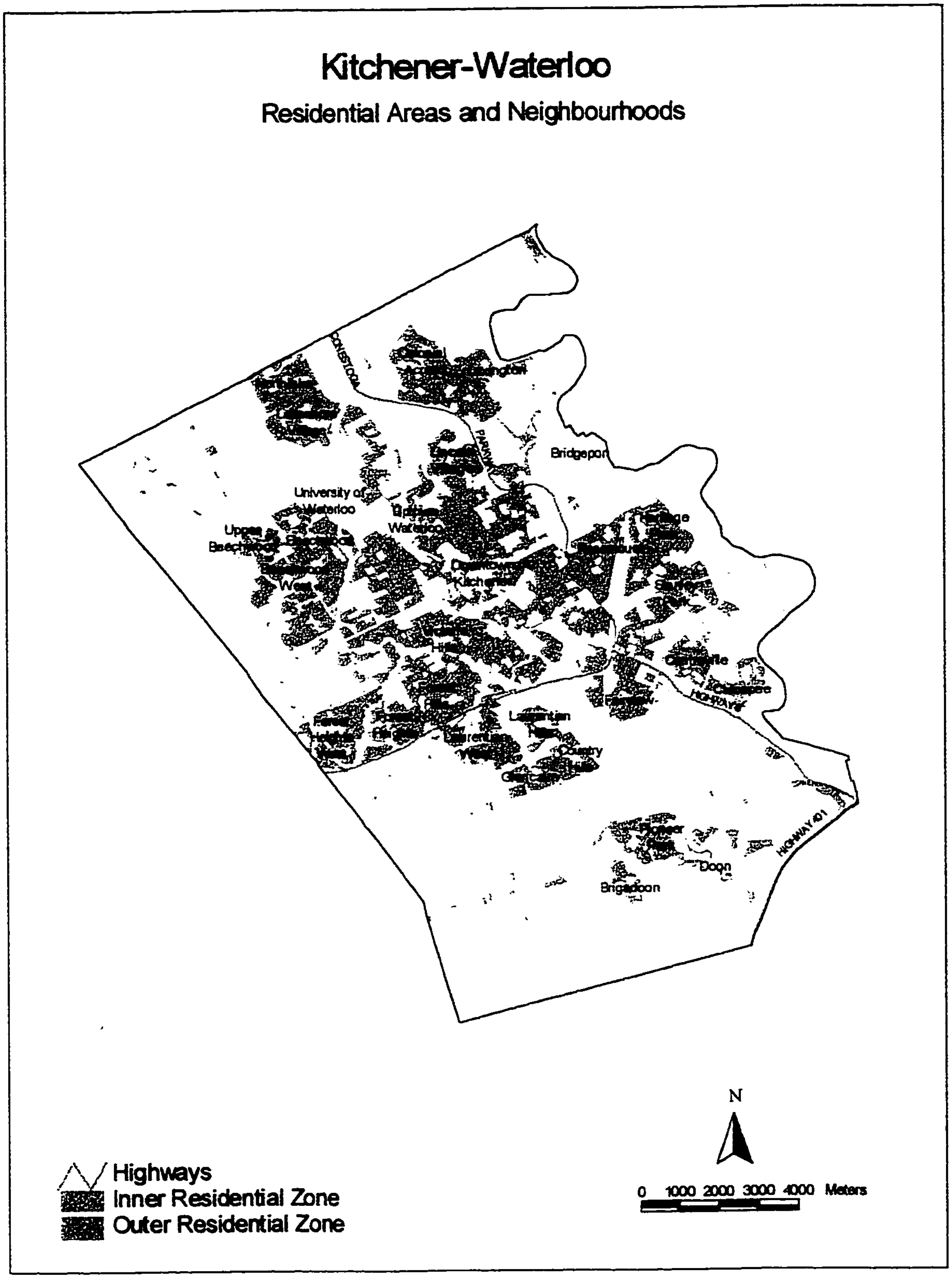
Figure 7.1) Participation Rates in Kitchener-Waterloo



seems to follow the inner city and suburban structure of Kitchener-Waterloo. Participation rates higher than the regional average of 63.4% are associated with the outer suburban regions. A reference map (Figure 7.2) shows the residential neighbourhood of the regions as well as some of the major roads. High rate neighbourhoods are Forest Heights, Lincoln, Heritage Park, Beechwood, Stanley Park, Westvale and Forest Heights West. Much lower participation rates occur in the inner zone of the city. This region is centered along King Street and runs north right from its entry into the urbanized part of the study area at Fairway up to its intersection with the Conestoga Parkway in the north of the study region. Uptown Waterloo, downtown Kitchener, Victoria Hills, Fairview, Chicopee, and the regions surrounding the universities all exhibit participation rates that are lower than the regional average.

Kitchener-Waterloo has several clusters of apartment buildings where population is highly concentrated. The population characteristics of these regions are often different from the surrounding areas, with higher population turnover (more people moving in and out) and less social cohesion. The regions around Westmount Plaza in the north and Fairview Mall have especially dense concentrations of apartment buildings but similar concentrations are also found in other parts of the study region. Participation rates in apartment blocks are generally quite low even when the polling station is located within the building. Only one apartment block located near the Dutch Boy supermarket on Margaret Street had a turnout rate higher than the regional average. Interestingly its sister building, with similar demographic characteristics, has a participation rate that is 10% lower. Apartment blocks generally seem to have a much lower participation rate even

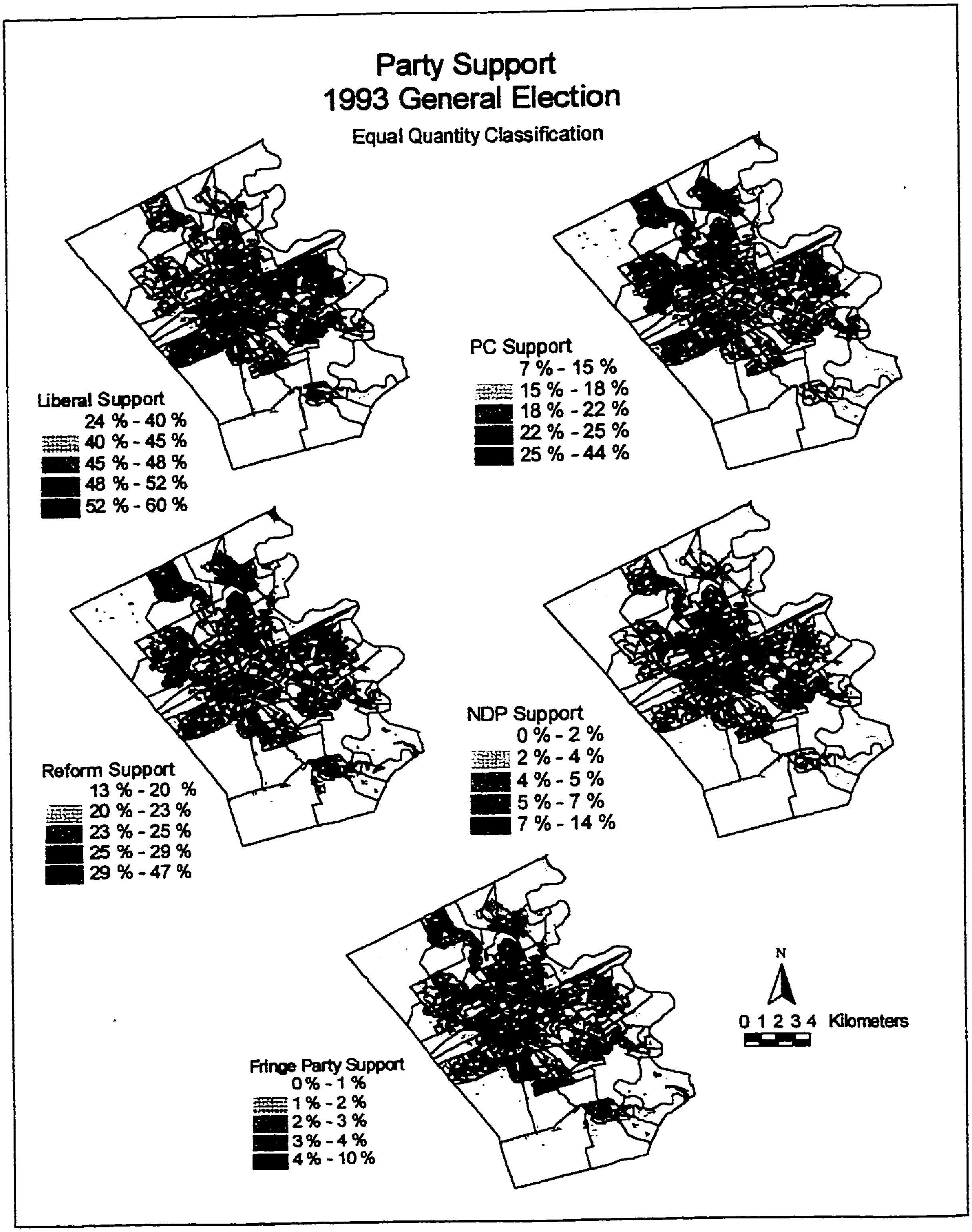
Figure 7.2) Kitchener-Waterloo Neighbourhoods



when surrounded by regions with higher rates. This can be seen in the Stanley Park area where the housing blocks near Stanley Park Mall are lower than the rest of the immediate area. For spatial modeling purposes however, most apartment buildings were purged from the analysis since the socio-economic data obtained from the area weighted interpolation was suspect. The effect of these denser areas of higher mobility and reduced social cohesion is not included in the models presented.

The difference in the inner city regions of Kitchener-Waterloo and the outer suburbs suggest that demographics and social characteristics are determinants of political participation since this mirrors the patterns shown in income, education, house ownership, and age distributions. It remains to be seen if the core/periphery pattern in participation is the result of separate processes in these zones requiring a spatial regimes model that examines each region separately or simply a result of the distribution of population characteristics which would be accounted for by one regional model. The spatial variation in turnout contrasts strongly with the patterns of support for political parties (Figure 7.3) which are not strongly divided according to a core and periphery relationship but rather a north/south divide that seems to correspond with the riding boundaries. However, support for the NDP and the fringe parties seems to echo the pattern displayed in the participation rates. The distribution of support for the large political parties is interesting in that it seems to support the theory that the brokerage politics practiced by Canadian political parties does not elicit strong support along socio-economic cleavages. The spatial patterns indicate that the decision to cast a ballot and the decision on how to mark that ballot appear to be motivated by different processes.

Figure 7.3) Party Support in the 1993 Event



Correlations and Descriptives

Several tables (Appendices A to E) provide descriptive statistics for the variables examined in this research. Providing maps for all of these variables would consume too much space and is redundant since a basic trend seems to appear regarding the spatial distribution of social and economic characteristics in the study area. The basic inner city/outer suburban pattern can be seen in all of the variables indicating that these areas are quite different in nature. Appendix A lists the descriptive statistics for the complete study area as well as for the inner and outer zones. The pattern in the maps supported by the descriptive statistics indicates that the population of the outer region of the Kitchener-Waterloo is generally older and more likely to have a higher income and a better educational achievement than the population of the inner city as well as being more likely to be married and owning their accommodation. The inner city population is less stable and less cohesive since it consists largely of a younger population that is in flux. The youth population will be less likely to have stable employment and accommodation, and generally fewer opportunities to influence events that affect their lives, as an older population. Movement is common among the young as seen by the positive correlations between mobility and the younger population (Appendix B). Community ties are not likely to be strong with this form of population and political awareness may not be that high. Furthermore, we see that there are also substantial geo-structural differences between these areas. The outer suburbs are more spread out since distance to polls and size of residential areas are larger than inner city zones. The inner zone also has a greater population density. Any model of political participation must, in order to be valid, be able to describe both areas. This may not be possible if the processes and concerns that motivate the population to vote or to abstain from voting are quite different in both regions. The strong differences in the two regions suggest that some form of spatial regimes model may have to be used if an overall model for the region cannot be developed.

The correlation matrix provided (Appendix B) shows the interrelations between different variables. It is important to know how they vary with each other since this will help deal with multicollinearity problems. Inclusion of highly interrelated variables in multivariate statistical analysis can result in an unstable model with insignificant variables, and non-normal errors. Some relationships are obvious such as the strong correlation of eccentricity to residential area since area is a component in the calculation of the compactness of the polling division. Others such as single, youth population, and rented housing are all related since they are all associated with a younger life cycle where individuals are not firmly established in life. Especially strong correlations (those that are stronger than ±0.600) are marked in boldface. The strength but not direction of the correlation is important in knowing how variables are interrelated for the purposes of multicollinearity. In large databases even low correlations can be statistically significant. During the multivariate modeling, problems often seem to arise if variables that are correlated at more than ±0.600 are included and this figure is used to indicate strong relationships. The correlation figures presented here are derived from the entire study region. The correlation between variables in the inner or outer zones can change depending on the spatial distribution of the indicator. This is dealt with during the application of models to the two zones.

Scatter plots and the purging of high leverage observations.

The variables were plotted against participation in order to examine the strength and direction of the relationships. These relationships can be obtained also from the correlation matrix but scatter plots (Appendix C) allow a visual inspection of the dispersion of observations. Those observations that are quite different will stand alone from the general pattern. Only scatterplots of variables that proved to be important in the course of the research are provided in the appendix for space considerations. In some cases a transformation was deemed necessary to straighten the relationship between the variables. This is necessary if it appears that the relationship is not linear but rather curvilinear. This appears to be the case with residential area and household income where points do follow some sort of linear relationship but are more heavily concentrated in one part of the graph. These scatter plots serve to show how the variables behave on their own but it should be remembered that multivariate behaviour is much more complex and variables that appear to have no direct relationship may indeed prove to be important when combined with other variables. This appears to be the case regarding the Majority variable that measures the strength of victory in a polling division. On its own it appears to be randomly scattered with no relationship to participation, but, in some multivariate tests performed on the data set Majority often proved to be a significant determinant of participation once the effects of other important variables were controlled for. It seems that most variables are inversely related to participation. Especially strong relationships seem to be with rental and population movement variables, unmarried, and single parent family indicators. Younger ages groups are inversely related to participation while the older groups above age 40 are positively associated. Income and residential area are also positively associated with high turnout. Political variables do not seem to have as strong relationships as other indicators examined. The PC and Reform vote seems to be positively associated while the NDP and fringe parties are inversely related. The Liberal vote seems to be quite dispersed with no strong relationship present.

The scatter plots presented show that most of the variables have a linear relationship to participation although it appears that in a lot of cases it is not a strong one. Several variables (residential area, density, and household income) were transformed to achieve a more linear relationship by application of a logarithmic transformation. This transformation accounts for the non-linear relationship that seems to be present in these variables where the increase in participation drops off after these variables reach a certain threshold. Overall, little was done to transform variables since mathematically transforming variables and expressing them in standardized units can greatly complicate interpretation and understanding. Standardization of variables (in deviations from the mean) can aid in minimizing the statistical impact of numerically large variables (e.g. distance expressed in meters) when used with other rated variables. In some cases variables with large units will dominate a model. In order to evaluate this effect a copy of the database was made and all variables transformed by means of power transformations to yield a completely normalized database where all variables followed a Gaussian distribution and were expressed in standard deviates. Results obtained from the original database were evaluated on the standardized data and very little differences were noticed. Models showed the same variables as being significant and they seemed to behave in the same manner. The standardized data was eventually discarded as it was redundant. We can be confident that the models presented in this research are a true result of the spatial distributions and interactions of the observations themselves rather than the measurements of the units.

A further use of scatter plots involves spotting observations that appear to be very different from the rest of the database. These outliers are points that are located away from the main pattern of point dispersion. In a least squares type regression model these outliers will have a disproportionate effect on the parameter calculation for a best fit of a model to the database. The stability and ruggedness of a statistical model can be evaluated by removing the high leverage points and rerunning it. If the results are widely different then it can be safely assumed that the observations have tremendously disproportionate effects. This does not necessarily mean that the observations are "wrong" but it does mean that they would appear to be very different from the rest of the observations. The most appropriate strategy in these cases is to remove them from the general model and examine them separately in order to try and determine why they are so different. Spatial statistics can also be used to check a data set for extreme values by means of a Local Indicator of Spatial Autocorrelation (LISA) statistic. This was employed in residual analysis but basic exploratory data analysis was restricted to the scatter plots.

This study is not concerned with identifying every single outlier in each scatterplot since there is always an element of randomness in human geography that will lead to observations behaving differently. Of interest are observations that are consistent outliers throughout the database. The observation numbers for the outliers in each scatterplot are shown and several seem to show up consistently as being quite different from the rest of the database. Seven of these points (observations 42, 184, 197, 205, 239, 309, 333) were detected with number 239 being the most obvious to spot. This particular observation proved to be the "Waterloo County Home for the Aged" that had quite different age and social characteristics from the rest of the area. The turnout was surprisingly low in this polling division (26%) despite the presence of a mobile poll to ease access of the electors. Five of the other outliers were densely populated polling divisions where an error in assessing the extent of the residential area would have major consequences. These observations were located in apartment blocks in central Kitchener and Fairway. One observation was a polling division composed of mostly uninhabited areas such as large school yards and conservation areas. These observations should have been deleted with the other old age homes, hospitals, and apartment blocks where the journey to the poll was effectively non-existent. The remaining outlier consists of an apparently normal suburban area near Fischer-Hallman that should have been well handled during the interpolation. The researcher is not sure why this particular polling division seems to be so different from the rest of the database. Overall, it is surprising that there were so few outliers in such a large database.

Removal of the outliers did prove to be justified since the results from the basic stepwise regression were influenced by their presence. Exclusion of the outliers increased the R^2 goodness of fit measure by 2%. The distribution of residuals was also non-normal and the histogram of residuals seemed leptokurtic in shape. Exclusion of these seven cases results in a less extreme range of residuals and the appearance of the classic Gaussian distribution curve. This indicates that these seven polling divisions are quite different and had a high influence on the fit of the model. Their exclusion makes for a better understanding of the processes occurring in the other 377 polling divisions.

Correlograms

Spatial insights into the database are needed to get an indication of the scale at which variation occurs. This is necessary for evaluating the appropriateness of the spatial units used. A poorly specified model may be the result of incorporating variables that seem to operate at different scales. The Moran statistic discussed above provides a measure of the clustering of like values. This is usually calculated for only first order connectivity (or lags) but can be extended to examine spatial association at higher order lags. The Moran values can be bar charted as correlograms that indicate if the spatial variation is highly localized or if it occurs at a more regional scale. Persistently high Moran values that carry on over a number of lags would indicate that the variable is highly regional and the presence of heterogeneity in the variance (Bailey, 1994, p.272) which could cause problems in statistical analysis. A rapid drop off after the first lag would indicate a much more homogenous spatial distribution with any clustering occurring at a highly localized level. The correlograms give the analyst some indication of the appropriateness of the spatial units as well as clues as to which variables may be candidates for spatial lagging in order to counter spatial dependence. The correlograms presented (Appendix D) all have the same scale so as to facilitate comparisons. It should be noted that the significant values of Moran's I will vary according the size of the sample and lag order, and similar values cannot always be compared between data sets. In this case comparison is possible since the number of observations is the same for all variables. The correlogram serves as a useful exploratory tool for examining the spatial structure of the variables. The values of Moran's I for neighboring lags are highly correlated since the correlation at higher lags is a product of the lower lags (Bailey, 1994, p.271). It would be very unusual to see a high value at one lag followed by a quite low one which is in turn followed by a return to a high value.

A glance at the correlograms (Appendix D) reveals that most of the variables are fairly localized in nature although some political as well as educational variables clearly show evidence of interacting at a more regional scale. The differences in the income and educational characters of Kitchener and Waterloo are probably responsible for the spatial interaction exhibited in the correlograms for the educational and income variables. The regional trends exhibited in average household size and unmarried population can be attributed to the different social and demographic composition of the inner city and outer suburbs. Of particular interest is the difference in spatial dependence in the political variables. The PC and Liberal parties clearly have a more regional character than do the NDP and the Reform Party. Support for the fringe parties shows little persistence and is quite localized. Plurality, indicative of the strength of victory, shows regional variation which is similar to the support for the two largest traditional parties but the correlograms depicting participation show little persistence over space and is not at all regional in nature. It appears to be quite different from the other political variables and is more similar spatially with social variables depicting the residential stability of the area such as age, home ownership, and residential mobility. The correlograms for these variables indicate that they are more fragmented and occur at a finer scale with pockets of similarity but with no regional trends. These correlograms indicate that political participation in this event is more related to social phenomenon rather than political motivation. Furthermore, the polling division is an appropriate unit for modeling participation since the spatial interaction does not appear to persist beyond the neighboring polling divisions. Also, lagged variables will probably contribute little to the modeling of participation although an examination of support for the political parties would probably require lagged explanatory variables to "regionalize" the social, economic, and geographic variables so as to be comparable to the regional support for these parties. It is evident that spatial statistics can contribute to the modeling and understanding of political support as well as voter participation.

Bivariate Regression Analysis

A useful exploratory tool consists of running a series of bivariate regressions for all of the variables with participation as a dependant variable. This is to show which variables are good predictors of participation. It is also possible that a basic bivariate model may provide a good model of turnout without the need to resort to more complex multivariate techniques. This would appear to be unlikely since human spatial behaviour is usually the result of many complex interactions and is very rarely motivated by any one factor. The results of the bivariate regressions must be viewed with caution since variables can behave differently in a multivariate environment. The rented variable is the best single predictor of turnout. The unmarried population, lone parent families, population age variables, income, and residential area variables all offer R^2 fits that are fairly high compared to the other variables in the data set. The fit of the logarithms of residential area, income, and density are all higher than the untransformed variables indicating that changes in these variables do not have a constant affect upon participation but tend to drop off once a critical value is reached. No political variables really seem to offer much explanatory power in themselves with the possible exception of the NDP at an R^2 of 0.236. It is interesting that there is no relationship with participation with the Liberal vote despite their very strong showing in 1993 in general and in Kitchener-Waterloo in particular. Educational and employment type variables are also poorly related to political turnout indicating that participation is not based upon a class based cleavage.

Clearly participation is a result of a complex process since none of the variables, with the exception of rented, seems to provide good explanatory power. Furthermore,. all of the univariate models display significant degrees of spatial autocorrelation in their residuals indicating mispecified models and poor fit in many parts of the study region. Multicollinearity is not a problem since only one explanatory variable is used. No one variable is able to reliably explain variance in participation on its own and a multivariate model is clearly needed.

The exploratory research using scatterplots, correlograms, correlation matrices, and bivariate regressions allows some insights to be gained into what determines participation. The social variables seem to be related more strongly to turnout than the political variables and the maps also reveal different spatial patterns. The single best indicator of turnout is house ownership expressed in terms of the percentage of population renting accommodation. This is an inverse relationship as is unmarried population, youth population, and percentage of single parent families which are also related to participation. Others that are related in a positive way are household income, average household size, and older population. These variables can be thought of life cycle variables since they are indicative of a stage in life. Others, like education and profession are more likely to be indicators of social class. The correlation matrix shows that these life cycle variables are all interrelated and it is highly probable that only a few will prove to be significant since some of them (e.g. rented and population age) could each substitute for the other. Exploratory data analysis also serves to identify high leverage points that can disproportionately affect the performance of the model. Examination of the database also helps determine how the analysis will proceed. The different characteristics of the inner and outer zone mean that care will have to be taken to ensure that the model predicts uniformly throughout the complete study area. Poor fit of an overall model would suggest that fundamentally different processes and relationships are at work in the different zones and each would have to modeled separately. The strong differences in descriptives for the two zones suggest that this may be the case. Voter participation is the result of complex interactions between many factors and a multivariate model with spatial diagnostics is required since no single variable is able to adequately predict this behaviour.

Chapter Eight

Multivariate Analysis

Spatial statistical analysis is highly data intensive and there is a need to coordinate many different variables and databases. Many models and combinations of variables must be examined to try and obtain a well specified model that supplies high explanatory power. Furthermore, models must also be applied to regional subsets to evaluate how the model performs spatially. GIS proved invaluable in matching spatial databases and in creating geographic variables. It can also serve as a database manager and proved very useful in exchanging information between statistical software packages and in visualizing results.

Three different software packages were needed to develop and evaluate the statistical models. *SPSS* has advanced modeling capabilities and was especially useful since it supports stepwise regressions as well as residual histograms and scatterplots. It lacks provisions for spatial statistics and *SpaceStat* had to be used to evaluate models for global and local spatial autocorrelation. *ArcView GIS* was used as a centralized data manager and acted as a link between the two statistical packages. The extensive mapping capabilities of *ArcView* were used to present the final results. A utility obtained via the Internet from the *ArcView* community loosely coupled *ArcView* with *SpaceStat* and allowed easy visualization of statistical results. This adds considerably to the capabilities of *SpaceStat* which has no visualization procedures and displays its results in long streams of often difficult to interpret numbers. Ideally, one software package would have performed all of the functions listed above but, in the absence of this, the loose connection of three highly different programs worked fairly well.

A basic model can be derived using statistical software that implements the stepwise regression procedure discussed earlier. The number of observations is actually quite high in the data set and it is quite easy to get statistically significant relationships between variables even when the product moment correlation values appear to be quite low. Multicollinearity was also a danger since some of the variables proved to be quite highly interrelated but use of stepwise procedures helped to minimize this problem. Stepwise regression tests whether the increase in the R^2 value which represents the explanatory power of the model is significantly increased by addition of the variable to the model. Generally, the iterative stepwise process will find the best fit and if two variables are highly interrelated the one that fits best will be selected. The other will probably not add much explanatory power to the model and will be rejected. A stepwise regression will give the best fit but care must be taken to verify that spurious correlations have not occurred and that the model is theoretically sound. Building a multivariate model using stepwise procedures is an iterative process where many different combinations are tested and retested using differently specified variables, substitutions, etc. with the aim of achieving a well specified model with good explanatory power using variables that make sense and work according to theory. The first attempt to model participation involved inclusion of all variables to establish a base model from which to work. From this a final model was developed and applied to the different zones in the study area to examine the fit. Table 8.1 lists the parameters and diagnostics of the different models while Table 8.2 lists the relative impact of variables upon the R^2 measure of explained variance. The order of inclusion of variables in the regression is always top to bottom.

The Models

The basic model derived from the stepwise regression (Model X1) seems to have several components: a life cycle component represented by rented accommodation, population aged 40-64, and population aged 18-24; a political motivation factor represented by support for the PC party; and a structural-geographic component represented by residential area and distance to polling station. After these come a series of variables that proved to be fairly unstable and highly sensitive to the other variables included in the equation. On occasions substitution of the PC support for another political party would completely change the variables that came into play following distance to polls. Support for fringe parties was one that seemed to be more consistent than others such as unmarried population, proximity to roads, density, or public service employment which all showed considerable sensitivity to the other variables. Model X1 shows the parameters for the basic stepwise regression where the *SPSS* package chose variables that fit best. Some variables (residential area, household income, and density) were transformed using their logarithm to simulate a non-linear relationship but, in a multivariate environment, these did not provide better fits than the untransformed versions.

Model X1 as a first attempt provides a fairly good model of participation. The explained variance in turnout is quite high with an R^2 of 0.692. The residuals are normally distributed and will be discussed in greater detail later on. All of the variables provide statistically significant contributions to the R^2 while the diagnostics for normality of errors, heteroskedasticity, robustness, and spatial dependance are all insignificant. Table 8.1 lists the significance levels of the variables and the tests by using the probability that the results of the test occurred randomly. A value of less than p=0.05 means that there is less than a 5% chance of the result occurring randomly. Impact of variables upon a model should have low p values while the diagnostics should not differ from a random distribution and should have high p values greater than p=0.05. Also, a multicollinearity condition number is provided that indicates the degree of multicollinearity present in the model. The precise value at which multicollinearity becomes a problem is subject to interpretation and for this analysis the value of 30 recommended by Anselin (Anselin, 1995, p.26-8) is used and Model X1 does fail the multicollinearity test. This is not surprising since four of the five political variables for party support are included. All four parties in Model X1 are statistically significant which would seem to indicate that they

TABLE 8.1) Regression Model Specifications

Model Fit

	Model X1	Model X2	Model X3	Model X4	Model X5	Model X6	Model X7
R^2	0.692	0.671	0.641	0.621	0.610	0.664	0.629
F-Test	32.0418 p = 0.000	107.458 p = 0.000	109.930 p= 0.000	152.276 p= 0.000	69.1364 p= 0.000	32.2384 p= 0.000	33.5046 p= 0.000
Number of Observations	377	377	377	377	272	105	105

Variables Included

Variable	Model X1 b Coef.	Model X1 Stand. β Coef.	Model X2 b Coef.	Model X2 Stand. β Coef.	Model X3 b Coef.	Model X3 Stand. β Coef.	Model X4 b Coef.	Model X4 Stand. β Coef.	Model X5 b Coef.	Model X5 Stand. β Coef.	Model X6 b Coef.	Model X6 Stand. β Coef.	Model X7 b Coef.	Model X7 Stand. β Coef.
CONSTANT	72.0634	-	62.7838	-	66.2409	-	62.3935	-	65.6617	-	67.7711	-	58.6564	-
	(3.002)	p= 0.000	(2.164)	p= 0.000	(2.172)	p= 0.000	(2.027)	p= 0.000	(2.478)	p= 0.000	(5.29201)	p= 0.000	(4.665)	p= 0.000
RENTED	-0.0841	β -0.229	-0.1059	β -0.288	-0.1200	β -0.352	-0.1200	β -0.327	-0.1390	β -0.373	-0.1093	β -0.264	-	-
	(0.018)	p= 0.000	(0.016)	p= 0.000	(0.017)	p= 0.000	(0.017)	p= 0.000	(0.020)	p= 0.000	(0.034)	p= 0.000	-	-
POP40_64	0.3057	β 0.211	0.341147	β 0.235	0.3522	β 0.243	0.3935	β 0.271	0.3521	β 0.256	0.3762	β 0.218	0.5794	β 0.333
	(0.052)	p= 0.000	(0.016)	p= 0.000	(0.055)	p= 0.000	(0.055)	p= 0.000	(0.064)	p= 0.000	(0.141)	p= 0.000	(0.131)	p= 0.000
POP18_24	-0.7626	β -0.252	-0.6597	β -0.280	-0.9095	β -0.301	-0.9778	β -0.323	-0.8688	β -0.276	-0.9194	β -0.365	-1.155	β -0.458
	(0.113)	p= 0.000	(0.121)	p= 0.000	(0.119)	p= 0.000	(0.121)	p= 0.000	(0.157)	p= 0.000	(0.201)	p= 0.000	(0.195)	p= 0.000
PC	0.1230	β 0.082	0.2321	β 0.154	0.2893	β 0.193	0.3221	β 0.214	0.280263	β 0.193	0.2837	β 0.187	0.3238	β 0.213
	(0.054)	p= 0.000	(0.050)	p= 0.000	(0.051)	p= 0.000	(0.120)	p= 0.000	(0.061)	p= 0.000	(2.854)	p= 0.000	(0.103)	p= 0.002
CLIPAREA	0.000030	β 0.222	0.000029	β 0.216	-	-	-	-	-	-	-	-	-	-
	(5.0E-6)	p= 0.000	(5.0E-6)	p= 0.000	-	-	-	-	-	-	-	-	-	-
DISTANCE	-0.00312	β -0.160	-0.0034	β -0.175	-0.00196	β -0.100	-	-	-0.00164	β -0.095	-0.00651	β -0.185	-0.00509	β -0.144
	(6.4E-4)	p= 0.000	(6.4E-4)	p= 0.000	(6.2E-4)	p= 0.000	-	-	(6.7E-4)	p= 0.015	(0.002)	p= 0.001	(0.002)	p= 0.025
NDP	-0.4330	β -0.128	-	-	-	-	-	-	-	-	-	-	-	-
	(0.123)	p= 0.000	-	-	-	-	-	-	-	-	-	-	-	-
REFORM	-0.175752	β -0.096	-	-	-	-	-	-	-	-	-	-	-	-
	(0.063)	p= 0.000	-	-	-	-	-	-	-	-	-	-	-	-
OTHER	-0.427609	β -0.086	-0.5038	β -0.102	-0.4937	β -0.100	-	-	-0.3355	β -0.068	0.8358	β -0.183	0.7599	β -0.166
	(0.160)	p= 0.000	(0.162)	p= 0.000	(0.169)	p= 0.000	-	-	(0.210)	p= 0.111	(0.280)	p= 0.003	(0.291)	p= 0.010
PARENT	-0.580134	β -0.109	-	-	-	-	-	-	-	-	-	-	-	-
	(0.220)	p= 0.000	-	-	-	-	-	-	-	-	-	-	-	-

Regression Diagnostics

Variable	Model X1	Model X2	Model X3	Model X4	Model X5	Model X6	Model X7
Multicollinearity Condition Number	38.3608	23.4047	21.4421	17.4792	19.9176	32.3608	19.9176
Test on Normality of Errors (Kiefer-Salmon Test)	KS value 1.8915, df 2 p = 0.388	KS value 3.2905, df 2 p= 0.192	KS value 0.9360, df 2 p= 0.626	KS value 1.9638, df 2 p= 0.375	KS value 1.8941, df 2 p= 0.322	KS value 2.1973, df 2 sig 0.333	KS value 2.2641, df 2 p= 0.322
Test for Heteroskedasticity (Breusch-Pagan Test)	BP value 7.6312, df 10 p= 0.665	BP value 14.2116, df 7 p= 0.047	BP value 9.5125, df 6 p= 0.147	BP value 4.8776, df 4 p= 0.300	BP value 10.7568, df 6 p= 0.096	BP value 2.5504, df 6 p= 0.863	BP value 0.94241, df 5 p= 0.967
Specifications for Robustness (White's Test)	W value 77.6728, df 65 p= 0.134	W value 68.4183, df 35 p= 0.000	w value 35.8122, df 27 p= 0.119	W value 15.9536, df 14 p= 0.316	W value 34.2075, df 27 p= 0.160	W value 18.45304, df 27 p= 0.889	W value 17.4601, df 20 p= 0.623
Diagnostics for Spatial Dependence (Moran's I Test)	I value -0.0519, z value -1.333 p= 0.183	I value -0.0330, z value -0.769 p= 0.442	I value -0.0418, z value -1.068 p= 0.286	I value -0.0347, z value -0.877 p= 0.380	I value -0.0073, z value -0.143 p= 0.886	I value -0.16114, z value -2.140 p= 0.032	I value -0.11431, z value -1.428 p= 0.153

Model X1: Model resulting from stepwise regression technique.
Model X2: Model with collinear political variables removed .
Model X3: Final model with only one geographic variable. This model provides the best balance between fit and statistical robustness.
Model X4: Model with only lifecycle and political motivation variables included.
Model X5: Final model applied to outer zone only.
Model X6: Final model applied to inner zone only
Model X7: Final model applied to inner zone only with rent variable removed

Table 8.2) Changes in R Square Values

Variable	Model X1		Model X2		Model X3		Model X4		Model X5		Model X6		Model X7	
	Change in R^2	Cumm. R^2	Change in R^2	Cumm. R^2	Change in R^2	Cumm. R^2	Change in R^2	Cumm. R^2	Change in R^2	Cumm. R^2	Change in R^2	Cumm. R^2	Change in R^2	Cumm. R^2
RENTED	.479	.479	.479	.479	.479	.479	.479	.479	.469	.469	.358	.358	-	-
POP40_64	.050	.528	.050	.528	.050	.528	.050	.528	.046	.515	.101	.459	.392	.392
POP18_24	.052	.580	.052	.580	.052	.580	.052	.580	.043	.558	.079	.538	.123	.515
PC	.041	.621	.041	.621	.041	.621	.041	.621	.038	.596	.059	.597	.064	.579
CLIPAREA	.014	.635	.014	.635	-	-	-	-	-	-	-	-	-	-
DISTANCE	.028	.662	.028	.662	.012	.632	-	-	.010	.606	.036	.633	.024	.603
NDP	.011	.673	-	-	-	-	-	-	-	-	-	-	-	-
REFORM	.006	.679	-	-	-	-	-	-	-	-	-	-	-	-
OTHER	.006	.685	.009	.671	.008	.641	-	-	.004	.610	.031	.664	.026	.629
PARENT	.006	.691	-	-	-	-	-	-	-	-	-	-	-	-

Model X1: Model resulting from stepwise regression technique.
Model X2: Model with collinear political variables removed .
Model X3: Final model with only one geographic variable. This model provides the best balance between fit and statistical robustness.
Model X4: Model with only lifecycle and political motivation variables included.
Model X5: Final model applied to outer zone only.
Model X6: Final model applied to inner zone only.
Model X7: Final model applied to inner zone only with a collinear variable removed to counter significant spatial autocorrelation

each cater to a different part of the electorate. However, since all of these political parties are going to be trying to get people motivated and involved in order to cast a ballot, including multiple parties in the model is in effect covering the political motivation factor several times. Dropping several of these variables should solve this problem. The single parent variable is also highly related with rented and contributed little on its own, although it did show as being significant.

Lack of spatial autocorrelation is significant in that it indicates that the model does a fairly good job in predicting participation throughout the study area and not in just one particular zone. This seems to indicate that although the average values of participation are much lower than those in the outer suburbs, they are still the result of the same basic process and that the same factors are at work in both zones. An overall model should work fairly well rather than some sort of zonal model. However, the overall model must be applied to each zone to in order verify this belief.

Removing NDP, Reform, and Parent from Model X1 results in Model X2. Some political models believe that fringe support is related to participation (Eagles, 1991, p.14) so this is left in the model but support for the other mainstream parties is removed. This model has fewer problems with multicollinearity but suffers from some problems of mispecification regarding robustness and heteroskedasticity. Some experiments with different variables showed that using two geographic variables caused heteroskedasticity in the data set. Inclusion of the residential area seems to conflict with the rented variable. This is probably because both variables are highly associated with each other in a geographic sense. The inner city tends to have a concentration of rented dwellings while the larger residential areas are located in the outer suburbs. Inclusion of all of the party variables may have helped suppress this effect but with only one political factor, the conflict becomes apparent. Use of a logarithmic transform on the residential area did not ease this problem either. This was attempted since it was theorized that the area represented some sort of critical density after which an increase in the size of the populated area contributed little to an increase in participation. Removal of one of the geographic variables eased the problem. Clearly the geographic factor seems to work with the life cycle variables to produce non-normal distribution of variance in the predicted model. Use of the distance to polling station variable (Model X3) instead of residential area (model not included) resulted in an R^2 value that is 0.005 smaller but is better specified in normality of errors and heteroskedasticity. For simplicity, only the distance model X3 is shown in the table. However, the results from inclusion of area instead of distance were highly similar and a case could have been made for use of area instead of distance in lieu of the slightly higher R^2 value. Instead the researcher opted for the model that explained slightly less dependant variable variance but was better specified. The final model (X3) used in this research for modeling political participation includes only distance to polls to represent the geo-structural component of turnout. Model X3 has an R^2 of 0.641 with significant variables and is well specified with all regression diagnostics for mispecification showing as insignificant.

In a fine scale examination, we see that inclusion of a geo-structural component does indeed contribute to the modeling of participation. However this contribution is quite small as can be seen by Model X4 where an R^2 of 0.621 is achieved when only life cycle variables and the PC support for political motivation are included. This model is well specified and does not have any problems with multicollinearity or even spatial autocorrelation. Inclusion of a geographical variable as well as the fringe vote adds only 0.002 increase (2%) to the R^2 value. From Table 8.2 it can be seen that the life cycle variables yield a cumulative R^2 of 0.580 before any other factors are included. Turnout is clearly related to what stage in life one is in and factors such as politics and the distance to the polling station, while contributing slightly to the decision to vote, are not really that important.

Model Application to Zones

The lack of spatial autocorrelation seems to indicate that the city can be examined as a whole rather than in zones but the final Model X3 should be applied to both zones separately to confirm this. Model X5 and X6 are the results of applying the final model to the outer suburban zone and the inner city zone respectively. Generally the overall model works fairly well in both zones. The different structural composition of the two zones has an effect on turnout as seen by the change in relative importance of the distance to poll measure. The standardized beta coefficient (β) is the measure of the relative importance of a change in one variable upon the dependant when the other independents are held constant. Distance is more of a factor in turnout in the inner city ($\beta = -0.185$) than in the suburbs ($\beta = -0.095$). The urban inner city is much more compact and since the average distance to the polling station is less and people are probably more inclined to walk than drive. The greater mobility of a car would tend to negate the effects of a longer distance to poll. Furthermore, voting could easily be done when driving to or from work and might not involve an extra trip. The population of the inner zone is more likely to be younger, and have less education and income. No data for car ownership was found for the enumeration area data but it is probable that the inner city population is less likely to have access to a car. The distance to the polling station is more likely to play a part in the decision to vote in the inner zone than in the suburbs. Another characteristic of the inner zone seems to be the greater importance of the youth population ($\beta = -.365$ for the inner zone, $\beta = -.276$ in the outer zone). The presence of a large youth population in the inner city (and their absence at the polls) is reflected in the β coefficients. Support for the fringe parties has almost the same relative importance as distance in the inner zone while in the suburbs it is statistically insignificant. The fringe vote and its associated disenchantment are clearly more a phenomena of the inner city than the suburbs.

Application of the final model to the different zones still results in fairly well specified models that pass the tests for normality of errors, robustness, and heteroskedasticity, whereas the application of the model to the inner city results in high multicollinearity as well as spatial autocorrelation. The spatial autocorrelation is interesting in that it is negative with clustering of opposite values in the manner of a chess board. Two variables in the final model work together fairly well in the overall region but interfere when the inner zone is considered separately. The inner zone has a much higher proportion of people renting as well as a younger population. For the entire region the correlation of rental population to population aged 40-64 is -0.515 but this rises to -0.635 when considering the inner zone. This increase in the correlation is made even more significant by the smaller number of observations in the inner zone. Removing the rental factor gives Model X7 where the multicollinear effect of these two variables is reduced to an acceptable level and spatial autocorrelation is no longer significant. The youth population is highly likely to rent rather than own their homes and removal of this variable increases the relative importance of the youth variable and the β value increases to -0.458 from -0.365. In the regional model it appears that some of the effect of the youth population upon the explained variance is being soaked up by the rental variable. Finally, the political motivation factor provided a higher level explanation when using the NDP support rather than the PC.

The models from the different zones should not replace the overall regional model since it appears that the same basic processes seem to be at work and the regional model handles the different social and geographic characteristics of the two zones without any problems. The relative importance of some of the variables changes according to the which part of Kitchener-Waterloo is examined. Voter disaffection measured by the fringe party support, and the journey to the polling station have a more important contribution to turnout in the inner city than in the outer suburbs. It also seems that different parties can be substituted as a measure of political motivation depending on which zone is examined. Substitution of the NDP for the PC party resulted in a slightly higher goodness of fit while the Reform Party could be substituted in the suburbs with little loss of explanatory power. Despite these observations, it appears that the same basic process of life cycle, followed by political motivation, structure, and then protest still seems to be occurring in all parts of the study area. However, application of the final model to these areas is useful in gaining insight into how the turnout is affected by the different population types and structural characteristics.

Results of the Models

The results from this analysis indicate a number of things. The polling division level is an appropriate unit for use in electoral studies. Although this form of study requires a tremendous amount of digital data, with many layers of information, manipulation of this information is rapidly becoming easier with advances in computer processing power, easier to use GIS packages, and the gradual coupling of GIS and spatial statistical techniques to yield powerful tools of spatio-statistical analysis. GIS was paramount since a model could be developed and then applied to regional subsets of the regions. Outlier observations that had unduly high leverage points could be removed or reintegrated easily and the connectivity matrix rebuilt quickly. The variance explained as measured by the R^2 values when examined at the polling division level is considerably higher than that achieved in ecological models performed for the Royal Commission on voting in 1991 which ranged between 0.41 and 0.45 (Eagles, 1991, p.18). This is quite surprising when one considers that the number of samples is comparable (282/295 vs 377) and also the smoothing of the variance by aggregation to the riding level. Furthermore, dummy variables, although used to test hypothesis, are not employed in the final model as used in the 1991 study. Clearly the evidence indicates that polling division level analysis is a better way of examining this phenomena than the regional analysis using ridings that are more common in Canadian electoral studies. The results would indicate that political participation at the federal level is based more upon demographic characteristics of an area at a fine scale rather than top down politicization of the public by the parties.

The lack of spatial autocorrelation also serves to indicate that the scale of study is appropriate to this social phenomena. The variation in voting is well described by the polling division and the determinants of participation do not seem influenced by the social and geographic characteristics of the surrounding region. Had the participation process been operating at a regional level rather than a more local neighbourhood level captured by the fine areal units used, then problems would have been encountered regarding variable significance since an important component was missing. Low R^2 values and spatial patterns in the residuals would also have been probable. Initially participation was thought to be regional in nature since the correlograms of the variables seemed to indicate that the political variables had more spatial association, indicated by longer lags in the correlograms, and seemed to be regional phenomena over the entire study area. A possible explanation for this would be that both political and social phenomena follow the same variance structure and vary together. Some experimentation was performed with a spatially lagged regression model using lagged explanatory variables but the results were clearly inferior to the regular OLS regression. Use of spatially lagged variables was clearly inappropriate and their use was quickly dropped. However, the correlograms seem to indicate that this may be an appropriate technique to employ when modeling political support at this scale.

The residuals (discussed in detail later) for the final model are distributed normally and are not spatially autocorrelated which would indicate that the model is well specified, the residual variance is homoskedastic, and that no major component is being missed. These factors and the lack of spatial autocorrelation indicate that the fit of the model is good over the entire region and that it can be applied to the entire study area. A pattern of over prediction and under prediction would indicate that some areas would have been following different processes in the participation process. Particular attention was paid to the inner city zone area to see if it behaved in a significantly different way from the outer zone which is more heavily dominated by suburbs and the characteristics associated with suburban living.

Demographics and Lifecycle

Scale issues aside, the most important determinant of voter participation appears to be the social cohesiveness of a region. The variables that explain the most are house ownership and demographics which can be thought of as working together to act as an indicator of the social cohesiveness in the area. This process was also observed in the national study by Eagles and the local effects mirror the broader processes observed at the national scale. It can be seen from the results that a highly mobile population characterized by high numbers of tenants who rent and a large youthful component have little interest in political participation while more interest is aroused when there is an older population that is more firmly established in an area. The coefficients for the percentage renting and the population aged 18-24 are negative indicating lower participation associated when these variables are high. The coefficient for the variable representing population aged 40 - 64 is positive showing higher participation rates when this variable is high.

Issues of lifecycle are clearly a factor and it would appear that politics has little to do with the concerns of a younger population. Furthermore, a highly mobile population is less likely to be affected by politicization. Any positive affects that a popular and well known candidate would bring to the election are negated with a highly mobile population as the electorate is less likely to be familiar with the candidate. The importance of demographics rather than income determinants as being important components of Canadian political activity seems to support the theory of brokerage politics and the idea that politics has transcended the class interests from which political parties arose and used to represent.

Multicollinearity is an important consideration in statistical analysis and it would seem that inclusion of two variables representing population age in a regression model would cause statistical problems. In most circumstances, a particular phenomena would be represented by only one variable. However, the youth category corresponds closely with the age group associated with students who make up such a strong component in the population fabric of Kitchener-Waterloo. During the iterative process of regression model building and evaluation, the youth variable was removed to see if the age component could be simplified by using only one variable. The population aged 40 to 64 was the preferred variable since it had more explanatory power. However, models that did not explicitly incorporate a youth variable often had unstable variables and exhibited positive spatial autocorrelation among the residuals. This is probably due to the geographic distribution of this age group and its concentration about the universities and the inner city zone centered along King Street. Generally, it is not a good idea to have two variables representing similar phenomenon but it is justified in this instance and is even critical to a well specified model. It is not known if this youth component and its spatial distribution are important in political participation in other cities or whether it is a unique characteristic of the university dominated region of Kitchener-Waterloo.

Political Motivation

Political factors also need to be addressed in this model since not all people in the same stage of their life cycle will behave in the same manner or will have the same motivation to vote. Issues and beliefs will have a role to play as well as the promises made by parties during the campaign. In the multivariate models of participation derived using stepwise regression (allowing the statistical program to choose best fit) a political variable, measured as percentage of support for a party, always seemed to come into play after the life cycle variables of age and accommodation. The PC, Reform, and NDP votes could all be incorporated at this stage to measure the politicization of the area. The electorate will have some sort of motivation, perhaps related to but not determined by life cycle, to participate and support their party on polling day. In the personal decision to participate, this will tend to be motivated by usually one party but in an aggregate study all of the parties are going to have some political influence on the inhabitants in the riding. Due to multicollinearity problems, not all of the parties could be incorporated into a measure of political motivation to participate and only one was used. A more sophisticated measure of political motivation could have been derived from a factor analysis of the study area which would have probably resulted in one component that would have been associated with all of the parties. Often in the stepwise regression, use of two political parties resulted in both being significant with both having a positive increase in the R^2 but the model would start to suffer from multicollinearity. The statistically significant contributions from the political parties in Model X1 could be interpreted as meaning that the political motivation to vote is quite complex and transcends any one party when examined in an aggregate fashion. However, it was decided to stick to only one variable that resulted in a lower R^2 value but a more robust statistical model. The PC party was chosen since it was one of the major parties and it explained more variance in the multivariate environment than any of the other parties. Reform and the NDP could equally have been chosen but curiously enough not the Liberals. Support for the Liberal party was not significant at any stage. Liberal support, unlike the other parties, was quite strong and seems to be relatively constant over the area, showing little relative variance that could be related to participation. Alternatively, this could indicate that in spite of the unprecedented win of the Liberals in this election, their support may have been due to voter disenchantment regarding other parties or a result of strategic voting where a party receives support in order to deny a split vote that would allow another party to get a plurality. Overall, it would appear that the Reform was a beneficiary of this at the national level to counter the Bloc Québécois. In the study region, the NDP seems to have been a loser in this respect. The scatterplots seems to indicate a NDP political behaviour similar to that of the fringe parties rather than the other national parties. In the Kitchener-Waterloo region, support for the Liberals, in spite of their strong showing, does not seem to be related to participation.

Protest Factor

The stepwise modeling revealed a protest component acting inversely to turnout that always seemed to come into play once life cycle, structure, and mainstream political support were incorporated. This factor was usually represented by support for the fringe parties or the NDP. Incorporation of the fringe party vote yielded a slightly better increase in the R^2 value. Some voter studies have revealed that alienation from mainstream politics is often associated with support for smaller parties that have little chance at being elected. The strength of the fringe vote serves as an indicator of political alienation which is associated with low participation levels (Eagles, 1991 p.14). The inverse relationship observed in this research would seem to confirm this theory and incorporation of the fringe party support into the participation model does provide a small but statistically significant increase in the explained turnout variance.

Geographic Variables

Several geo-structural variables showed up as being significant during the modeling process although only one was adopted for the final model. The size of residential area consistently showed up as being significant in a positive relationship after lifecycle and political motivation were accounted for. When area was dropped from the models, population density took its place, followed by eccentricity which served as a measure of the compactness of the polling division. This would appear to indicate that there is some sort of structural component to the participation process that is best described by the area measure but is accounted for with density, which is highly related to area, when area is dropped from the model. The shape of the polling division is also highly related to area. Furthermore, the relationship is stronger when logarithmic transformations are applied to area and density which seems to imply some sort of saturation point after which an increase in the urban density and size of residential area has little positive impact on turnout. This positive relationship with the area, shape, and density of population is curious and could be related to the visible signs of an election campaign as manifested in terms of candidate signs on lawns and billboards, polling, and visits by enumerators. These highly visible signs of an election campaign will be concentrated in areas of high visibility and in places that are highly accessible to people going about their daily business. Lower density areas that are more isolated will not be as saturated in political coverage as more dense regions. It would appear that political interest does not occur in a vacuum and that there is greater interest in areas where there is a greater concentration of population. Higher density areas would allow more contacts to occur between the inhabitants and greater opportunities to engage in discussions and invigorate political interest. The researcher has encountered several people whose motivation to vote was to cancel out the ballot of someone else. For this to occur, some sort of contact and discussion has to have occurred. It seems as if residential area and density might be acting as surrogates for a critical density of human interaction where interest in an election is generated. These two variables were dropped because of statistical conflicts with the rented variable despite significant contributions to explained variance in turnout, Another geographic variable that was consistently significant was the distance to the polling station. This relationship was negative and implies a classic geographic distance decay relationship. The likelihood of participating decreases as the distance one must travel to cast a ballot increases. The distance decay appears to be uniform as distance from poll increases. Several experiments were performed by applying power transformations to examine this decay. The distance was squared to simulate a rapid drop off in interest as distance increased but this worsened considerably the fit of this variable and rendered it insignificant. A logarithmic transformation was also examined to simulate a lag of high participation before a rapid drop off at a certain threshold distance. This also was found to have a poor impact upon the model fit and instead a straight forward uniform decline seems to appear as the distance to poll increases. This indicates a difference between individual and aggregate spatial behaviour. Individual behaviour is influenced by close proximity and ability to visually see destinations. For an individual trip, turnout would probably follow some sort of power relationship with distance. It is possible that this individual effect is present in the data set but is being averaged out over the polygon when examined in an aggregate fashion.

Some caution must be exercised regarding distance since it appears that some of the values exercise high leverage over the effect of this variable. Some of the polling divisions are very large with a relatively small inhabited area which may be dispersed in only a few sections. This may lead to a distance measure that may not be entirely appropriate for the majority of the population of that polygon. In order to try and remedy this, some of the centroids were shifted for these polygons. The centroid point was shifted to the center of the inhabited region of the polygon from its geometric center. A visual examination of the distribution of the residential areas revealed that only 32 needed to be shifted. This is less than 10% of the entire database of 377 polygons, but the modified distance measure was no longer found to be significant in stepwise regressions. This shows that these 32 points have a very strong effect upon the entire model. It could also indicate that, in these polygons, the real population distribution is not being adequately represented by the residential areas digitized and that the basic measure of distance from the centroid is adequate. Distance to poll does not act uniformly over the entire study area and it can be seen from the application of the participation model to the different zones that the distance component is a more important determinant of participation in the inner city than in the suburbs.

The perceived distance measure was not used in the final model and never made a significant contribution to explained variance at any point in the modeling process. It is possible that this measure of perceived distance was not an appropriate measure of perceived distance. Certainly the research seemed divided regarding the use of intersections and barriers in distance perception. Some postulated that more intersections represented a greater distance while others postulated that a greater number of intersections make for a more interesting journey and decreased the perceived distance to the destination. Furthermore, a precise route must be known for this type of measure to be effective in any sense and this is not possible with the aggregate behaviour of hundreds of electors following many different routes. In aggregate studies there is no fixed route for all of the voters to travel and it is probably impossible to get a simple friction measure of the barriers that must be overcome. The research by Taylor as well as discussions with various people about voting behaviour seems to indicate that perceived distance is a factor in electoral participation for an individual but it cannot effectively be measured in this type of aggregate analysis. An analysis at the neighbourhood level analysis whereby actual street addresses of participants are known may reveal the effect of perceived distance. Taylor's research in distance effects on turnout in Swansea involved only three polling stations and this ultra fine scale of geographic research may be needed to reveal perceived distance effects on turnout.

The proximity of the polling station to a major road was also dropped like the perceived distance indicator. This was incorporated as a measure of accessibility and in recognition of the important role that cars play in modern life. It was reasoned that ease of access by car from a main road might increase the likelihood to participate. This measure of distance did show up as being significant in some models but proved to be highly unstable and was dropped from the analysis.

Other Discarded Variables

The multivariate modeling process allowed some variables to be dismissed fairly quickly as being of little significance to participation. The religion variables were not highly associated with participation and also did not work with other indicators in a multivariate environment. Religion has been associated with political behaviour so its initial inclusion in the study was not inappropriate. Religious denominations have been observed to behave in a group manner politically but religion is thought to be more of an indicator of values passed on by family rather than as a conscious factor in a political decision. Religious denomination may have proved to be more important had the study area been extended to include the rural region north of Waterloo with its large population of non-voting Mennonites. In addition to religion, occupation does not seem to be related to the decision to vote. Overall income was related in a positive manner but the manner in which income is earned has no bearing to electoral turnout. A common theme during national election campaigns is the performance of the economy and issues of taxation and it is possible that those with higher incomes may feel that they have more at stake. Such issues are unlikely to be of great interest to those who have less income.

Several variables appear to be related to turnout according to the exploratory data analysis but are not included in the final models. Most of these variables seem to be related to the social cohesion of a polling division. The percentage of single parents, unmarried population, household size, and percentage of population that recently moved are all inversely related to turnout but also to the rent variable as seen in Appendix D. Rent appears to be the best single indicator of social cohesion and, when used in combination with population age variables, provides a very well specified model of turnout with a high proportion of explained variance with an R^2 of 0.580. The increase in importance of the youth population variable, when rent was removed from the model application to the inner zone, also shows how the rent variable is explaining some of the turnout variance that is associated with youth. In a different city with fewer university age students it would appear that the youth factor could probably be dispensed with as a rental variable would be highly associated with youth.

The final proposed model incorporates life cycle, political motivation, geo-structural, and protest component. Percentage of the population renting accommodation, the youth population, and the more mature population (represented by age groups 18-24 and 40-64 respectively) serve as indicators of life cycle and these three variables alone result in a well specified model explaining 58% of the observed variance in turnout. Political motivation is represented by the support for the PC party, and the geo-structural component of the polling division is represented by the distance to the polling station. Finally, the protest component is represented by the support for the fringe parties who do not really stand much chance of ever actually winning an election in the face of the larger, national parties. This model explains 64% of the variance in turnout. All variables are significant and the model passes all tests for statistical robustness, multicollinearity, normal residuals, and spatial autocorrelation. This model also provides a high level of explanation throughout the entire study area and doesn't break down in either the inner city or the outer suburbs.

Local Effects

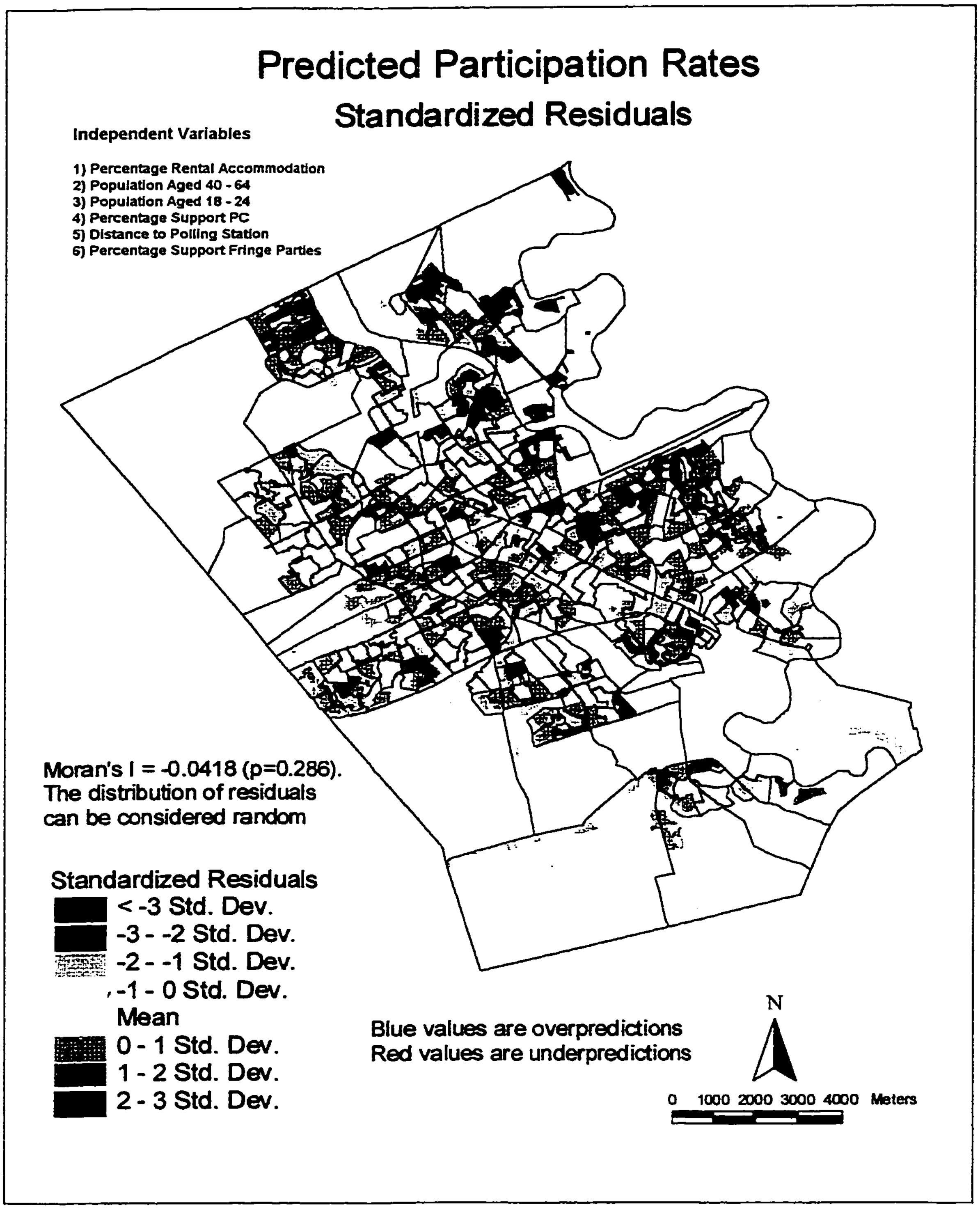
The political motivation factor so far examined and incorporated into the model seems to be effective throughout the entire study region. This motivation is probably national in nature since a national election campaign involves national issues reported in the national media. Outside of the national organizations, local candidates and their campaigns are going to try to motivate electors and turn out the vote. Some campaigns may have local high profile candidates who are well known while others will have unknowns. An incumbent candidate will also have the advantage of being well known in the community and will have a series of tangible accomplishments to present to the electorate for perusal. Also, not all local campaigns will be run in an effective manner. The effect of local candidates is difficult to measure from aggregate data although if local candidates have a effect on the turnout some sort of riding effect would be observed. When other factors are controlled some of the residual variance could be explained by the riding in which a polling division is located. A control variable representing riding membership of a polling division was used in several different stages of the modeling process but the effects of the riding in which the participation occurs seems to have no statistical effect that can be measured. Some tests were performed to model party support and in these cases the riding variable proved highly significant for predicting support for the PC, Liberals, and Reform parties. When controlling for household income and university education, the dummy riding variable made large significant contributions to the variance explained in the support for most of the large parties. This indicates that local party organizations and candidates are unable to motivate people to participate but do have an effect once people have already decided to undertake the process. Simply put, an effective local campaign can rob other parties of votes but not get any support from otherwise apolitical electors. It would appear that a great many voting decisions are made at the ballot box itself where an elector will vote for the familiar name of a high profile candidate who nevertheless had little effect on why the voter went to the polling station. Certainly the pattern of support seems to generally follow the riding boundaries as indicated in Figure 7.3. The party support shows more clearly the riding structure with the strong Tory support stopping dead at the Kitchener riding boundary. The Liberal party's strongest support in the region clearly coincides with the Kitchener riding while Reform received its strongest support in the Cambridge Federal district. Some research in electoral geography (Johnston, 1977b) has examined the contagion effects of candidates upon the political behavior of the population surrounding their homes. This approach was not adopted since any effect would be washed out by the national campaign issues and would be impossible to examine in this manner. However, an examination of the Conservative support showed that it was at its highest in the study region in the polling divisions surrounding the home of Lynne Woolstencroft who was the 1993 and also 1997 PC candidate in Waterloo. It would appear that the parties and their candidates can clearly have an effect on how people choose to vote but have a much reduced impact on peoples initial decision to participate in the process.

Residual Analysis

Much can be learned about the performance of a model by an examination of the residuals. A well specified model should have normally distributed residuals whose histogram should follow a bell shaped Gaussian distribution. A scatterplot of the standardized residuals and the predicted values should reveal no patterns. Appendix F shows these plots for residuals from the different model examined in Tables A and B although in the interests of space only residuals from the regional models are shown. In all cases the residuals are normally distributed and there are no patterns present in the scatterplots and histograms. The standardized residuals can also be mapped and checked for spatial autocorrelation. Table 8.1 includes the Moran statistic for spatial autocorrelation for each model and there appear to be no spatial patterns in the residuals detected by the statistic for global spatial autocorrelation. All of these factors indicate that the model is well specified and that there are no missing indicators. A pattern in the residuals can be indicative of an important missing variable. The standardized residuals from the overall model (Model X3) can be mapped (Figure 8.3). This visual appraisal is useful since it does yield insights that the scatterplots and the Moran statistic cannot reveal. Some localized patterns can be seen and examination of these areas can give insight into how well the model performs.

The large rural polling divisions with sparely distributed populations are generally poorly predicted. This appears to be because of their location on the fringes of the city. The residuals in these areas are generally negative (blue) indicating that the actual turnout was below that predicted according to the social and geographical characteristics of the polling divisions. This particular region was expected to cause problems for the model since it was expected to behave in a politically different way from the rest of the study area. These areas are quite isolated politically and physically since they are part of the city of Waterloo but assigned at the Federal level to the riding of Cambridge whose bulk of electors is located to the southeast on the other side of Highway 401 which serves as a physical and psychological barrier. One polling division is almost entirely cut off from the surrounding region by highways 401 and 86, and the Grand River and this particular polling division is very poorly predicted. These areas do appear to be suffering from some form of isolation that does not appear to be present in similar polling divisions in the

Figure 8.3) Standardized Residuals



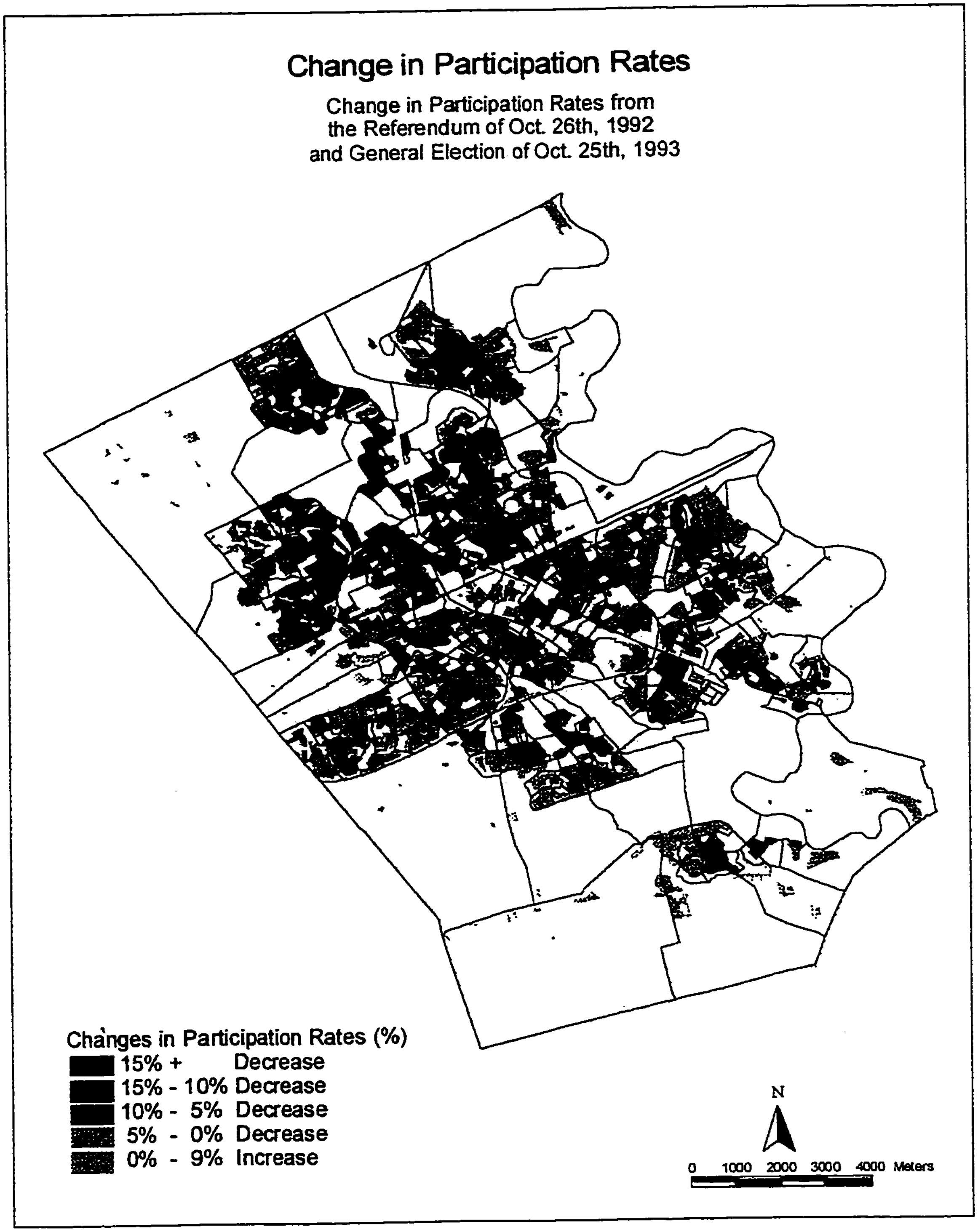
northern part of the study region. This particular area was severely over predicted. This serves to illustrate some of the problems that can occur when political boundaries do not match the distributions of the communities and populations they are meant to serve. This is not to suggest that this has been done deliberately during the 1987 redistribution upon which the 1993 election was based but it does show how effective gerrymandering could be when communities are broken up into inappropriate political units.

A further area where the model seems to be over predicting participation is in the regions surrounding the University of Waterloo. This appears to be in spite of the inclusion of the population aged 18 to 24 variable that should help account for the student distribution. This may be because the proportion of students in the population is under estimated due to the census being performed in June when most students are gone. Another related reason why the participation rates are lower in the student areas could be due to overestimation of the number of eligible voters. Since the list was based on the 1992 list and then revised it is possible that many people who were enumerated in 1992 may still have been on the list even if they had left the area. This would lead to an inflation of the estimate of the number of electors in the riding and results in an artificially low participation rate. Although it is illegal to vote twice it is not illegal to be on the voters list in two ridings. An artificially low participation rate due to overestimation of the electorate is probably a problem throughout the study area and would expect to be part of the random error involved in the modeling process. However, it is expected that areas that are highly associated with students would be especially susceptible to this form of administratively induced low participation. There is little that can be done to estimate the effect of this phenomenon although we can examine a map of the change in participation rates between 1992 and 1993. Comparisons between events at the polling division level are quite difficult since the polling division boundaries will usually change even if the riding boundaries do not. The 1992 and 1993 events both used the same spatial units so geographic comparison are possible (Figure 8.4). Overall there was a drop in participation between the two events but we see from the map that the pattern is not random and that the drop is at its highest in the residential areas surrounding the two universities in Waterloo. This could be as a result of overestimation of the eligible electors in the 1993. Alternatively it could due to the nature of the event. The 1992 event involved a direct input into the acceptance or rejection of a national issue while the 1993 event, although still national in scope, involved local ridings and candidates which may not be relevant to students who are often from out of town.

LISA Statistic

A local indicator of spatial autocorrelation (LISA) statistic was employed to determine if the residuals displayed areas of local spatial autocorrelation. The Moran statistic provides a global measure but it does not reveal more localized areas where values are interacting spatially. Furthermore, local areas of positive and negative spatial autocorrelation may cancel out each other in the global measure. A LISA statistic is often used on univariate data in order to detect outliers. In residual analysis, the LISA statistic gives an indication of the spatial clustering of highly similar values around an observation. (Anselin, 1995, p.42). These values can be negative or positive. This allows the Moran

Figure 8.4) Change in Participation Rates from 1992 to 1993



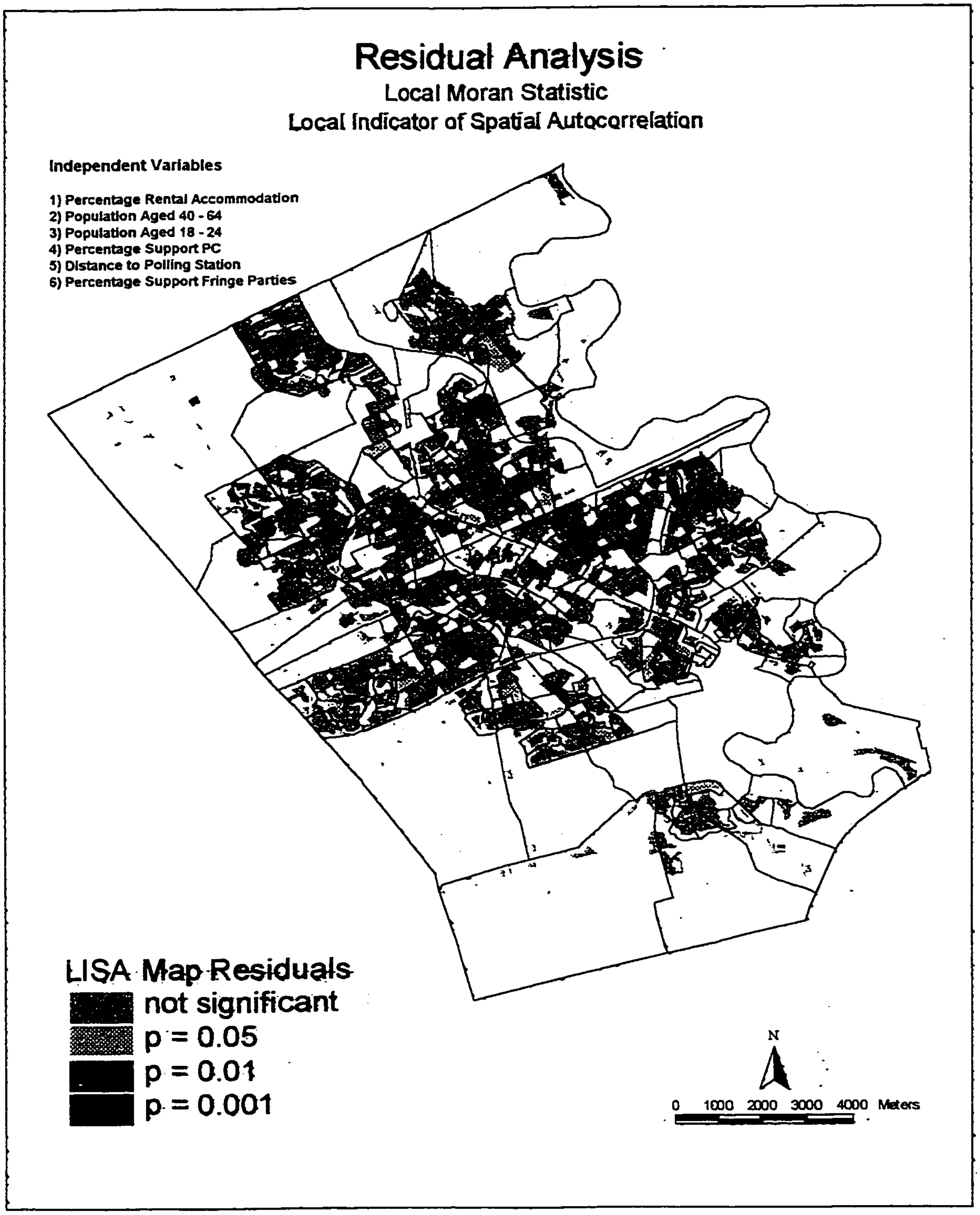
statistic to be decomposed and examined for regions where an observation is surrounded by statistically significant values that are highly similar or highly different. The Local Moran statistic is defined as:

$$I_i = z_i \sum_j w_{ij} z_j \qquad (5)$$

where z_i and z_j are values expressed in deviations from the mean and W_{ij} is a row-standardized spatial weights matrix.
Equation from Anselin (1995) p.42.

The LISA statistic gives a researcher some indication of how well the model has predicted an area although interpretation can be sometimes be difficult. Polling divisions with significant LISA probabilities (Figure 8.5) are areas where either a value has been well predicted but is surrounded by poorly predicted polling divisions, or vice-versa. A map of the LISA values must be used in conjuction with the z-score residuals map in order to determine which situation is occurring. Some cases where a significant probability occurs can be attributed to an extremely different value in the polling division i.e. a low rent complex with a young population surrounded by an older neighbourhood. Other cases may be caused by a clustering of similar poorly predicted values indicating some sort of local trend affecting participation. Generally, it seems that more values were under predicted than over predicted and the LISA and residual maps indicate that there are some polling divisions surrounded by areas that are poorly predicted. An examination of some of these regions reveals that the PC support in some of these is quite low while the Reform Party support is quite high. The PC support is used as an indicator of the political motivation and acts in a positive manner. Therefore, when the PC support is low the

Figure 8.5) Local Indicator of Spatial Autocorrelation



model tends to under predict. This is a problem in some polling divisions that have characteristics associated with the outer suburban region, e.g. low percentage of the population renting, older population, longer distance to poll, etc. but with low PC support. This demonstrates one drawback of using a single party indicator for political support instead of some sort of composite indicator derived from a principal component analysis. When considering how the PC party lost political support of the far right to the Reform party, it is surprising that so few polling divisions suffer from under prediction of this nature. This is probably because the political dimension to the participation process is quite small. The life cycle variables (rented accommodation and population age) account for 58% of the variance on their own and the political dimension, when added in the form of the PC vote, adds only 4% to the increased explanatory power of the model. If the political motivation to vote was especially strong, a more complex indicator would have to be constructed. As it stands, using only one political variable seems to work fairly well without the statistical problems associated with highly collinear variables that would occur if other mainstream political party support was added. The LISA plot also revealed polling divisions that were extremely different from their surroundings because of the housing characteristics. These areas had demographics similar to the outer suburban region but had housing characteristics more commonly associated with the inner zone. Some regions in the Heritage Park neighbourhood near Victoria Street seem to be under predicted because of higher Reform support at the expense of the PC party. This also appears to be the case in the Pioneer Park region. The Concordia Park area in Forest Hills was under predicted because of a high percentage of renters in its population despite other it's other demographic characteristics that are similar to suburban areas where there is generally a low proportion of renters. This also shows up in the LISA statistics since the surrounding areas themselves are all highly similar in having low PC support although in these cases, as in other parts of Kitchener, the Liberals enjoy especially strong support at the expense of Conservatives. A better specified measure of political support may have reduced some of the pockets of under prediction that seem to be present in the southern Kitchener region. Use of the LISA statistic aids in residual analysis since it does force an analyst to look at characteristics of regions surrounding an extreme value as well as characteristics of that particular polling division.

Examination of the residuals from the final model (X3) reveal a well specified model that has normally distributed errors with no global pattern to the residuals. A close examination using a LISA statistic of the residual patterns does reveal that there are some pockets of local autocorrelation that may be indicative of some problems with a mispecified political variable. This does not seem to affect the overall performance of the model but some caution must be used when examining regions with especially low support for the PC party in the outer suburban regions. These areas may be better modeled using the Reform support as a guide to political motivation. The residual analysis does reveal problems with using a one party indicator to measure support and further work would be needed to better specify this variable if there were indications that it was highly important to the prediction of turnout. However, in light of the extremely high proportion of variance already explained by life cycle variables, a single party indicator is an easy to implement measure that seems to work fairly well at the regional level. Spatial statistics play an important role in evaluating the performance of the model since pockets of similar residuals do not show up in regular non-spatial residual analysis and were also too minor in scope to be detected by the global Moran statistic.

Chapter Nine

Conclusions and Discussion

Conclusions from the Research Results

The results obtained from this examination of the 1993 event in Kitchener-Waterloo suggest that political participation is largely motivated by life cycle. Political and spatial influences are also determinants but play a minor role in relation to demographics. Younger populations that are more mobile are less likely to be interested in politics and are less likely to vote than older populations who are more established in their lives. Many measures of this social cohesion are available to the spatial analyst and a combination of house ownership data and demographics serves as a good indicator of political participation. The most important single determinant was the percentage of the population renting accommodation. The life cycle variables of rented accommodation, population aged 40-64 and aged 18-24 can be used without any other indicators to get a well specified regression model that explains a high proportion of turnout variance. Other factors also appear to be involved in the decision to participate but life cycle is by far the most important.

In addition to life cycle, there appears to be a political factor whereby people are motivated to cast a ballot. A dummy variable indicating riding membership of a polling division was not a significant determinant of participation and riding specific candidates and local campaign organizations seem to have little impact upon participation rates in the study area. It would seem that the political motivation factor is probably the result of national issues, national party platforms, and the images and abilities of national leaders. A lack of motivation based upon riding specific candidates could be related to the lack of power that elected members of parliament actually have. In the present system policy decisions are made in cabinet and votes in the parliament are bound by party unity. As a result, a large part of the population does not have much faith in the representative abilities of their MPs (Clark in Aubry, 1998, A1). This could also explain the presence of a disaffected electorate whereby a percentage of the population just cannot be bothered with the process. Disaffected voters can be best measured by the support for the fringe parties that are not expected to be elected. This research did find that areas with low participation are associated with higher support for fringe parties. Incorporation of the fringe parties into the regression model explained just under 1% of the observed variation in turnout.

These findings are not radically different from those obtained from other studies of participation in Canadian politics. What is different is the incorporation and the significance of the structural component of the city and electoral units. Many of these variables were examined and it was found that the distance to polls and the size of the residential area within the polling division contribute to the participation process; the former in an inverse manner and the latter in a positive manner. Incorporating a structural/geographic component to the participation is new to this study. These two variables, residential area and distance to polling station, are certainly not the most important ones in determining participation but they do account for 4% of the explained variance in the observed participation. The area component was dropped in the final model because of multicollinearity problems with other variables that did a better job of predicting turnout despite the significant contribution it made to the explained variance. Examination of how the model performed in different parts of the city showed that the distance component to voting was much less important in the outer suburban areas than in the inner city region of Kitchener-Waterloo. This is probably due to the greater importance of the automobile in suburban regions where distance is not really an issue. Measures of perceived distance did not prove to be important. This may be due to the aggregate nature of the data and techniques used since some geographical research does suggest that this type of distance perception is important at an individual level. In the final proposed model, the structural component of electoral turnout is represented by a straight forward distance decay measure.

Overall, the model presented here is quite simple. A basic least squares regression model is used with spatial diagnostics. From the point of view of a spatial scientist, such a model is non-spatial and quite unexciting since it does not incorporate any spatially lagged variables, spatial regimes models, or spatial error models. The results and diagnostics from the regular OLS models did not suggest that any spatial contagion effects were present and there just didn't seem to be any justification to apply more complex models when a regular OLS regression provided a well specified model. The final model (X3) explains a fairly high proportion of the observed variance in voter participation. The R^2 value of 0.641 is indicative of a good fit to the data. The model also has significant variables that behave in a logical manner and residuals that are distributed normally with no spatial autocorrelation. The model fits the entire study region and appears to handle both the outer suburbs and the inner city with few problems. All of these factors suggest that political participation is, despite the incorporation of distance to polls and residential area measures, largely an aspatial process. The participation rate of a polling division can be predicted with confidence based upon the characteristics present in that polling division with no reference to its neighbors.

This lack of spatial contagion in the model is a function of scale and the level of analysis adopted in this research. The polling division is a much finer spatial unit than that used in most Canadian voter participation research and it does a good job of capturing the variance in the data set for both the socio-economic variables as well as participation. Most of the correlograms (Appendix D) indicate that the social-economic variables show little interaction beyond their immediate neighbors. The use of appropriate spatial units seems to preclude the use of sophisticated spatial statistical techniques. The correlograms also indicate that the variables representing support for political parties seem to be more regional in nature. If political support were to be modeled rather than participation the conventional OLS regression techniques used in this analysis would probably break down due to the mismatch in the scale at which the socio-economic and political variables are operating. Further political analysis will probably benefit considerably from spatial statistical techniques but they are not required for modeling voter turnout. The basic OLS model using fine scale spatial units provides results that are better than that of the recent Royal Commission on Electoral Reform that specifically looked at electoral turnout. The levels of explanation achieved in that study ranged from 0.41 to 0.45 according to how the model was specified (Eagles, 1991, p.18). This indicates that political participation is a result of social and life cycle processes that occur at the city and neighbourhood level rather than the coarser national level that uses regions and ridings as the units of analysis. However, this coarser level of analysis may be appropriate for an examination of whom the electorate support. There is clearly an aspect of political motivation involved in participation but it does not seem to be a result of local processes. The maps of political support in the study area do seem to indicate riding biases and a dummy variable indicating riding membership proved significant in exploratory models of political support undertaken during the course of this research. Overall, it appears that turnout during a national election is not highly related to political motivation on a riding level but rather the result of social and demographic factors that are better examined using fine scale units at the city level.

Study Limitations

Caution must be exercised when interpreting the results of this research. Firstly, the aggregate level approach works well for looking at spatial behaviour for an area or a region but this cannot be used to predict individual behaviour. There are advantages to this approach since it is data rich and data from actual electoral and census events can be used to examine aggregate behaviour. Such an approach avoids the ambiguities, personal motivations, and misinterpretation that are so common with survey data that is used to research individual behaviour. Furthermore, a region can be examined and regional trends and differences can be noted and modeled. This cannot not be done effectively with survey data. The aggregate approach is limited in that some potentially important aspects of participation cannot be explored. An example of this would be difference in turnout between men and women. This cannot be looked at an aggregate level since variations in the distribution of the sexes throughout the population are so weak that any statistical relationships to variation in turnout are so weak as to be invisible. This is not to say that the relationship is not present but rather that the methods used in this research do not allow it to be explored. Individual motivations are impossible to examine using aggregate methods of analysis and this is in large part why they are often criticized.

Secondly, this research is derived from only one urban area and the results may not be applicable to another region that followed a different form of development. The results showed that the number of young people aged 18 to 24 was an important determinant to participation rates. The may indeed be applicable to other cities or it may be a result of the large student population living in the Kitchener-Waterloo region. The model will probably work fairly well in other small Canadian cities similar to Kitchener-Waterloo that have large university districts. Peterborough, Kingston, and St. Catherines are places where the model may be applicable. A city that is not dominated by universities and their associated demographics may behave in a different manner. Furthermore, the region researched is an urban area and application of the model to a study region that has rural polling divisions is not recommended. Different processes may be at work in a rural region that has different structural and demographic composition to urban areas as well as slightly different electoral rules regarding enumeration.

Thirdly, only one event has been modeled and this limits the ability of the researcher to observe temporal trends and predictions. An election involves personalities, issues, and even political parties that are related to only one event and may not have been present in other events. The issue of expensive military helicopters was unique to the 1993 event while the Reform Party and Bloc Québécois were absent from the 1988 event. The lack of temporal data also makes it difficult to understand some of the processes that seem to be at work in the event examined. The research indicates that participation can be broken down along demographic groupings with an older population being more likely to vote than a younger one but with only one event to examine we cannot explore this any further. Will the young become more likely to participate as they age and develop different needs and lifestyles or is the drop off in participation related to generations born in different eras with different concepts of good citizenship? This relationship can only be examined by analyzing several events over time. The demographic component of the model is quite important and, since it is not known how this behaves over time, application of the proposed model to other events should be limited to fairly recent elections until the temporal behaviour of the population is better understood.

Fourthly, this research is also limited by the quality of the enumeration. The calculation of accurate participation rates depends upon a good knowledge of how many electors live in the polling division. The quality of the enumeration may be uneven and a highly mobile population is difficult to record. Participation is based upon who is enumerated and the enumeration card presented to the polling clerks at the polling station is mailed to the elector. This assumes that one has a home. The homeless, living in shelters and the streets of the city, are unlikely to be enumerated and will not be able to vote. Discussions with social workers indicate that some efforts are made at shelters to get the homeless enumerated but with little success. Free trade agreements and helicopters are of little interest if one does not know where to sleep that night.

Discussion

This study has focused upon determinants of participation in Kitchener-Waterloo in the 1993 Federal Election. In addition to these findings, which are specific to a particular place and event, other broader issues must be examined that will have an effect upon political participation. Changing administrative procedures, emerging technologies, and the very nature of the political system will all have an impact upon future turnout in Canadian electoral events and in how participation is analysed.

Data Quality Issues

Participation rates may fluctuate due to administrative reasons in addition to the factors listed above. Turnout is impossible to gauge accurately if the information regarding the number of eligible voters is inaccurate. An overestimation of the number of eligible electors will have the effect of underestimating the actual turnout since some listed as living in the polling division may no longer be there and will be listed as not voting. This is a phenomena that is difficult to measure since the enumeration is not a perfect process and any list of voters is out of date almost from the moment it is created. The list used in the 1993 event was a year old although polling divisions with highly mobile populations were re-enumerated. It is likely that the much of the participation figures used in this research are underestimated. The results from this research show that younger populations are less likely to vote than older more stable populations. These younger populations are also more mobile and are more prone to administrative overestimation of the electoral list and it is difficult to know just how much the decline in participation for the younger population can be attributed to genuine disinterest or to administrative problems in estimating the electorate. Furthermore, the present system of enumeration is heavily dependent upon the voter having a residence and an address. This makes it very difficult to register the homeless and other needy elements of society whose voices are often unheard in politics.

Clearly there is a need for some sort of voter registration system that is constantly updated. Without such a system turnout will remain difficult to accurately estimate and model. Elections Canada has taken a step in this direction with its creation of a permanent list of electors that uses tax information to maintain an up to date digital database of electors and their locations. This is an area where GIS technology is contributing to the management of this vast spatial database. The permanent list of electors project when completed will greatly ease the administrative problems of running an election and offers the potential to save a considerable amount of money. However, while clearly being a more efficient way of keeping track of electors than the present system, it is not necessarily a better way. This system when fully implemented is supposed to eliminate the need for a full enumeration but the use of tax records will still tend to favor an older population which is more stable with homes and steady jobs rather than the young or the poor. Driver information and birth records are being supplied by provincial governments but obtaining additional voter information from welfare rolls and student loans offices may offer some possibilities for improving the quality of elector information. An additional problem with this type of system hinges on the fact that some people just don't want to reveal personal information to the government under any circumstances. The present system is voluntary and the 1998 T4 tax forms were the first to ask a taxpayer for their consent to be added to the electoral roll. However, aside from these concerns, the permanent list of electors offers the potential to improve the quality of the electoral rolls and to make it easier to examine the process of political participation.

Use of GIS and Spatial Statistics for Social Analysis

GIS and spatial statistics have much to offer in political analysis. They are particularly well suited to small scale analysis which is data intensive and highly spatial in nature. GIS allows models to be developed and applied to an entire area or a subset of the region. Spatial statistics aid in the development of statistical models that account for spatial dependence in the database that may invalidate results obtained with traditional techniques. GIS enables mismatched spatial databases to be integrated and fine scale analysis performed when previously it was inhibited by lack of information at this scale. The correlograms indicate that this scale of spatial unit is good at capturing the variation in social and political data. Certainly the higher levels of explained variance in this research seem to indicate that this is a better way of examining political participation than the use of federal ridings. GIS is already contributing extensively in the planning and administration of elections in Canada (Chief Electoral Officer 1993a; 1997b) and has much to offer in the analysis and understanding of this type of urban spatial behaviour. However, spatial data analysis using GIS is still a cumbersome task.

This research made use of many different computerized programs to manage and analyze the data. Despite the numerous GIS and statistical packages available on the market, none are able to fully integrate the power of advanced GIS functions with statistical analysis in an effective and easy to use manner. *Arc/Info* proved useful for combining the social data obtained from the census and the political data digitized separately by means of the area weighted interpolation. However, since the *Arc/Info* data manager was extremely archaic and difficult to use, outside database managers had to be employed as well. *ArcView*, being a Windows package, was much easier to use and was very useful for graphic display and the management of the interpolated database. However, *ArcView* has limited GIS functions and no advanced analysis capabilities.

SpaceStat has many advanced spatial statistical functions but no display or data management capabilities. Furthermore it lacked the more common statistical functions like the stepwise regression, scatterplots, box plots, etc. that are essential to the development of a robust statistical model. These features were supported by *SPSS*. These packages were chosen primarily because they were commonly available to students and faculty in the geography departments of the two universities that provided support but it appears that there are no packages commercially available that effectively combine all of the functions used. Much time was required to make systems communicate with each other across different platforms and in some cases different operating systems. The couplings used in this research are often referred to as "loose coupling" whereby information is exchanged via data files that are exported and imported into other systems. A system that would tightly integrate all of these functions would greatly ease the ability to explore a spatial database and reduce the time wasted getting information from one system to another. This would allow more time to be spent on analysis and would reduce the time spent on data management. Ironically, the GIS community on the Internet provided much support for this research in terms of advice and also utilities and scripts that extended the capabilities of different packages. One such script allowed *ArcView* to create spatial weights matrices and data files for *SpaceStat* and also display *SpaceStat* report files. This proved to be invaluable to the successful conclusion of this research. Nevertheless, the lack of a package effectively combining GIS and statistical analysis will discourage many researchers from fully exploring their databases and examining the spatial dimension which can prove to be a very important aspect to understanding and modeling a process.

A large part of the effort involved in this research was devoted to building a database where social and economic data from the census could be compared to the electoral information. This was a complicated process since the spatial units didn't match and an area weight interpolation had to be used to match the decomposed census data according to the polling divisions. It would greatly aid political research if Elections Canada and Statistics Canada were to adopt the same spatial units. Mismatched spatial units probably discourage many researchers from attempting this type of fine scale analysis (Reynolds, 1969, p.122). Both organizations require spatial units in the order of 300-400 people per unit and both undertake enumerations on this level on a regular basis. Furthermore each organization often requires information from the other in order to update their databases. Using common spatial units would seem to be a more efficient way of organizing their operations. This, however, is easier said than done since organizations, even within the federal government, often evolve separately and have different agendas, as well as different mandates and corporate cultures. Negotiations had been on going for several years during the time the researcher worked at Elections Canada regarding sharing of digital geographic databases between Statistics Canada and Elections Canada. One of the main problems regarded distribution of data to third parties since Statistics Canada charged for their data but Elections Canada had to provide data to outside parties free of charge. Intellectual property rights of the combined database was also a problem regarding joint projects.

Information Technology and Participatory Democracy

Information technology is changing the way elections are administered and also how they are analyzed. The changes discussed above regarding the compilation and administration of the list of electors offer more efficient ways of administration and delivery of electoral services (Chief Electoral Officer, 1997, p.22). However, it is still fundamentally the same voting process that has occurred since Confederation. An area is divided into polling divisions, electoral lists are produced, and polling stations are designated. The elector must still undertake a trip to the polls. The geographic dimension is present in electoral participation as demonstrated by the results in Kitchener-Waterloo, but this may not be the case for much longer. Modern information technology could remove the geographic component to voting and widespread introduction of communications technology may have a strong impact upon future turnout levels for all types of political participation although, as the following section reveals, precisely how has yet to be determined.

Recent experiences in Europe show how a voting system based up a voter registration card system run via the Internet can be implemented (Tsagarousianou, 1998, p.52). Enabling people to vote via computer or over the phone is certainly more convenient and will remove any access barriers that may prevent people from bothering with the electoral process. It may also help negate the weather factor that researchers believe contributes towards lower participation during winter events. This research does reveal that the longer journeys to the polling station result in a lower participation when other factors are controlled for in a multivariate model. However, this geographic component to the participation process seems to be very small and is clearly not important when compared to age and social cohesion. As a result of the recommendations of the 1991 Royal Commission, Elections Canada introduced a number of changes to the voting process that were designed to increase participation. The introduction of mobile polls and improved accessability at polling stations were designed to minimize the impact of access problems to the polling stations but the results have been a decrease in participation for the 1993 event and a further drop in 1997. Technology, when applied directly to administration and execution of an electoral event, will probably result in little direct increase in participation but it does have a role to play in the stimulation of interest in the political process.

Recent advances in information and communications technology have enabled the evolution of many projects involving participatory democracy via cyberspace. Many hope that this will recoup interest in politics by removing barriers to participation and enable people to interact directly with representatives, law makers, and even the bureaucrats responsible for the implementation of projects. Whether or not this input is welcome is another matter. The 1992 referendum was meant to be a rubber stamp type event and Canadian political parties have been slow to embrace new technologies to stimulate democratic participation. Parties adopted web pages but the Internet is being used simply as another medium for distribution of centrally defined goals and policies. New medium; same message with its emphasis on received wisdom. E-mail to MPs is allowed but still designed for one way, top down interaction that is common in Canadian federal brokerage politics. However, there is some evidence that communication will play a greater role in the formation of government policy and a recent example of this was consultation via the Internet regarding whether peacekeepers should be sent to Haiti (Cribb, 1998, J2). Many initiatives are originating outside the government and political establishment. One such example is the Canada by Design project at the University of Toronto which can be found at *http://www.candesign.utoronto.ca/ab_intro.html* on the Internet. This is a project designed to facilitate the exchange of ideas and views using live events and online discussions. Technology has been much more readily embraced for participatory democracy in the United States and especially Europe. Many of these particpatory democratic initiatives are occurring at the local level rather than the national. A discussion of the details and philosophies involved in this form of cyberdemocracy is beyond this research but can be found in Tsagarousianou et al's Cyberdemocracy (1998) which details citizen initiatives in Manchester, Philadelphia, Berlin, Athens, and Amsterdam. The organizers of these schemes believe that technology will reinvigorate politics and enable a return to the Athenian ideal of democracy from Ancient Greece.

Some of the problems with this form of participatory democracy involve universality of access and also require a technologically literate public. There is some danger that these initiatives will be hijacked for the purposes of special interest groups. Certainly technology and the electronic mail tend to be dominated by specific groups namely white, educated, male youths (Bryan, 1998, p.160). However, there is no dispute that cyberspace provides a powerful medium for political motivation. This was recently demonstrated in Minneapolis concerning the use of public funds to build a new stadium for the Minnesota Twins. The plan had originally been backed by government and the business community and the formal process of public consultation had even been canceled. A cyber based movement was able to bring such pressure to bear on elected officials that the scheme was eventually dropped (Cribb, 1998, J2).

Information technology has also been labeled as the culprit in regards to recent drops in political participation in the apathetic apolitical nineties. Modern communications technology such as the Internet, cellular phones, and the home office serve to isolate the individual from contact and physical interaction with the rest of the society (Tsagarousianou, 1998, p.54). While this can be tremendously liberating on a personal level, people cease to be members of a society designated geographically and instead become cybercitizens in a cyber based society of their own choosing. Issues and concerns will be entirely personally defined. Concerns such as initiatives for the renewal of the inner city will be meaningless to someone working in a home office who never has to visit the area. Technology has been especially embraced by the young and this research has confirmed that the youth of society are the least likely to participate in the democratic process. Overall, information technology has much to offer in the administration of electoral events and can make the process of voting easier. It is seen by some as having the potential to stimulate political debate by challenging the monopoly of existing political hierarchies (Tsagarousianou, 1998, p.167) but technology is also seen as a fearsome threat to participatory democracy as it cuts the social ties and concerns that turn a collection of individuals into a social community.

The Future of Participation

The present electoral system is heavily controlled by the political parties and there is potential for political tampering in the voting process although there is no evidence of this occurring in Kitchener-Waterloo in 1993.The staff that run the elections are political appointments and control the delineation of polling divisions and the enumeration process as well as the polling station on polling day and could certainly use their influence to interfere if they so choose (Brook, 1991, p 216). Elections Canada is an independent, apolitical organization and allowing it more control over the staff that run elections may help to shed the poor light in which politicians and their parties have been viewed recently.

Representation of the electorate is important since administration of government is increasingly bureaucratized with programs being developed, implemented, and administered by public servants and it is quite easy for MPs to lose contact with their constituents and adopt the goals, priorities, and methods of these public servants (Irvine, 1979, p.78). There is a danger of marginalization of the electorate and their representatives to the priorities of unaccountable technocrats. An important question to consider is how better participation and consequently representation can be achieved with the existing electoral system. Much research has been done regarding alternate forms of electoral systems. Some research has suggested that a system involving proportional representation rather than current "first past the post" system would result in a political system that is far more representative. Such a system would better address the priorities of the supporters and resuscitate political interest (Irvine, 1979, p.79). Discussion of merits or disadvantages of proportional representation and allocation is beyond bounds of this research. However, it is the feeling of the researcher that participation will not increase by simply changing the mathematics of getting MPs into Parliament. There needs to be some way that MPs can remain accountable after the 36 days of campaigning are over. Some form of electoral and parlimentary reform is needed to make MPs more accountable to the views of their constituents and make citizens feel they had more input in to the process of government. Detachment and alienation from the political and decision making process has often been cited as a prime cause of the declining turnout out rates in the West (Bryan, 1998). A detailed discussion of parliamentary and electoral reform is beyond the scope of this work, although readers are directed to Chapter 6 of John R. Short's The Humane City (1989) which extensively examines political reform and its geographical implications.

Further Research

This study looks at political participation during one electoral event in one particular region. The results of this research show that fine scale political analysis using GIS and statistical analysis works well. The results are only appropriate for the Kitchener-Waterloo study area and research should be performed on other cities to see if similar levels of explanation using the same variables can be achieved. Kitchener-Waterloo is quite different from other Canadian urban regions in having a highly dispersed structure as well as having different population characteristics because of the student population. In addition to analysis of other places, research over time should also be performed. It is highly possible that temporal trends in participation may exist that are not evident when a single event is examined. Fine scale temporal analysis of electoral events is hampered by polling divisions and even ridings that shift boundaries between events. GIS offers a way of spatially joining data from many events to allow fine scale temporal analysis to be performed.

Inclusion of different variables may also yield different results. Distance to the polling station was much less important in the car based suburbs than in the inner city. Incorporation of some sort of indicator of parking availability at the polling station may have a greater impact on the decision to vote in an automobile based society than the distance involved. Finally, the use of spatial statistics seems appropriate to the analysis of political support as well as participation. The effect of individual candidates and the effectiveness of local campaign organizations will be riding specific and the riding in which a polling division occurs will probably be important. This is something that is not significant when modeling participation.

Concluding Remarks

The research undertaken shows that participation in the political process is largely a result of life cycle. The structural components of the riding and polling division play a part but it is not that large in comparison to life cycle factors. The results indicate that a general model can be applied to both inner city zones as well as suburban areas. The population characteristics of these regions are quite different but the same combination of demographics, political motivation, geography, and voter disaffection is at work in both areas of the city. However, the importance of these factors seems to vary throughout the city as shown by the coefficients for distance. Electoral turnout in the inner city is much more subject to geographic influence than it is in the automobile dominated suburbs. This research fills a void in Canadian electoral studies by applying fine scale spatial units to the analysis of political behaviour instead of the larger, highly aggregated spatial units more commonly used. Use of the polling divisions reveals the fine scale structure to electoral turnout. This research indicates that use of these units is appropriate and the results show that spatial contagion is not present in the final model presented. Analysis at a fine scale shows how participation can be reasonably determined by the unique social characteristics of a polling division. The final model is quite simple in term of statistical technique but considerable effort was involved in creating a database that captured variation at the intra urban level to allow this model to be applied. The problems involved in matching disparate spatial databases and the time and expense involved in digitizing information may well discourage further analysis of other areas and events. Without further studies, it will not be known just how the results of this research are representative of Canadian cities or how much they are unique to Kitchener-Waterloo. The importance of the youth population in the model combined with the large student population of the region suggests that the results may be quite unique to the region of study.

Voting as a geographic behaviour may well be supplemented or possibly replaced by voting as cybernetic behaviour. New technologies may remove geo-structural barriers and reinvigorate political interest which has tended to stagnate at the end of the twentieth century. However, it does appears that the greatest barriers to be overcome regarding participatory democracy are the very nature and structure of political power.

What can be done to increase citizen participation? Better information, better location of polls, better access to polling stations, or even some sort of electronic system where any geographic barriers to voting are removed may help increase access but there is little evidence that declining participation is a result of these structural components. Changes will have to be made to the way politics function in Canada for there to be any sort of renewal of interest in participation and for citizens to exercise their rights of citizenship. In the present system MPs serve more as ambassadors of a political party to a riding rather than as representatives in Parliament of the population in question. Furthermore, unless the MPs are members of the cabinet, they have little to contribute to party policy (Clark in Aubry, 1998, A1). When the supposed representatives of a population are themselves excluded from the corridors of power it is hardly surprising that people are apathetic and would rather not bother with the entire process. MPs effectively lose the ability to represent their constituents when they are not allowed to vote freely in Parliament. Recent events such as the party voting over Hepatitis C compensation have served to show just how powerless MPs are. Such a system with its top down approach and received wisdom is also incompatible with recent trends in society and technology. Whether information technology will play a role in making politicians and political parties more accountable to their constituents remains to be seem. However, as long as the present system continues, there is little prospect for any sort of increase in citizen participation and the recent trends of declining turnout will continue.

BIBLIOGRAPHY

Anselin, Luc. SpaceStat User Manual (Santa Barbara: National Centre for Geographic Information and Analysis, 1995).

Anselin, Luc. *"The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association"* in Spatial Analytical Perspectives on GIS.,(eds.) Manfred Fischer, Henk J. Scholten, and David Unwin., (London: Taylor & Francis Ltd., 1996) GISDATA IV, series editors I. Masser and F. Salgé.

Archer, J. Clark, and Fred M. Shelley. American Electoral Mosaic, (Washington D.C.: Association of American Geographers, 1986).

Aubry. Jack. *"Give MPs Power to Block Government Spending: Clark"* in The Ottawa Citizen., Wednesday, September 9th, 1998, A1.

Bailey, Trevor C. *"Review of Statistical Spatial Analysis in GIS"* in SpatialAnalysis and GIS., (eds.) Stewart Fotheringham, and Peter Rogerson., (London: Taylor & Francis Ltd., 1994).

Barnett, J.Ross. *"Regional Protest and Political Change: Social Credit in New Zealand, 1954-1972"* in.People, Places & Votes: Essays on the Geography of Australia and New Zealand.(ed) R.J. Johnston, (Armidale, Australia: The University of New England, 1977).

Bakvis, Herman (ed.) Voter Turnout in Canada., (Toronto: Dundurn Press, 1991). Research Studies, Vol.15., Royal Commission on Electoral Reform and Party Financing.

Batty, Michael. *"Predicting Where we Walk"* in Nature (1997), vol. 388, p.19-20.

Batty, Michael and Yichun Xie. *"Urban Analysis in a GIS Environment: Population Density Modeling Using Arc/Info"* in Fotheringham, Stewart and Peter Rogerson (Eds.) Spatial Analysis and GIS., (London: Taylor & Francis, 1994).

Bone, Hugh A., and Austin Ranney. Politics and Voters, (New York: McGraw-Hill Book Company, 1981).

Black, Jerome H. *"Reforming the Context of the Voting Process in Canada: Lessons from Other Democracies"* in Bakvis, Herman (ed.) Voter Turnout in Canada., (Toronto: Dundurn Press, 1991). Research Studies, Vol.15., Royal Commission on Electoral Reform and Party Financing.

Brown, Daniel G. *"Spatial Statistics and GIS Applied to Internal Migration in Rwanda, Central Africa"* in Practical Handbook of Spatial Statistics. (eds) Sandra Lach Arlinghaus, Daniel A. Griffith, William C. Drake, and John D. Nystuen., (Boca Raton: CRC Press LLC, 1996).

Brook, Tom. Getting Elected in Canada: A Political Insider Demystifies the Canadian Electoral Process., (Toronto: The Mercury Press, 1992).

Brunk, Gregory G. *"On Estimating Distance-Determined Voting Functions"* in Political Geography Quarterly., (1985) Vol. 4, No. 1, pg. 55-65.

Bunting, Trudi E. Kitchener-Waterloo: The Geography of Mainstreet., (Waterloo: Department of Geography, Faculty of Environmental Studies, University of Waterloo, 1984).

Busteed, M.A. Geography and Voting Behavior, (London: Oxford University Press, 1975).

Bryan, Cathy. *"Manchester: Democratic Implications of an Economic Initiative"* in Cyberdemocracy: Technology, Cities and Civic Networks., (Eds.) Roza Tsagarousianou, Damian Tambini, and Cathy Bryan., (London: Routledge, 1998).

Cadwallader, Martin. *"Problems in Cognitive Distance: Implications for Cognitive Mapping"* in Environment and Behavior, (1979), Vol. 11, No. 4., pg. 559-576.

Campbell, C., and Christian, W. Parties, Leaders, and Ideologies in Canada., (Toronto: McGraw-Hill Ryerson Ltd., 1996).

Canadian Federation of Mayors and Municipalities. Survey of Municipal Voting in Fourteen Canadian Urban Centres of 100 000 or More Population (Ottawa: Canadian Federation of Mayors and Municipalities, 1967).

Chatterjee, Samprit, and Bertram Price. Regression Analysis by Example., (John Wiley & Sons, 1977).

Chief Electoral Officer (1992a). The 1992 Federal Referendum: A Challenge Met, (Ottawa: Office of the Chief Electoral Officer, 1992).

Chief Electoral Officer (1992b). 1992 Federal Referendum: Official Polling Results, (Ottawa: Office of the Chief Electoral Officer, 1992).

Chief Electoral Officer (1993a). Towards the 35th General Election, (Ottawa: Office of the Chief Electoral Officer, 1993).

Chief Electoral Officer (1993b). 35th General Election: Official Polling Results, (Ottawa: Office of the Chief Electoral Officer, 1993).

Chief Electoral Officer (1997a). A History of the Vote in Canada, (Ottawa: Minister of Public Works and Government Services Canada, 1997).

Chief Electoral Officer (1997b). Report of the Chief Electoral Officer: 36th General Election, (Ottawa: Office of the Chief Electoral Officer, 1997).

Chou, Yue-Hong. Exploring Spatial Analysis in Geographic Information Systems., (Santa Fe: OnWord Press, 1997).

Clarke, Harold D., Jane Jenson, Lawrence LeDuc, and Jon H.Pammett. Absent Mandate: Canadian Electoral Politics in an Era of Restructuring. Third edition., (Vancouver: Gage Educational Publishing Company, 1996).

Clarkson, Stephen. *"Yesterday's Man and His Blue Grits: Backwards into the Future"* in , The Canadian General Election of 1993., (eds.) Alan Frizzell, Jon H. Pammett , and Anthony Westell., (Ottawa: Carleton University Press, 1994).

Courtney, John C. (eds.) Voting in Canada, (Scarborough: Prentice-Hall of Canada Ltd., 1967).

Courtney, John C., Peter MacKinnon, and David E.Smith., Drawing Boundaries: Legislatures, Courts, and Electoral Values, (Saskatoon: Fifth House Publishers, 1992).

Cox, Kevin R. *"The Voting Decision in a Spatial Context"* in Progress in Geography: International Reviews of Current Research Vol.1., (eds.) Christopher Board, Richard J. Chorley, Peter Haggett, and David R. Stoddart., (Edward Arnold (Publishers) Ltd., 1969).

Cox, Kevin R. *"Residential Relocation and Political Behavior: Conceptual Model and Empirical Tests"* in Acta Sociologica., (1970) Vol. 10, pg. 40-53.

Cox, Kevin R. *"Residential Mobility, Neighbourhood Activism and Neighbourhood Problems"* in Political Geography Quarterly., (1983) Vol. 2, No. 2, pg. 99-117.

Cribb, Robert. *"Power to the People"* in The Toronto Star.(1998), April 16th, 1998., J1.

Cribb, Robert. *"Who Needs Politicians?"* in The Toronto Star.(1998), April 16th, 1998., J2.

Eagles, Monroe *"An Ecological Perspective on Working Class Political Behavior: Neighbourhood and Class Formation in Sheffield"* in Developments in Electoral Geography., R.J. Johnston, F.M.Shelley, and P.J. Taylor., (London: Routledge, 1990).

Eagles, Munroe. *"Voting and Non-Voting in Canadian Federal Elections: An Ecological Analysis"* in Voter Turnout in Canada.,(ed.) Herman Bakvis., (Toronto: Dundurn Press, 1991). Research Studies, Vol.15., Royal Commission on Electoral Reform and Party Financing.

English, John, and Kenneth McLaughlin., Kitchener: An Illustrated History., (Waterloo: Wilfrid Laurier University Press, 1996).

Environmental Systems Research Institute, Inc., Understanding GIS: The ARC/INFO Method., Fourth edition, (Cambridge, UK: GeoInformational International, 1997).

Fisher, P.F. and M. Langford. *"Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation"* in Environment and Planning A., (1995), Vol. 27, pg. 211-224.

Flowerdew, Robin, and Mick Green. *"Areal Interpolation and Types of Data"* in Spatial Analysis and GIS., (eds.) Stewart Fotheringham and Peter Rogerson., (London: Taylor & Francis Ltd., 1994).

Forrest, J., Marjoribanks, E. and Johnston, R.J. *"Local Effect in New Zealand Local Elections"* in People, Places & Votes: Essays on the Geography of Australia and New Zealand.. (ed) R.J. Johnston., (Armidale, Australia: The University of New England, 1977).

Frizzell, Alan, Jon. H. Pammett, and Anthony Westell. (eds.) The Canadian General Election of 1993., (Ottawa: Carleton University Press, 1994).

Frizzell, Alan, and Anthony Westell. *"The Press and the Prime Minister"* in The Canadian General Election of 1993., (eds.) Alan Frizzell, Jon H. Pammett, and Anthony Westell., (Ottawa: Carleton University Press, 1994).

Gaile, Gary L., and Cort J. Willmott. Spatial Statistics and Models. (Dordrecht: D. Reidel Publishing Company, 1984).

Gargrave, Anthony, and Raymond Hull. How to Win an Election: The Complete and Practical Guide to Organizing and Winning Any Election Campaign., (Toronto: MacMillian of Canada, 1979).

Gatrell, Anthony. Distance and Space: A Geographical Perspective., (Oxford: Clarendon Press, 1983).

Gidengil, Elizabeth. *"Canada Votes: A Quarter Century of Canadian National Elections Studies"* in Canadian Journal of Political Science, (1992) Vol. 25, No. 2, pg. 219-248.

Griffith, Daniel A. *"Some Guidelines for Specifying The Geographic Weights Matrix Contained in Spatial Statistical* in Practical Handbook of Spatial Statistics., (eds) Sandra Lach Arlinghaus, Daniel A. Griffith, William C. Drake, and John D. Nystuen., (Boca Raton: CRC Press LLC, 1996).

Griffith, Daniel A., and Ayse Can. *"Spatial Statistical Econometric Versions of Simple Urban Population Density Models"* in Practical Handbook of Spatial Statistics., (eds) Sandra Lach Arlinghaus, Daniel A. Griffith, William C. Drake, and John D. Nystuen., (Boca Raton: CRC Press LLC, 1996).

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

Hall, Tony. *"The Silent Vote: Why Three-Quarters of Registered Indians Won't Show Up on Election Day"* in Canadian Forum., (1993), September, 1993., pg. 30-31.

Harrop, Martin, and William L. Miller. Elections and Voters: A Comparative Introduction., (New York: The Meredith Press, 1987).

Helbing, Dirk, Joachim Keltsch, and Peter Molnar. *"Modelling the Evolution of Human Trail Systems"* in Nature (1997), vol. 388, p.47-50.

Hodge, David C. and Lynn A. Staeheli *"Social Relations and Geographic Patterns of Urban Electoral Behavior"* in Urban Geography., (1992), Vol. 13, No. 4, pg. 307-333.

Horton, Frank E., and David R.Reynolds. *"Effects of Urban Spatial Structure on Individual Behavior"* in Economic Geography., (1971), Vol. 47, pg. 36-48.

Huxhold, William E. Introduction to Urban Geographic Information Systems., (Oxford University Press, 1991).

Irvine, William P. Does Canada Need A New Electoral System?., (Institute of Intergovernmental Relations, Queen's University, 1969).

Johnston, J. Paul, and Harvey E. Pasis. Representation and Electoral Systems: Canadian Perspectives., (Scarborough: Prentice-Hall Canada Inc., 1990).

Johnston, R.J. (1977a) (ed). People, Places & Votes: Essays on the Geography of Australia and New Zealand. (Armidale, Australia: The University of New England, 1977).

Johnston, R.J. (1977b) *"Contagious Processes and Voting Patterns: Christchurch, 1969-1972"*. in People, Places & Votes: Essays on the Geography of Australia and New Zealand., (ed) R.J. Johnston, (Armidale, Australia: The University of New England, 1977).

Johnston, R.J. (1986) *"Information Provision and Individual Behavior: A Case Study of Voting at an English General Election"* in Geographical Analysis., (1986), Vol. 18, No. 2, pg. 129-141.

Johnston, R.J., F.M. Shelley, and P.J. Taylor. Developments in Electoral Geography., (London: Routledge, 1990).

Johnston, Richard., André Blais, Henry H.Brady, and Jean Crête., Letting the People Decide: Dynamics of a Canadian Election., (Stanford University Press, 1992).

Jordan, Terry G., and Lester Rowntree. The Human Mosaic: A Thematic Introduction to Cultural Geography., Fourth edition., (New York: Harper & Row Publishers, 1986).

Knoke, David, and Geogre W. Bohrnstedt. Statistics for Social Data Analysis, Third edition., (F.E. Peacock Publishers, Inc., 1994).

Kreis, Bruce D. Arc/Info Quick Reference: Quick Answers to Arc/Info Questions., (Redlands: OnWord Press, 1995).

Lam, Nina Siu-Ngan. *"Spatial Interpolation Methods: A Review"* in The American Cartographer., (1983), Vol. 10, No. 2, pg. 129-150.

Land, Brian. Eglington: The Election Study of a Federal Constituency., (Toronto: Peter Martin Associates, 1965).

Landes, Ronald G. The Canadian Polity: A Comparative Introduction., Second edition, (Scarborough: Prentice-Hall Canada Inc., 1987).

Laponce, J.A. *"Ethnicity, Religion, and Politics, in Canada: A Comparative Analysis of Survey and Census Data"* in Quantitative Ecological Analysis in the Social Sciences., (eds) Dogan Mattei and Stein Rokkan.., (Cambridge, Mass.: Massachusetts Institute of Technology, 1969).

Laurini, Robert, and Derek Thompson. The A.P.I.C. Series Number 37: Fundamentals of Spatial Information Systems., (Academic Press Limited, 1992)., A.P.I.C. Series., (general editors) M.J.R. Shave and , I.C. Wand.

Leduc, Lawrence. *"The Leaders' Debates: Critical Event or Non-event?"* in The Canadian General Election of 1993., (eds.) Alan Frizzell, Jon H. Pammett, and Anthony Westell., (Ottawa: Carleton University Press, 1994).

Ley, David. A Social Geography of the City., (New York: Harper and Row, 1983).

Lumb, Lionel. *"The Television of Inclusion"* in The Canadian General Election of 1993., (eds.) Alan Frizzell, Jon H. Pammett, and Anthony Westell., (Ottawa: Carleton University Press, 1994).

Levine, Ned. *"Spatial Statistics and GIS: Software Tools to Quantify Spatial Patterns"* in Journal of the American Planning Association., (1996), Vol. 62, No. 3, pg. 381.

MacEachren, Alan. *"Travel Time as the Basis of Cognitive Distance"* in Professional Geographer., (1980), Vol. 32, No. 1, pg. 30-36.

Martin, David. Geographic Information Systems and Their Socioeconomic Applications., (Routledge, 1991).

McLaughlin, Kenneth. Waterloo: An Illustrated History., (Windsor Publications, Ltd., 1990).

Meisel, John. *"Religious Affiliation and Electoral Behavior: A Case Study"* in Voting in Canada., (ed.) John C. Courtney, (Scarborough: Prentice-Hall of Canada Ltd., 1967).

Montello, Daniel R. *"The Measurement of Cognitive Distance: Methods and Construct Validity"* in Journal of Environmental Psychology., (1991), Vol. 11, pg. 101-122.

Mox, Alex Yin-Cheung., The Spatial Variation of Voting., (1970), Unpublished Master's Thesis: Department of Geography, Waterloo Lutheran University.

Nasar, Jack L. *"Environmental factors, perceived distance and spatial behavior"* in Environment and Planning B: Planning and Design., (1983), Vol. 10, pg. 275-281.

Nasar, Jack L., Hugo Valencia, Zainal Abindin Omar, Shan-Chy Chueh, and Ji-Hyuan Hwang. *"Out of Sight, Further from Mind: Destination Visibility and Distance Perception"* in Environment and Behavior, (1985), Vol. 17, No. 5, pg. 627-639.

Niemi, Richard G., and Herbert F. Weisberg (eds.). Controversies in Voting Behavior., Second edition., (Washington D.C.: Congressional Quarterly Inc., 1984).

O'Loughlin, John, Colin Flint, and Luc Anselin. *"The Geography of the Nazi Vote: Context, Confession, and Class in the Reichstag Election of 1930"* in Annals of the Association of American Geographers, Vol. 84, No.3. (1994), pg. 351-380. (Cambridge: Blackwell Publishing).

Pammett, Jon H. *"Voting Turnout in Canada"* in Voter Turnout in Canada.,(ed.) Herman Bakvis, (Toronto: Dundurn Press, 1991). Research Studies, Vol.15., Royal Commission on Electoral Reform and Party Financing.

Prescott, J.V.R. Political Geography., (London: Methuen & Co. Ltd, 1972).

Reynolds, David R. *"A Spatial Model for Analysing Voting Behavior"* in Acta Sociologica, (1969), Vol. 12, pg. 122-130.

Rose, Richard.(ed.) (1980a). Electoral Participation: A Comparative Analysis., (London: Sage Publications Ltd., 1980).

Rose, Richard (1980b). Do Parties Make a Difference?., (Chatham, N.J.: Chatham House Publishers Inc., 1980).

Sadalla, Edward K., and Lorin J. Staplin. *"The Perception of Traversed Distance: Intersections".*, in Environment and Behavior., (1980), Vol. 12, No. 2, pg. 167-182.

Sãisã, Jouko et al. *"Intraurban Cognitive Distance: The Relationship between Judgements of Straight-Line Distances, Travel Distances, and Travel Times"* in Geographical Analysis., (1986), Vol. 18, No. 2, pg. 167-174.

Saul, John Ralston (1992). Voltaire's Bastards: The Dictatorship of Reason in the West., (Toronto: Penguin Books, 1992).

Saul, John Ralston (1995). The Doubter's Companion: A Dictionary of Aggressive Common Sense., (Toronto: Penguin Books, 1995).

Saul, John Ralston (1998). Reflections of a Siamese Twin: Canada at the End of the Twentieth Century., (Toronto: Penguin Books, 1998).

Scarrow, Howard A. *"Patterns of Voter Turnout in Canada"* in Voting in Canada., (ed.) John C. Courtney, (Scarborough: Prentice-Hall of Canada Ltd., 1967).

Shaw, Gareth, and Wheeler, Denis. Statistical Techniques in Geographical Analysis. (Chichester: John Wiley & Sons Ltd., 1985).

Short, John R. The Humane City. (Oxford: Basil Blackwell Ltd., 1989).

Smith, David Edward. *"Questionnaire Response, Voter Turnout and Party Support"* in Voting in Canada., (ed.) John C. Courtney, (Scarborough: Prentice-Hall of Canada Ltd., 1967).

Statistics Canada (1992a). 1991 Census of Canada: Profile of Census Tracts, Part A & B (CD-ROM). (Ottawa: Ministry of Industry, Science and Technology, 1992).

Statistics Canada (1992b). 1991 Census Dictionary. (Ottawa: Ministry of Industry, Science and Technology, 1992).

Staplin, Lorrin J., and Sadalla, Edward K. *"Distance Cognition in Urban Environments"* in Professional Geographer., (1981), Vol. 33, No. 3, pg. 302-310.

Swinton, Katherine. *"Federalism, Representation, and Rights"* in Drawing Boundaries: Legislatures, Courts, and Electoral Values.,(eds.) John C. Courtney, Peter MacKinnon, and David E. Smith., (Saskatoon: Fifth House Publishers, 1992).

Taylor, Alan *"Journey Time, Perceived Distance, and Electoral Turnout - Victoria Ward, Swansea"*, in Area, (1973), vol. 5, No. 1, pg 59-62.,(London: Institute of British Geographers).

Taylor, M.J. *"Perceived Distance and Spatial Interaction"* in Environment and Planning:A., (1978), Vol. 10, pg. 1171-1177.

Taylor, Peter J., and R.J. Johnston. Geography of Elections., (New York: Holmes & Meier Publishers Inc., 1979).

Tiefelsdorf, Michael. *"Modelling Spatial Processes: The Identification and Analysis of Spatial Relationships in Regression Residuals by Means of Moran's I"*., Unpublished PhD Thesis., (Wilfrid Laurier University, Dept. of Geography, 1997).

Thomas, R.W., and R.J. Huggett. Modelling in Geography: A Mathematical Approach., (Harper & Row, Publishers, 1980).

Thompson, Allan. *"Top Bureaucrat Urges Role for Ordinary Citizen"* in The Toronto Star, (1988), May 26th, 1998.

Tindall, C. Richard., and Susan Nobes Tindall. Local Government in Canada., Fourth edition, (Toronto: McGraw-Hill Ryerson Ltd., 1995).

Tsagarousianou, Roza, Damian Tambini, and Cathy Bryan. (eds.) Cyberdemocracy: Technology, Cities and Civic Networks., (London: Routledge, 1998).

Underhill, Frank H. *"Our Fantastic Electoral System"* in Voting in Canada., (ed.) John C. Courtney, (Scarborough: Prentice-Hall of Canada Ltd., 1967).

Walford, Nigel. Geographical Data Analysis., (Chichester: John Wiley & Sons, 1995).

Walmsley, D.J., and G.J. Lewis. Human Geography: Behavioral Approaches., (New York: Longman, 1984).

Wearing, Joseph (ed.). The Ballot and its Message: Voting in Canada., (Toronto: Copp Clark Pitman Ltd., 1991).

Woolstencroft, Peter. *""Doing Politics Differently": The Conservative Party and the Campaign of 1993"* in The Canadian General Election of 1993., (eds.) Alan Frizzell, Jon H. Pammett, and Anthony Westell., (Ottawa: Carleton University Press, 1994).

Wong, David W. S. *"Aggregation Effects in Geo-referenced Data"* in Practical Handbook of Spatial Statistics. (eds) Sandra Lach Arlinghaus, Daniel A. Griffith, William C. Drake, and John D. Nystuen., (Boca Raton: CRC Press LLC, 1996).

APPENDIX A Descriptive Statistics

VARIABLE	Units	Minimum Value			Maximum Value			Mean Value			Standard Deviation		
		Study Area	Inner Zone	Outer Zone	Study Area	Inner Zone	Outer Zone	Study Area	Inner Zone	Outer Zone	Study Area	Inner Zone	Outer Zone
AREA_LOG	log m^2	3.73	3.73	4.17	5.71	5.52	5.71	5.10	5.05	5.12	0.22	0.22	0.21
CLIPAREA	m^2	5417.83	5417.83	14818.75	512100.44	334026.10	512100.44	141408.69	124236.89	148037.51	67222.90	49682.49	71862.20
DENS_LOG	people/km^2 (logged)	2.94	3.32	2.94	4.72	4.72	4.61	3.65	3.67	3.65	0.19	0.18	0.20
DENSITY	people/km^2	862.69	2087.00	862.69	52604.09	52604.09	40691.69	5114.30	5356.85	5020.67	4180.27	4990.64	3828.25
DISTANCE_M	m	10.44	23.75	10.44	4105.66	1113.33	4105.66	525.10	384.14	579.51	504.37	236.26	566.45
DISTANCE	m	23.71	23.71	29.62	3370.85	1112.97	3370.85	529.63	390.17	583.47	467.97	235.97	521.56
ECCENTRY	area / perimeter	1.14	4.90	1.14	55.56	55.56	26.06	11.21	12.23	10.81	4.74	6.07	4.06
EDU_9	% of Population	0.00	3.00	0.00	23.80	23.80	17.30	7.86	11.47	6.47	4.63	4.49	3.87
EDU_HS	% of Population	9.20	9.20	16.80	45.80	45.00	45.80	31.68	32.62	31.31	5.97	5.83	5.99
EDU_UNIV	% of Population	3.60	3.60	6.00	60.30	60.30	48.70	18.88	19.28	18.72	9.37	10.74	8.81
HS_SIZEA	Mean size of household	1.50	1.50	1.50	3.80	3.40	3.80	2.75	2.27	2.94	0.53	0.28	0.48
IMMIGRANT_	% of Population	8.40	8.40	11.20	47.20	41.80	47.20	22.02	22.58	21.81	6.17	7.44	5.60
INC_HOUSE	dollars	23312.00	23312.00	26229.00	121742.00	70340.00	121742.00	51935.57	40215.40	56459.90	17419.01	8806.84	17817.19
INC_LOG	dollars (logged)	4.37	4.37	4.42	5.09	4.85	5.09	4.69	4.60	4.73	0.13	0.09	0.13
INTERSECTI	# per polling division	2.00	2.00	2.00	80.00	40.00	80.00	17.46	17.75	17.35	10.00	7.31	10.87
JOB_FINA	% of Population	17.90	17.90	19.30	48.60	46.50	48.60	33.05	31.63	33.60	5.73	5.78	5.63
JOB_MAN	% of Population	10.20	10.20	11.90	57.90	57.90	55.90	32.30	32.38	32.27	8.71	9.96	8.20
JOB_PUB	% of Population	0.00	7.20	0.00	58.20	58.20	46.30	20.72	20.77	20.70	7.45	6.97	6.79
JOB_SERV	% of Population	3.10	4.50	3.10	24.30	24.30	23.90	11.78	13.02	11.31	3.61	3.85	3.41
LIBERAL	% of Ballots Cast	24.89	34.63	24.89	59.94	59.07	59.94	46.23	47.78	45.63	6.48	6.00	6.56
MAJORITY	% Plurality	0.00	8.00	0.00	45.00	45.00	43.00	21.94	26.11	20.32	9.36	8.01	9.36
MOVE_1	% of Population	2.20	2.60	2.20	46.70	45.60	46.70	16.40	20.87	14.67	7.97	7.25	7.57
MOVE_5	% of Population	20.20	21.80	20.20	94.70	94.70	89.50	48.02	47.00	48.42	13.83	11.35	14.68
NDP	% of Ballots Cast	0.00	1.36	0.00	14.00	14.00	11.37	4.80	6.96	3.97	2.71	2.74	2.19
OTHER	% of Ballots Cast	0.00	0.42	0.00	9.74	8.38	9.74	3.04	3.46	2.86	1.84	1.82	1.83
PARENT	% of Population	0.40	1.80	0.40	12.30	12.30	11.10	3.39	3.84	3.22	1.72	1.45	1.79
PARTICIP	% of Electors	41.39	43.90	41.39	80.30	79.90	80.30	63.74	59.96	65.20	9.14	8.32	9.03
PC	% of Ballots Cast	7.92	9.09	7.92	43.97	40.81	43.97	20.56	19.17	21.09	6.08	5.48	6.22
POP15_24	% of Population	4.20	7.00	4.20	23.30	23.30	21.10	11.47	12.15	11.20	3.02	3.30	2.87
POP25_39	% of Population	8.50	10.50	8.50	44.30	44.30	43.00	27.36	28.64	26.87	6.07	5.12	6.34
POP40_64	% of Population	8.80	8.80	10.20	42.10	42.10	41.90	26.16	23.76	27.08	6.30	4.82	6.57
POP65OVR	% of Population	1.10	1.80	1.10	55.10	42.60	55.10	10.23	16.48	7.81	8.03	7.39	6.89
PROXROAD	m	0.87	12.73	0.87	994.84	420.88	994.84	205.70	120.25	238.68	158.70	83.59	168.35
REFORM	% of Ballots Cast	13.07	14.35	13.07	46.19	32.13	46.19	24.52	21.64	25.63	4.97	3.90	4.89
RELI_CATH	% of Population	14.00	14.00	14.10	61.10	52.70	61.10	32.93	33.06	32.88	7.56	8.45	7.21
RELI_NONE	% of Population	1.60	5.20	1.60	31.20	31.20	29.40	11.96	14.29	11.07	4.85	5.08	4.45
RELI_PROT	% of Population	23.60	23.60	25.10	72.60	72.60	70.70	49.34	47.56	50.03	8.55	9.81	7.92
RENTED	% of Population	0.00	1.40	0.00	100.00	100.00	100.00	36.26	50.87	30.62	24.86	20.12	24.26
SINGLE	% of Population	12.10	15.80	12.10	45.70	45.70	38.80	23.15	27.24	21.58	5.34	5.18	4.50
UNEMPLOY	% of Population	0.00	1.50	0.00	15.50	13.20	15.50	4.95	5.79	4.62	2.35	2.20	2.33

APPENDIX B Correlation Matrix

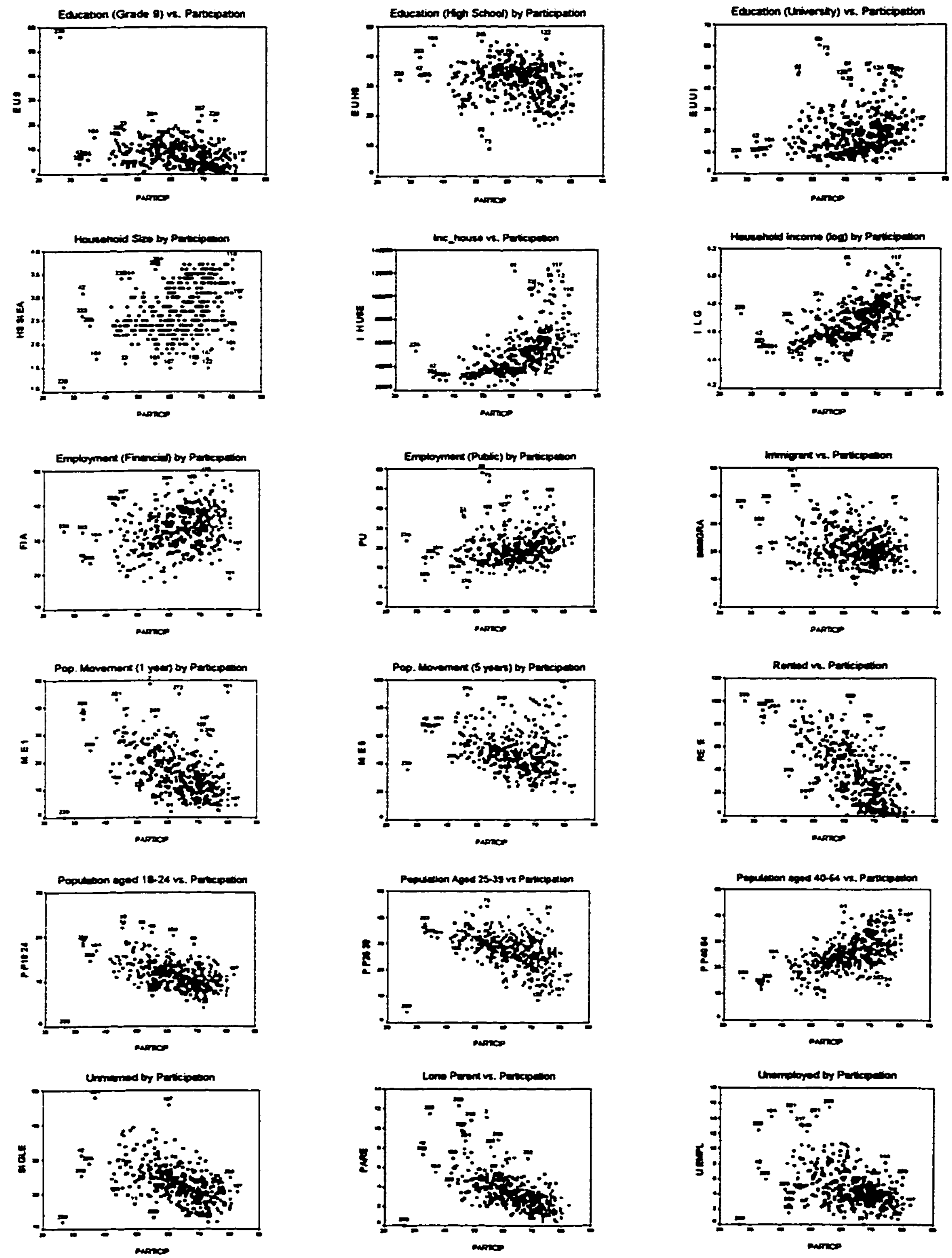
Variable	CLIPAREA	DENSITY	DISTANCE	ECENTRY	PROXROAD	EDU_9	EDU_HS	EDU_UNIV	HS_SIZEA
CLIPAREA	1								
DENSITY	-0.392	1							
DISTANCE	0.409	-0.133	1						
ECENTRY	-0.505	0.417	-0.423	1					
PROXROAD	0.209	-0.087	0.220	-0.280	1				
EDU_9	-0.260	0.000	-0.147	0.204	-0.284	1			
EDU_HS	-0.254	0.074	-0.176	0.136	-0.152	0.344	1		
EDU_UNIV	0.223	-0.095	0.062	-0.098	0.015	-0.436	-0.668	1	
HS_SIZEA	0.312	-0.183	0.186	-0.242	0.472	-0.601	-0.320	0.007	1
INC_HOUSE	0.538	-0.234	0.223	-0.216	0.340	-0.522	-0.479	0.460	0.642
MOVE_1	-0.284	0.383	0.027	0.131	-0.249	0.140	0.030	-0.007	-0.434
MOVE_5	-0.066	0.291	0.146	0.011	0.078	-0.186	-0.282	0.074	0.092
PARENT	-0.390	0.344	-0.084	0.154	-0.212	0.136	0.308	-0.341	-0.300
POP18_24	-0.284	0.233	-0.093	0.086	-0.174	-0.023	0.018	0.214	-0.276
POP25_39	-0.170	0.135	0.081	-0.098	-0.050	-0.088	-0.138	-0.159	-0.032
POP40_64	0.284	-0.208	0.043	-0.039	0.133	-0.108	0.050	0.211	0.143
POP65OVR	-0.158	0.023	-0.213	0.270	-0.319	0.645	0.294	0.016	-0.756
RENTED	-0.475	0.351	-0.169	0.235	-0.354	0.409	0.252	-0.138	-0.708
SINGLE	-0.315	0.172	-0.179	0.127	-0.347	0.235	0.153	0.171	-0.615
UNEMPLOY	-0.276	0.234	-0.039	0.177	-0.257	0.201	0.285	-0.267	-0.362
LIBERAL	-0.318	0.145	-0.228	0.123	-0.068	0.356	0.370	-0.529	-0.097
MAJORITY	-0.297	0.131	-0.313	0.174	-0.125	0.402	0.258	-0.321	-0.222
NDP	-0.336	0.114	-0.149	0.136	-0.309	0.296	0.217	-0.073	-0.451
OTHER	-0.118	-0.014	0.054	-0.010	-0.114	0.127	0.077	-0.086	-0.164
PARTICIP	0.479	-0.235	0.015	-0.169	0.261	-0.310	-0.180	0.198	0.407
PC	0.248	-0.122	0.037	-0.002	0.056	-0.295	-0.453	0.637	0.094
REFORM	0.355	-0.093	0.319	-0.239	0.242	-0.338	-0.082	-0.007	0.329

Variable	INC_HOUSE	MOVE_1	MOVE_5	PARENT	POP18_24	POP25_39	POP40_64	POP65OVR	RENTED
INC_HOUSE	1								
MOVE_1	-0.483	1							
MOVE_5	-0.036	0.556	1						
PARENT	-0.573	0.533	0.162	1					
POP18_24	-0.314	0.444	0.112	0.422	1				
POP25_39	-0.451	0.494	0.577	0.360	0.182	1			
POP40_64	0.544	-0.539	-0.526	-0.468	-0.173	-0.831	1		
POP65OVR	-0.301	0.036	-0.307	-0.091	-0.157	-0.421	0.126	1	
RENTED	-0.725	0.679	0.265	0.644	0.537	0.385	-0.539	0.305	1
SINGLE	-0.476	0.517	0.060	0.447	0.778	0.174	-0.202	0.199	0.676
UNEMPLOY	-0.518	0.499	0.190	0.534	0.395	0.319	-0.325	0.006	0.556
LIBERAL	-0.386	0.063	-0.030	0.232	-0.080	0.038	-0.082	0.090	0.161
MAJORITY	-0.367	0.101	-0.046	0.229	0.018	-0.032	-0.073	0.228	0.243
NDP	-0.544	0.361	-0.069	0.421	0.298	0.265	-0.329	0.210	0.464
OTHER	-0.256	0.146	-0.002	0.258	0.192	0.184	-0.284	0.057	0.223
PARTICIP	0.574	-0.533	-0.222	-0.604	-0.495	-0.435	0.571	-0.076	-0.731
PC	0.483	-0.145	0.049	-0.398	0.017	-0.247	0.254	0.040	-0.208
REFORM	0.332	-0.167	0.016	-0.163	-0.165	0.024	0.110	-0.311	-0.310

Variable	SINGLE	UNEMPLOY	LIBERAL	MAJORITY	NDP	OTHER	PARTICIP	PC	REFORM
SINGLE	1								
UNEMPLOY	0.434	1							
LIBERAL	0.002	0.213	1						
MAJORITY	0.158	0.221	0.829	1					
NDP	0.504	0.328	0.030	0.172	1				
OTHER	0.235	0.144	-0.129	-0.060	0.304	1			
PARTICIP	-0.535	-0.445	-0.083	-0.118	-0.485	-0.365	1		
PC	-0.050	-0.273	-0.641	-0.371	-0.315	-0.263	0.309	1	
REFORM	-0.320	-0.193	-0.494	-0.775	-0.335	-0.079	0.167	-0.097	1

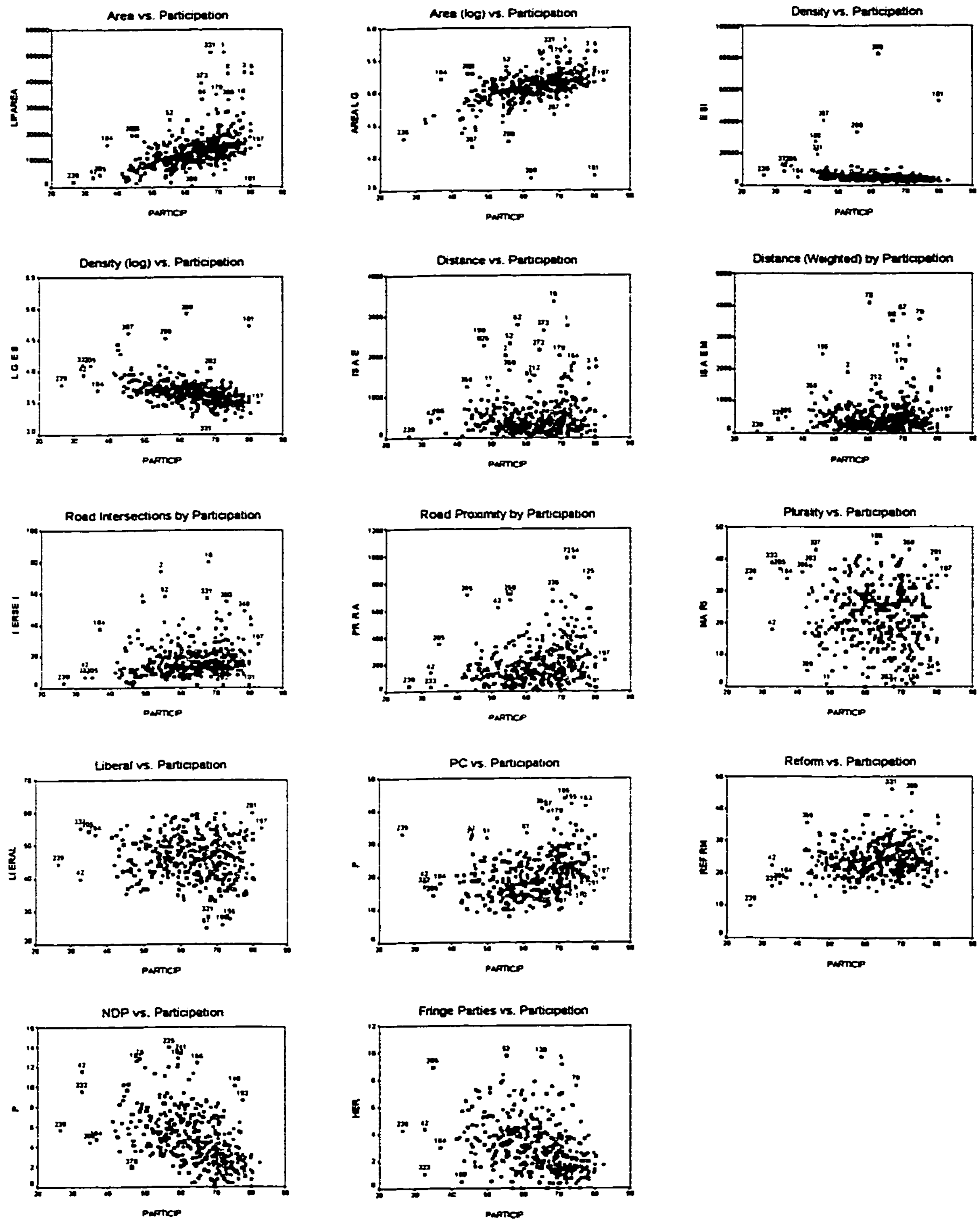
APPENDIX C Scatter Plots

Social and Economic Variables



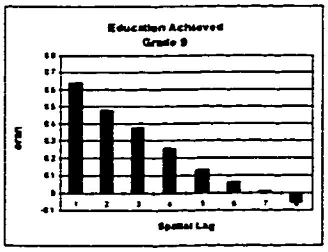
APPENDIX C (contd) Scatter Plots

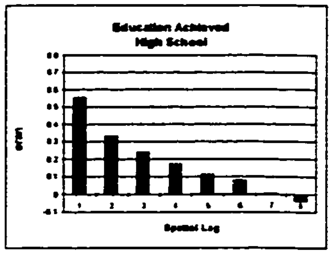
Geographic and Political Variables

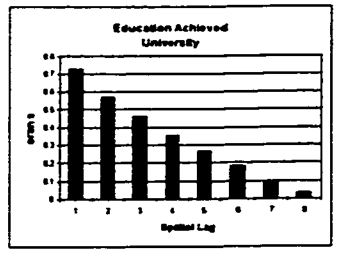


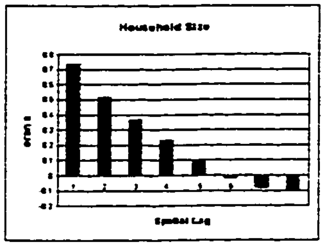
APPENDIX D Correlograms

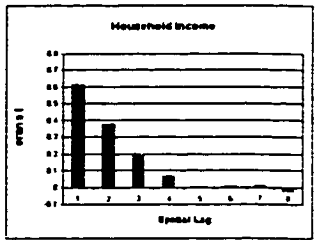
Social and Economic Variables

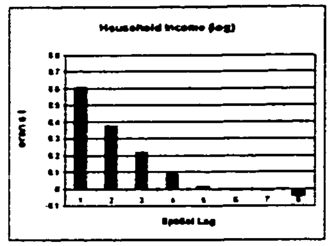


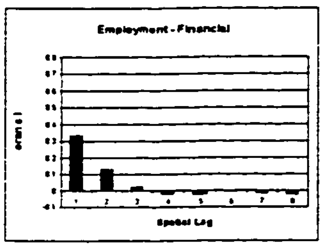


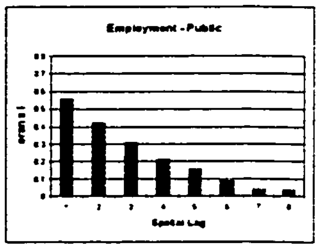


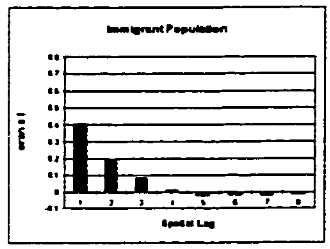


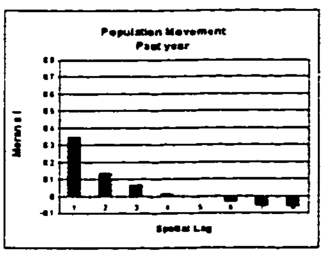


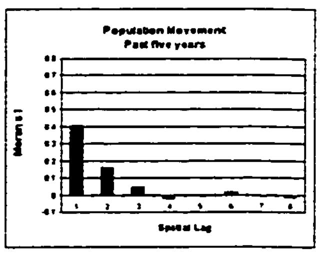


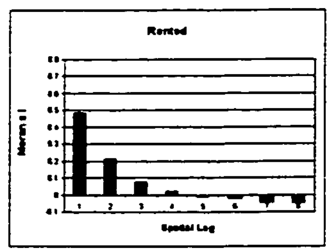


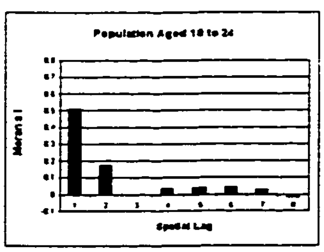


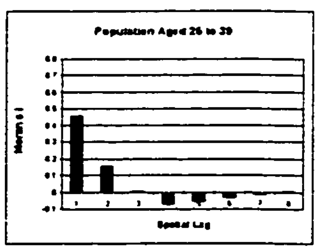


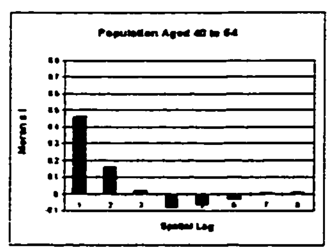


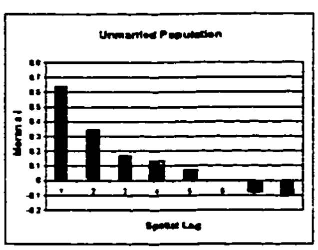


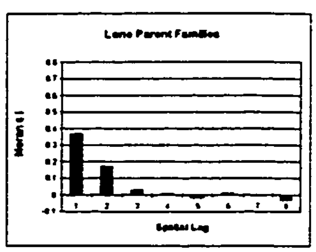


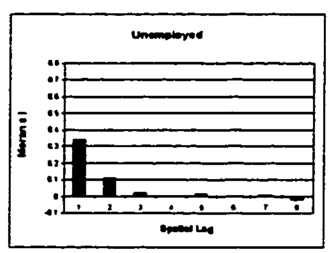






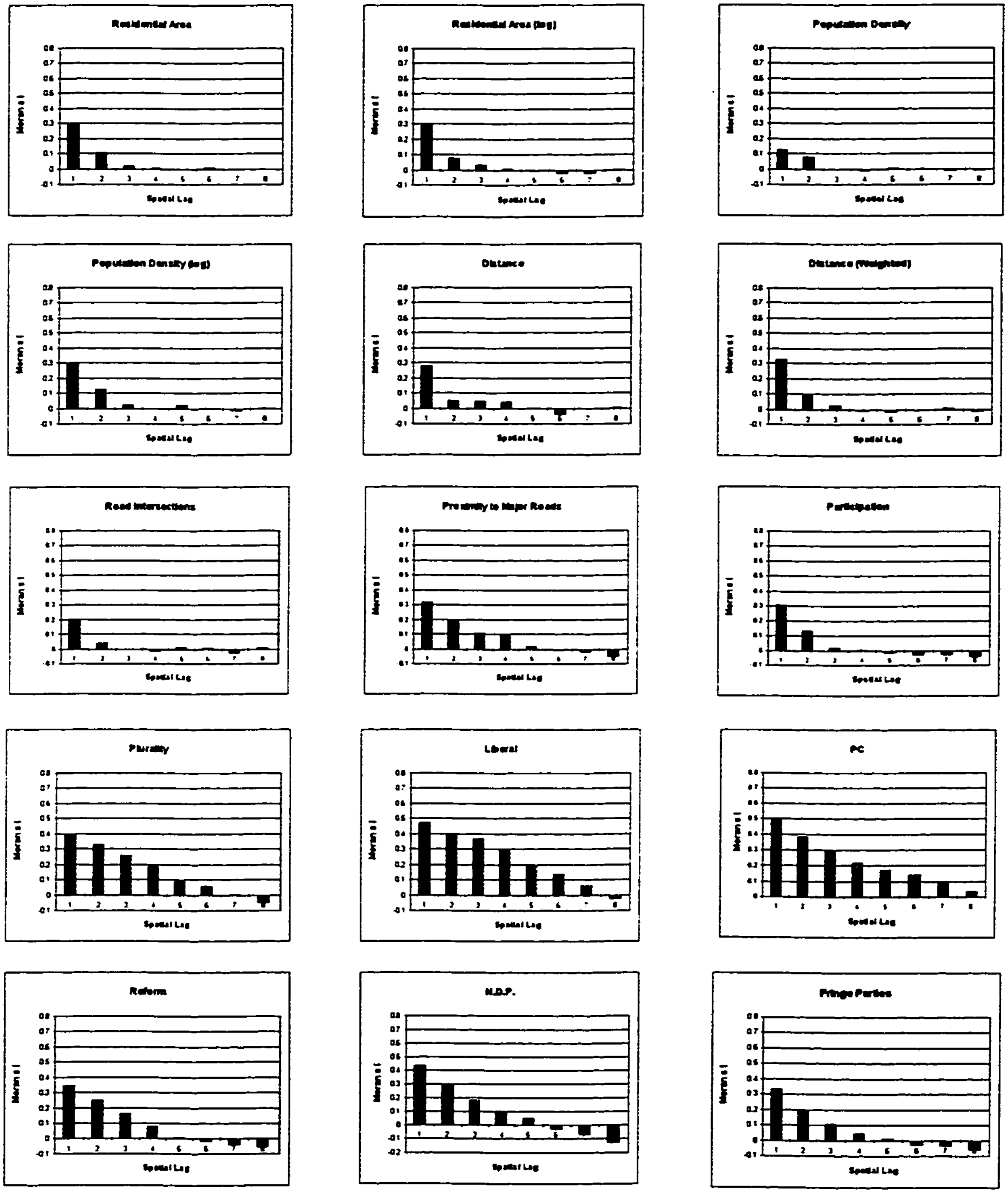






APPENDIX D (contd.) Correlograms

Geographic and Political Variables

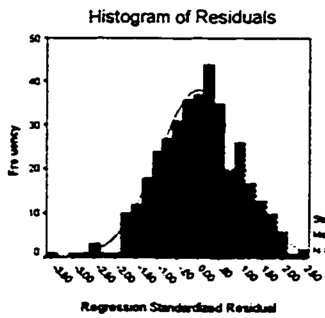


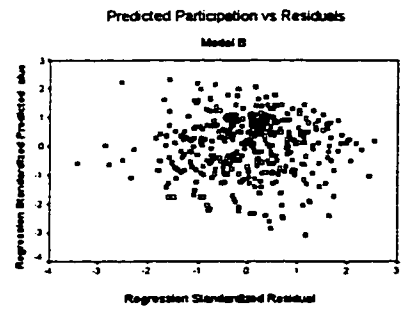
APPENDIX E Bivariate Regressions - Dependant Variable Participation

VARIABLE	Goodness of Fit				Coefficients						Spatial Autocorrelation	
	R^2	Std. Error	F Value	Prob. Sign.	Constant	B Coeff.	Std. Error	Std Beta	t value	Sign. Level	Moran's I	Sign. Level
Pop18_24	0.287	7.7128	151.275	.000	82.329	-1.621	0.132	-0.536	-12.299	.000	0.2218	.000
Pop25_39	0.227	8.0443	109.924	.000	83.356	-0.717	0.068	-0.476	-10.484	.000	0.2766	.000
Pop40_64	0.300	7.6553	160.463	.000	42.988	0.794	0.063	0.547	12.667	.000	0.2382	.000
Pop65ov	0.000	9.1473	0.031	.861	63.849	-1.03E-02	0.059	-0.009	-0.175	.861	0.3070	.000
Liberal	0.006	9.1194	2.332	.128	68.871	-0.111	0.073	0.339	-1.527	.000	0.3067	.000
P.C.	0.115	8.6050	48.788	.000	53.265	0.510	0.073	0.339	6.985	.000	0.2758	.000
Reform	0.014	9.0850	5.194	.023	58.472	0.215	0.094	0.117	2.279	.023	0.2957	.000
N.D.P.	0.236	7.9983	115.525	.000	71.614	-1.638	0.152	-0.485	-10.748	.000	0.1635	.000
Other	0.135	8.5070	58.607	.000	69.271	-1.821	0.238	-0.368	-7.656	.000	0.2288	.000
Majority	0.008	9.1127	2.880	.091	65.612	-8.52E-02	0.050	-0.087	-1.697	.091	0.3016	.000
Area	0.219	8.0846	105.107	.000	54.752	6.36E-05	0.000	0.468	10.252	.000	0.2721	.000
Area_log	0.250	7.9208	125.164	.000	-44.622	21.233	1.898	0.500	11.183	.000	0.2859	.000
Density	0.090	8.7262	37.096	.000	67.097	-6.56E-04	0.000	-0.300	-6.091	.000	0.2698	.000
Density_log	0.237	7.9919	116.304	.000	148.302	-23.144	2.146	-0.487	-10.784	.000	0.2269	.000
Distance	0.000	9.1474	0.019	.892	63.817	-1.37E-04	0.001	-0.007	-0.136	.892	0.3098	.000
Distance_mod	0.001	9.1418	0.484	.487	63.402	6.504E-04	0.001	0.036	0.696	.487	0.3041	.000
Ecentry	0.016	9.0754	5.994	.015	66.451	-0.242	0.099	-0.125	-2.448	.015	0.2929	.000
Proxroad	0.066	8.8385	26.685	.000	60.692	1.484E-02	0.003	0.258	5.168	.000	0.2581	.000
Edu_9	0.072	8.8131	29.017	.000	67.901	-0.529	0.098	-0.268	-5.387	.000	0.2829	.000
Edu_hs	0.027	9.0234	10.397	.001	71.711	-0.251	0.078	-0.164	-3.225	.001	0.2954	.000
Edu_univ	0.030	9.0105	11.502	.001	60.570	0.168	0.050	0.173	3.391	.001	0.3135	.000
Inc_hous	0.339	7.4368	192.388	.000	47.883	3.054E-04	0.000	0.582	13.870	.000	0.2175	.000
Inc_log	0.398	7.0954	248.298	.000	-139.155	43.224	2.743	0.631	15.757	.000	0.1914	.000
Job_fin	0.087	8.7387	35.923	.000	48.171	0.471	0.079	0.296	5.994	.000	0.2619	.000
Job_man	0.076	8.7951	30.663	.000	73.054	-0.288	0.052	-0.275	-5.537	.000	0.3032	.000
Job_pub	0.065	8.8447	26.129	.000	57.258	0.313	0.061	0.255	5.112	.000	0.3078	.000
Job_serv	0.065	8.8447	25.861	.000	71.310	-0.642	0.126	-0.254	-5.085	.000	0.2544	.000
Reli_cath	0.000	9.1473	0.027	.869	64.084	-1.03E-02	0.063	-0.009	-.166	.869	0.3094	.000
Reli_prot	0.091	8.7238	37.326	.000	47.877	0.322	0.053	0.301	6.109	.000	0.2849	.000
Reli_none	0.134	8.5125	58.052	.000	72.001	-0.690	0.091	-0.366	-7.619	.000	0.2027	.000
Rented	0.479	6.6056	344.157	.000	72.957	0.254	0.014	-0.692	-18.551	.000	0.1200	.000
Single	0.315	7.5725	172.230	.000	85.978	-0.960	0.073	-0.561	-13.124	.000	0.1885	.000
Unemployed	0.200	8.1813	93.818	.000	72.357	-1.741	0.180	-0.447	-9.686	.000	0.1807	.000
Move1	0.292	7.6978	154.561	.000	73.894	-0.619	0.050	-0.540	-12.432	.000	0.1873	.000
Move5	0.039	8.9664	15.317	.000	70.027	-0.131	0.033	-.198	-3.914	.000	0.3212	.000
Intersec	0.011	9.0958	4.288	.039	62.048	9.710E-02	0.047	0.106	2.071	.039	0.2990	.000
Parent	0.376	7.2245	226.224	.000	74.749	-3.254	0.216	-0.613	-15.041	.000	0.1669	.000
Household	0.156	8.4041	69.294	.000	45.037	6.797	0.817	0.395	8.324	.000	0.2642	.000
Immigrant	0.033	8.9933	12.986	.000	69.712	-0.271	0.075	-0.183	-3.604	.000	0.2899	.000

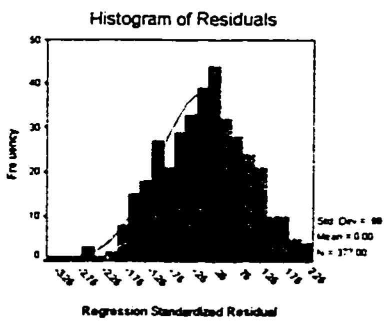
APPENDIX F RESIDUAL PLOTS

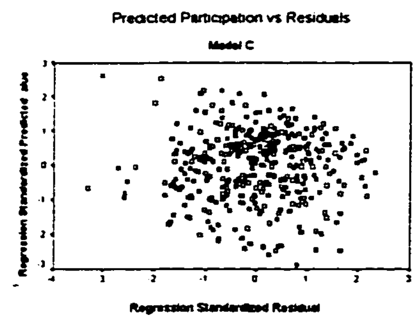
Residuals Model X1



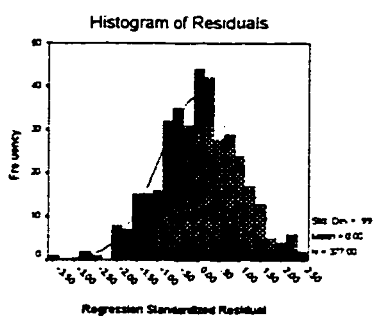


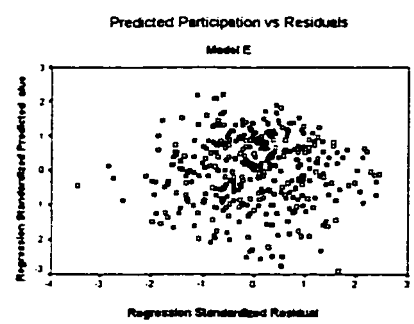
Residuals Model X2





Residuals Model X3





Residuals Model X4