1-1-2005

Demand modeling of successful park and ride planning: multivariate spatial regressive analysis

Hsin-I Yu
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Recommended Citation

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Demand modeling of successful park and ride planning: Multivariate spatial regressive analysis

by

Hsin-1 Yu

A thesis submitted to the graduate faculty in partial fulfillment of the requirements for the degree of
MASTER OF COMMUNITY AND REGIONAL PLANNING

Major: Community and Regional Planning

Program of Study Committee:
Riad G. Mahayni, Major Professor
David Plazak
Shauna Hallmark

Iowa State University
Ames, Iowa
2005

Copyright © Hsin-1 Yu, 2005. All rights reserved.
Graduate College
Iowa State University

This is to certify that the master's thesis of

Hsin-I Yu

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy
TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................ iv
LIST OF TABLES ........................................................................................................... vi
ACKNOWLEDGEMENT .............................................................................................. vii
ABSTRACT .................................................................................................................... viii
CHAPTER I: INTRODUCTION .................................................................................. 1
Statement of Problems .............................................................................................. 1
Background ................................................................................................................ 2
Objective .................................................................................................................... 3
General Layout and Methodology .............................................................................. 4
Research Organization .............................................................................................. 8
CHAPTER II: LITERATURE REVIEW ..................................................................... 9
General Studies of Park and Ride Planning............................................................... 9
Spatial Data Analysis ............................................................................................... 16
CHAPTER III: MODEL CONSTRUCTION ............................................................... 22
The Context of the Twin Cities Metropolitan Area ................................................... 23
Research Hypothesis ............................................................................................... 25
Sampling and Zonal Structure ................................................................................. 27
Data Aggregation and Centroids .............................................................................. 32
CHAPTER IV: MULTIVARIATE SPATIAL REGRESSIVE ANALYSES ............. 54
Multivariate Regression Analysis ............................................................................. 54
Examination Spatial Autocorrelation ..................................................................... 60
Spatial Autoregressive Analysis (SAR) ................................................................... 63
CHAPTER V: CONCLUSIONS AND RECOMMENDATIONS ..................... 68
Summary of Research ............................................................................................. 68
Model Application .................................................................................................. 69
Limitations and Recommendations ........................................................................ 70
REFERENCES ........................................................................................................... 72
LIST OF FIGURES

FIGURE 1.1 General Methodology ..............................................................6
FIGURE 3.1 Seven-County Twin Cities Metropolitan Area.............................23
FIGURE 3.2 Development of Geodatabase..................................................27
FIGURE 3.3 Park and Ride Capacity vs. Built Year......................................28
FIGURE 3.4 Transportation Analysis Zones.................................................30
FIGURE 3.5 Thiessen Polygons..................................................................31
FIGURE 3.6 Overlaying of TAZs and Thiessen Polygons...............................32
FIGURE 3.7 Estimated FY2004 Population by TAZs.....................................35
FIGURE 3.8 Estimated FY2004 Population by Thiessen Polygons..................35
FIGURE 3.9 Estimated FY2004 Number of Household by TAZs.....................36
FIGURE 3.10 Estimated FY2004 Number of Household by Thiessen Polygons...36
FIGURE 3.11 Estimated FY2004 Employment by TAZs..................................37
FIGURE 3.12 Estimated FY2004 Employment by Thiessen Polygons..............37
FIGURE 3.13 Estimated FY2004 Retail Employment by TAZs.........................38
FIGURE 3.14 Estimated FY2004 Retail Employment by Thiessen Polygons.......38
FIGURE 3.15 Estimated FY2004 Non-Retail Employment by TAZs..................39
FIGURE 3.16 Estimated FY2004 Non-Retail Employment by Thiessen Polygons..39
FIGURE 3.17 Median Household Income by Census Tracts............................42
FIGURE 3.18 Average Median Household Income by Thiessen Polygons............42
FIGURE 3.19 Number of Households without Vehicle by Census Tracts............43
FIGURE 3.20 Number of Households without Vehicle by Thiessen Polygons.....43
FIGURE 3.21 Number of non-white people by Thiessen Polygons....................44
FIGURE 3.22 Number of Families under Poverty Level by Thiessen Polygons.....44
FIGURE 3.23 Annual Average Daily Traffic (AADT) and Density Raster............46
FIGURE 3.24 Accumulated AADT Density by TAZs.................................46
FIGURE 3.25 Accumulated AADT Density by Thiessen Polygons...............47
FIGURE 3.26 Park and Ride Proximity to Centroids of Thiessen Polygons........48
FIGURE 3.27 Park and Ride Proximity to Major Shopping Centers..................49
FIGURE 3.28 Statistics of Park and Ride Proximity to Major Shopping Malls........49
FIGURE 3.29 Park and Ride Proximity to Express Route................................50
FIGURE 3.30 Park and Ride Proximity to Fixed Guideways..........................51
FIGURE 3.31 Comparison of Distribution Statistics for Proximity..................53
FIGURE 4.1 BoxMap Comparison of Park and Ride Capacity and Usage............56
FIGURE 4.2 Spatial Weight Matrix Property............................................60
FIGURE 4.3 Moran Scatterplot of Park and Ride Usage..................................61
FIGURE 4.4 LISA Cluster Map.................................................................62
FIGURE 4.5 LISA Significance Map............................................................63
FIGURE 4.6 Regression Window in Geoda..................................................64
FIGURE 4.7 Quantile Maps of Usage and OLS Predicted Usage Values............66
FIGURE 4.8 Standard Deviational Map of the OLS Residuals..........................66
FIGURE 4.9 Moran Scatterplot of Residuals...............................................67
LIST OF TABLES

TABLE 3.1 Survey of Very Important Transportation Programs ...............25
TABLE 3.2 Predictions of Demographic Explanatory Data by TAZs ..........34
TABLE 3.3 Summary of Demographic Explanatory Data .......................34
TABLE 3.4 Summary of Economic Explanatory Data .........................41
TABLE 3.5 Summary of Traffic Explanatory Data ...............................45
TABLE 3.6 Summary of Proximity Explanatory Data .........................51
TABLE 3.7 Summary of Transformed Proximity Explanatory Data ..........53
TABLE 3.8 Summary of Transformed Proximity Explanatory Data ..........53
TABLE 4.1 Summary of Potential Independent Variables ..................54
TABLE 4.2 Pearson Correlation Matrix of All Potential Independent Variables .....55
ACKNOWLEDGEMENT

A journey is easier when you travel together. This thesis is the result of two years of work whereby I would like to express my sincere gratitude and appreciation to many people who made this master thesis possible.

I would like to first thank my major supervisor Riad Mahayni and former major advisor Ivan Suen. I have known both of them as sympathetic persons since 2003. Riad’s enthusiasm and integral view on research and his mission only provide high quality of work. Ivan could not even realize how much I have learned from him. They were always available when I needed their advices. They were as close as relatives and good friends to me. I am really thankful to them for keeping an eye on the progress of my work. I have furthermore to thank Monica Haddad who brought me a broader window for GIS. I want to thank them for all their help, support, interest and valuable hints because I am glad that I have come to get to know them in my life.

In my opinion, doing a master oversea is a sacred task and this was definitely one of the best decisions of my life. Additional energy and vitality for this research was provided externally through my involvement in several social activities. Special thanks go to my dearest friends Amy, Dale, Eric, Eva, Jonathan, Mike, Milk, Nora and Vcilia who shared many experiences and thoughts with me during my stay in Ames.

Specially, I feel a deep sense of gratitude for to my grandmother, parents, brothers, and all my families who formed part of my vision and taught me the good things that really matter in life and whose patient love enabled me to complete this work. Many more persons participated in various ways to ensure my research succeeded. I thank you all.
ABSTRACT

Park and ride facilities are designed to efficiently intercept traffic flow toward metropolitan business districts and help relieve traffic congestion in the central business areas. An attempt in this research was made to develop a successful park and ride demand model based on the distribution of park and ride usage, by applying geographic information system (GIS) and other spatial statistical packages.

The Minneapolis-St. Paul (Twin Cities) Metropolitan Area was selected as a study area for this research because its park and ride system has grown to become one of the nation’s largest systems in terms of the number of facilities and total capacity. Recently, the Twin Cities Metropolitan Area started to consider designating large-scale park and ride faculties in the region. There is a need to conduct a research for achieving successful park and ride planning.

This research involves multivariate regression demand forecast, spatial cluster identification, and spatial autoregressive analyses. Factors considered in the model for assessing the success and failure of park and ride facilities include socioeconomic characteristics, transportation network features, user behavior, and spatial statistics. These factors are analyzed for predicting the park and ride usages, i.e., the number of parked lots in the park and ride sites during the last visit in FY 2004.

In general, the contribution of this research for successful park and ride facilities demand forecast is to integrate spatial association with quantitative statistics by a series of GIS-based statistical techniques for future practice in the field of transportation facility planning.
CHAPTER I: INTRODUCTION

Statement of Problems

Travel demand in the urban areas has continued to grow since the mid-1970s, with increasing reliance on private automobiles. According to the Transportation Statistics Annual Report from the U.S. Department of Transportation Bureau of Transportation Statistics (2001:69), annual vehicle miles of travel (VMT) in the United States rose by nearly 30 percent to 2.8 trillion miles between 1990 and 2000, an annual increase of 3 percent. VMT per capita rose by just over 13 percent during the same period, an annual increase of 1.3 percent. With significant growth in both population and individual travel, highway usage has resulted in inefficient travel patterns and overall increases in regional vehicle hours of travel (VHT).

Frequently, additional congestion on freeways and arterials, resulting in increased travel times and pressure on local and regional arterials, is a byproduct of high automobile dependency in a restricted urban environment. Besides, air quality and low transit/high occupied vehicle (HOV) use are common concerns in some regions where there is rapid regional growth with low density development patterns. Hence, a question concerning how to incorporate the mass transit to reduce traffic and to face the fact that high vehicle dependency has become definitely critical in this era.

In theory, park and ride facilities can help to relieve congestion because of their ability to move travelers efficiently and cost-effectively between home, work, and various activities. This is done through shifting them from the private auto to transit and carpool modes. However, there is lack of literature to verify the impact of park and ride planning due to the hardship for certain reasons. Although we might believe and assume that park and ride planning is one possible solution to discourage the trip generation of personal vehicles and to meet the heavier transportation demand, it is
still very tricky to offer conclusions about the level of impact of park and ride facilities on transportation effectiveness.

Existing highways and transit systems are struggling to move people and goods efficiently and to meet the ever increasing travel demand. A successful park and ride demand modeling process will assist in reducing these pressures and in serving as one additional cost-effective alternative of transport system management (TSM).

This research is dedicated to people who are interested in transportation facility planning, and to present an innovative approach for successful planning for park and ride facilities. It is involved in determining significant factors affecting planning for park and ride facilities by applying GIS technology and spatial statistical analyses.

**Background**

Park and ride facilities, as an integral component of the public transportation system, serve as an intermodal alternative to encourage commuters to first use private transportation and then utilize public transportation network. As many cities might have current or projected traffic congestion, park and ride has emerged as a response to increasing global oil prices, relieving congestion, and to increasing interest in mass transit. These facilities are located either close to major activity centers, served by local bus routes, or relatively far from major activity centers, served by express transit service.

Spillar (1997:49) indicated that interest in the park and ride mode expanded rapidly during the oil crisis of the 1970's. A number of transportation agencies sought ways to make carpooling and transit service more convenient for suburban and rural residents working in central cities and major employment centers. However, according to Lamothe (2001), a park and ride planning coordinator from the Minnesota Metro Council, only few new techniques for estimating park and ride
demand have been developed since the mid-1980’s. Initial approaches to park and ride facilities were mostly in the form of site characteristic checklists for evaluating potential or available building sites without estimating facility demand. As time went by, transport management agencies have now completed or are in the process of using their park and ride planning as a strategy for increasing transit ridership and managing growth in many different regions (Lamothe, 2001). Agencies have two primary approaches for developing improved park and ride demand forecasts: (1) to estimate individual park and ride demand based on regional modeling approaches and (2) to develop site-specific forecasting tools, and utilize basic rules of thumb to design individual facilities at a conceptual level.

Many cities throughout the United States, such as Chicago, San Francisco, Los Angeles, San Diego, Washington, D.C., Portland, Seattle, and others, have built a number of new park and ride facilities to promote transit friendly communities (Puget Sound Regional Council, 1999:2). Also, on a global scale, nations with limited territory but highly developed land usage, including the United Kingdom, Taiwan, Singapore, Hong Kong, Japan, and others, are moving toward encompassing the concept of park and ride to serve their metropolitan population (Gilbert and Ginn, 2001).

**Objective**

The demand for travel is growing dramatically in many metropolitan areas. As a consequence, the interest in park and ride planning is correspondingly increasing. To quantify the effect of various park and ride policy decisions and to model “what if” scenarios would be welcomed by many public transportation agencies. Park and ride planning is not only a local problem but also a national issue.

Park and ride facilities have been utilized as a part of transportation demand
management in many urbanized areas, especially in compact places with rapid urban growth that lack road capacity. But, park and ride planning usually differ in the places where they are located and the ways how local agencies plan for these facilities.

Every geographic place has its own identity, network features, and unique characteristics so that we need to think of a way of taking into account such realities when formulating practical plans for each place. Unfortunately, there is no absolute correct answer for all different places. Some solutions which may be useful in the Midwest might not be suitable in the West Coast. One of the purposes of this research here is to formulate a universal approach for improving current park and ride facility planning through the use of demand estimation modeling.

The essential objective of this research is to propose a park and ride demand modeling method, and to test the research hypothesis of that whether the successful park and ride facilities can be predicted. A study area, the Minneapolis-St. Paul (Twin Cities) Metropolitan Area, has been selected to explore the question of “why some park and ride facilities are more successful than others in terms of intercepting more cars toward downtown areas (park and ride usage)”.

The goal of this proposed approach in the research is to offer recommendations for policy-sensitive park and ride demand forecast modeling, and to address how such an approach may assist facility planning for travel forecasters, transportation planners, and decision makers.

**General Layout and Methodology** This research has applied a series of computer technologies to solve the spatial related statistical issues, such as geographic information system (GIS), Geoda, SPSS, and JMP. Although these software and databases have been developed and used for some years all over the world, there are only a few professional applications that have integrated their use for
more comprehensive planning analyses for park and ride facilities. In fact, the well-organized uses of these combinations can bring many advantages that help conduct more efficient analyses and make more accurate decisions.

The proposed methodology for estimating demand for park and ride facilities involves a transportation demand model using Geographic Information System (GIS) and Geoda to make spatially overlapping analyses, defining service area, building centroids, data aggregation and disaggregation, calculating distances, and to examine spatial clusters of park and ride usage in the Twin Cities Metropolitan Area. Besides, SPSS and JMP are the tools for seeking a statistically significant model while taking into account the correlation between multivariate independent variables. Moreover, Geoda is adopted again in the end for testing the spatial dependence of park and ride actual usage and predicted usage, in order to re-verify the reliability of the park and ride usage model while including space dimension.

FIGURE 1.1 diagrams the general layout of the methodology proposed for park and ride demand modeling. The methodology is applied to the Minneapolis-St. Paul (Twin Cities) Metropolitan Area. Unlike traditional regression analyses, spatial autocorrelation has played a critical role to identify the clusters in terms of the park and ride facility usage. This research begins with developing geodatabase in three categories: (1) Transportation (park and ride lots, major highways, functional class roads, bus service and routes, bus stops, fixed guideways, and transit centers), (2) Planning and Development (FY2000 counties, FY2000 census tracts, and FY2000 transportation analysis zones), and (3) Demographics and Business (general demographic characteristics, selected economic characteristics, and major shopping centers). To accomplish spatial analyses for park and ride demand model, the functions in GIS are applied to simulate all types of phenomena happening in the space, and referenced spatially and quantitatively based on their conditions and
relative locations.

**FIGURE 1.1 General Methodology**

Transportation Network i.e. highway, functional roadway, fixed guideway, transit service, park and ride facility, aadt...  
Planning and Development i.e. counties, transportation analysis zones, census tracts,...  
Demographics and Business i.e. population, income, household size, vehicle ownership, retail, major shopping centers...

Development of Geodatabase and GIS Basemaps

- Zonal Structure and Data Aggregation
- Classic Regression Analyses
- Spatial Autocorrelation
  - Global Moran’s I
  - LISA Cluster
- Examine Spatial Autocorrelation
- Spatial Autoregression
- Spatial Dependence Test for Residuals
- Model Calibration

Park and Ride Planning Demand Model

Next, to testify the spatial autocorrelation of the service areas of park and ride facilities in the Twin Cities Metropolitan Area, defining the zonal structure for analyzing transportation demand and all active park and ride facilities is essential. Based on the data derived from Minnesota Metro Council and Minnesota Department of Transportation, there are 1201 transportation analysis zones (TAZs) and 139 active park and ride facilities in FY2004, which are selected for the demand estimation.
Each of the TAZs has its population, employment, land use acreage, and related attributes. Each of park and ride facilities has basic information such as established year, facility name and location, on-site capacity, the number of spots that were used in a recent site visit, and others. By GIS, data are integrated into 139 Thiessen Polygons for modeling.

After proposing potential independent variables, classic multivariate regression analyses are adopted to construct a linear regression model for predicting park and ride usage in the Twin Cities Metropolitan Area. Knowing the coefficient correlation between pairs of independent variables, the results from stepwise regression model, which is utilized for choosing statistically significant independent variables and best-fit model, are reviewed for predicting the park and ride usage.

In addition, to verify the model rigorously, spatial dependence of park and ride usage is tested in Geoda to see the global and local clusters. The results from stepwise regression model are incorporated with the results of spatial dependence from Geoda. If there is a positive (or negative) spatial autocorrelation, it means the distribution of park and ride usages are clustered and have similarity (or dissimilarity) with their defined neighbors. Conversely, no spatial autocorrelation means the phenomenon in the area is isolated from each other and there is no need to include spatial factor in the park and ride demand forecasting model.

Finally spatial autoregressive analysis is applied to test the accuracy of the prediction model while taking into account the spatial dimension simultaneously. A close examination of residuals spatial dependence, a more comprehensive park and ride forecasting model can be reached by the proposed methodology and the findings and critiques can provide for reviewing current and future park and ride planning scheme.
Research Organization

The remaining chapters are organized in the following manner: Literature Review, Model Construction, Case Study Model Analyses, and Conclusions and Recommendations.

The first section of Chapter II offers a review of the literature relevant to the park and ride studies in locating and sizing facilities. Then, the spatial statistics for geographical data analyses, and the revolution of transportation network modeling techniques are summarized to assist the model development of this thesis.

Chapter III: starts with the statement of research hypothesis, the context of study area – the Twin Cities Metropolitan Area, and park and ride data sampling for modeling. Then, this research also proposed a way to define the service areas for each park and ride facility, and aggregate and disaggregate the census data for regression analyses.

Chapter IV continues to compute the measures in each park and ride service area as exploratory independent variables. Through a process of multivariate regression analysis, knowing the correlation of variables, the results of stepwise fit are adopted and incorporated with the results of spatial dependence by using Geoda - a spatial statistical tool. The tests of the spatial autocorrelation help to re-verify the reliability of the park and ride usage prediction model.

Chapter V contains the conclusions, the limitations of the research, as well as the recommendations for further research.
CHAPTER II: LITERATURE REVIEW

The current literature on park and ride facilities can be divided into two types. The first type is general studies reviewing the factors making park and ride facilities feasible and successful, the incentives to park and ride users, and the past experience in developing park and ride facilities. In addition, the literature reviews some basics of multivariate spatial statistical analysis and documents various up-to-date geodata analysis applications such as Geographic Information System (GIS) and GeoDa.

General Studies of Park and Ride Planning

Park and ride is one transport-planning tool that can be used to encourage car users to switch to public transport (Traffic Advisory Unit – TAU, 2004). Associated with other traffic management measures, park and ride facilities and local public bus service complement each other, and when present together, serve as a traffic reduction implement in the central area. A well-designed and well-located park and ride facility can assist in reducing traffic levels in the town center. It can provide more sustainable access, improve attractiveness, and enhance the economic viability of a town center. Therefore, knowing that park and ride facilities are likely to help form a positive image of public transport, it is necessary to further re-examine them as one of the possible elements in a local transport strategy and consider their local circumstances. To be able to support park and ride facility planning, we start the literature review with park and ride planning history and trace its evolution through integration with the increasing use of other mainstream public transit services.

As indicated by the senior park and ride planner Robert Spillar (1997:49), interest in the park and ride mode expanded rapidly during the escalating oil and gas crisis of the 1970s, and park and ride demand estimation peaked in the late 1970s...
and early 1980s (Spillar, 1997:49). Initial approaches to park and ride facility planning were largely based on practical knowledge of the proposed service area and often depended upon available building sites (Spillar, 1997:49).

As Spillar argues, two primary approaches for developing improved demand forecasts are evolving in this industry: estimating individual park and ride demand based on regional modeling approaches and developing site-specific forecasting tools, tailored to the metropolitan region (Spillar, 1997:50). Each demand forecasting technique can generate erroneous forecast estimates if not applied with an extensive knowledge of the study area.

Spillar emphasizes that the use of a regional travel demand modeling approach for estimating park and ride demand must be viewed within the context of the overall transportation modeling technique (1997:51). This argument for park and ride demand forecasting is heavily dependent on the modal choice model within the overall transportation modeling structure. Post-modeling regional forecasting techniques for individual park and ride facilities suggested by Spillar closely follow the traditional transportation modeling methodology. That is, trip productions and attractions are first defined, followed by trip distribution, assignment and modal split. It begins with identifying the production (home) and attraction (employment) ends of potential trips that might use a proposed park and ride facility, distributing the trip between the two influence areas by a regional travel demand model, estimating the proportion of each trip interchange, applying interchange tabulations, and computing the required number of parking spaces (Spillar, 1997:51-53).

On the other hand, the argument for site-specific forecasting methodologies was an attempt by Spillar to estimate park and ride demand based on the attributes of the proposed park and ride location (1997:56). The forecasting methodologies generally revolve around defining a given service or market area for a number of
individual park and ride facilities, followed by explanatory equations through the use of a multivariate regression process. These models often define the attractiveness of the site to potential users by focusing on the attributes of the specific lot and the traffic on adjacent streets. His assumption was that a modal split or preference for one mode over another can be implicitly determined either by measuring the differences in service attributes between competing modes or through travel surveys (1997:57).

For this reason, defining park and ride service area becomes one vital element of all site-specific demand estimation procedures, which frequently include complex and strict socioeconomic data on the people living within the market shed. A study found that the average market shed for park and ride lots is typically more dispersed than for the suburban park and ride, and is more closely described by “concentric demand contours” (North Central Texas Council of Governments, 1979). Therefore, how facility planners define the service areas or market shed is most critical, as interest in estimating park and ride demand has been expressed by many agencies.

In summary, these site-specific forecasting models did provide some explanation of park and ride demand although they pay no attention to the regional transportation system, land uses, and other geographical location issues. Spillar does not intend to discredit the regional modeling approach but rather suggests that it be used to develop planning-level regional or corridor estimates. We must be aware of these invaluable limitations and merits of both approaches.

Furthermore, according to “A Comprehensive Planning and Design Manual for Park and Ride Facilities” (Spillar, 1997:19), it is typically more effective to plan park and ride facilities as part of a coordinated transportation system than to plan individual facilities and try to tie these facilities together after the fact. This is because park and ride facilities cannot function on their own without direct linkages to the
surrounding transit and highway infrastructure. As a result, a recent trend in many U.S. cities is to develop comprehensive system plans inclusive of park and ride facilities before developing the individual elements within that overall plan.

Success of the individual park and ride facility lies in its ability to connect with the regional transportation network and its spatial location within that network (Spillar, 1997:19). Therefore, it is extremely important to coordinate with regional or local transit agencies in the transportation planning process. The transit agency must be able to provide service to each individual park and ride lot if the lot is to serve as a transfer point between auto and transit modes.

One proposed park and ride design requirement is to make it an integral part of the surrounding community (Spillar, 1997:85). In fact, early park and ride facilities were often designed and built by the transit agencies operating service to the lot (Spillar, 1997:86). They were often associated with fixed guideway transit modes or similar capital-intensive modes such as commuter train, interurban trolley, or intercity ferry. As such, early park and ride facilities were often built as components of an intermodal station. By the late 1960s and early 1970s, park and ride development began to be largely associated with bus transit and the expanding system of interstate highways being constructed at the time (Spillar, 1997:87). Primary objectives and advantages of a community-compatible or integrated park and ride facility include: supporting the services and security of the transit agency operating the park and ride facility by adjacent residential, service-oriented and commercial activities; increasing visual perceived safety by multi-story buildings located near the site; encouraging a multimodal use of the lot if there is adequate attention to pedestrian and bicycle facilities, both on-site and in the surrounding developments; and serving as a focal point for suburban community development with public investment in an integrated transit facility.
If park and ride facilities are planned without the participation and commitment of the local transit agency, they would only be parking lots, not a public transit component. Also, it is preferable to address facility planning prior to site selection. In most of early park and ride planning, individual sites were proposed by developers and agencies without making efforts on regional demand forecast.

As listed in “MAG Park and Ride Site Selection Study” (KJS Associates, 2001:5), characteristics of successful park and ride lots include the following:

- High level of express bus service (every 15 minutes or less during peak hours);
- Location within close proximity of a freeway or light rail line (1 mile or less);
- Access to HOV lanes for at least a portion of the bus trip to the final destination;
- Express transit service available over at least a three hour period in morning and evening peak periods;
- Visibility from adjacent arterials (to facilitate marketing and patron safety);
- Substantially lower bus fares than parking costs at the destinations served by lot.

The sites were identified based on the following factors (KJS Associates, 2001, 2-40):

- Available land/capacity and potential for expansion;
- Opportunities for joint use;
- Availability of express transit service;
- Vehicle access;
- Freeway proximity and access to HOV lanes and ramps;
- Location relative to freeway congestion.

It is very common to see a list of criteria for park and ride site selection and facility maintenance. Some park and ride planning has applied this inventory-like method for implementation; however, how precise should these lists be? Perhaps they should also include a map of the area, showing the destination, major roads,
nearby landmarks, the closest rail station or bus stops, and recommended cycling and walking routes; information about transit service frequency, fares, first and last runs, and public transportation schedules; estimated travel time from nearby cities and towns along with different trip purposes? The success of park and ride planning based on this oversimplified method highly depends on the planners' experience, common sense, local knowledge, and project budget.

Additionally, modal split is an indispensable element of the transportation demand modeling process. A multinomial logit model was developed recently by Hess for Portland's commercial business district in 2001 to evaluate the probabilities that commuters who do and who do not receive free parking at work will choose to drive alone. The modal choice model predicts that a daily parking charge of $6 in the Portland CBD would result in 21 fewer cars driven for every 100 commuters (Hess, 2001:2).

The model analyzed in Portland had two key findings. First, parking cost and the transit travel time plays a part in mode choice decisions for commuters (Hess, 2001:19). This suggests that raising the cost of parking at work sites and decreasing the transit travel time (by providing service and decreasing headways) will reduce the percentage of people who drive alone to work. Variables such as income and vehicles per capita have an effect on mode choice, but whether the commuter is male or female is unimportant. Secondly, two land use variables in the modal choice model were found to be unimportant (Hess, 2001:19). In particular, neither the proximity of the commuter's residence to a light rail station nor the "pedestrian connectivity" of the streets and sidewalks surrounding the commuter's residence has a significant effect on mode choice. This finding supports the contention that urban form has little impact on mode choice decisions.

Another eminent park and ride researcher, Graham Parkhurst, has worked on a
series of park and ride studies since 1996. He indicates that park and ride was initially introduced into cities which combined two conditions: physical constraint due to the density of the urban fabric and hence the least opportunity for road and car park building, and strong gravitation of trip exerted by a range of rare or unique services and opportunities located in the central business districts (Parkhurst, 1996:4). Actually, there were a lot of park and ride studies previously proposed by many researchers and state agencies for different aspects. The variety regarding park and ride facilities involve individual site design and maintenance blueprints, environmental cost-benefits evaluation, economic and modal split impact assessment, and optional location selection from proposed candidate sites. However, compared to these types of studies, park and ride demand forecast modeling is still in the beginning of its evolution. Much of the assigning of trips to individual park and ride facilities is accomplished outside of a transportation modeling procedure, using outputs from a regional model as the basis for trip estimation. This is caused by the inevitable issue of the uncertainty of park and ride impact on traffic reduction.

As identified by Parkhurst and Stokes (1999:2), problems with the policy of park and ride demonstrated some "unintended effects" in English cities:

- Abstraction from modes other than car: not all park and ride users drove cars to the city centers prior to the provision of the facilities because park and ride lowered the generalized cost of travel, so a proportion of users had switched mode from public transport services (Parkhurst, 1999:2).

- Lack of evidence for decongestion: it was not possible to demonstrate that park and ride resulted in a net reduction in urban congestion "downstream" of the sites. The possible implication was that suppressed demand had re-filled the road space made available by car trips being intercepted at park and rides (Parkhurst, 1999:2).
• Trip generation: some extra journeys were made to the city centre via park and ride sites because providing the schemes lowered the cost of travel (Parkhurst, 1999:2).

• Increased car dependence: the previous factors added up to an overall increase in car travel, rather than a reduction. In addition, there was an increased feeling that investment in park and ride sites and dedicated bus services represented an investment in facilities for car users instead of an attempt to address the problems of declining public transport use head-on. By replacing the penalties of car use within the city with subsidies to park at the edge, park and ride might encourage car use in the city’s hinterland. In other words, it might contribute to residential dispersion and growing car dependence (Parkhurst, 1999:2).

A key message of English Historic Towns Forum in 1998 was “park and ride is capable of making a contribution to traffic reduction, but only as part of a suitable package of restraint measures” (Parkhurst, 2000:161). The debate concerning the influence of park and ride facilities on traffic congestion are political and technical. Whether the practical role of park and ride policies is to direct traffic restraint or to support car use needs to be testified by both persuasive “demand” forecast and empirical “evidence”.

**Spatial Data Analysis**

Fundamental to the operation of GIS are spatial data. Although geographers have been using spatial data long before the mid-1980s, there has been a marked diffusion of interest in spatial data handling and the problems associated with such data (Fotheringham, 2000:15). Spatial data comprise of observations, sampled from the real world, of some phenomena that possesses a spatial reference, which may be explicit or implicit. The nature of the sampling is not only the variation of some
phenomena, but also the location of that variation. In general, sampling is either
discrete entities (houses, roads, administrative units) in the form of points, lines, or
areas, or some phenomenon that varies continuously (air pressure, elevation,
population density) in forms such as a set of observations taken at regular intervals
on a grid or lattice. (Fotheringham, 2000:17). Conceptually, we refer to the former as
objects, and the latter as continuous fields.

Spatial data means different things to different users for different purposes. A
commonly used characteristic of spatial data is the distance. It is the measure of the
distance between two points on a plane. So-called Euclidean or straight-line distance
is computed from the coordinates of two locations. If \( p1=(x_1, x_2) \) and \( p2=(x_2, y_2) \) are
two points on the plane, their *Euclidean distance* is given by:

\[
\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}
\]

However, the Euclidean distance may not always be the most meaningful
measure. This could be overcome by what is known as the Minkowski metrics:

\[
\left[ \sum |Z_i|^p \right]^{1/p}
\]

where \( p \) is a constant that can have any value from unity to infinity. When \( p=1 \),
the distance is referred to as the Manhattan, city-block, or taxicab distance. The
city-block distance is greater than the Euclidean distance, but is not always less than
the route distance.

Another related operation in several GIS packages is to create the boundaries
of service areas or influence areas. As radius buffering is one of the commonest
methods, the creation of Thiessen (Voronoi) Polygons is also very useful in practices,
which is from the proximity computation of irregularly spaced point data
(Fotheringham, 2000:38-40). The boundaries of Thiessen Polygons are created such
that any location inside a polygon is closer to that polygon’s centroid than any other
centroid outside that polygon.
Although we realize “location” matters, the analysis of spatial data is not just running any computerized software without awareness of the geographical solecisms. For example, a foundational assumption of classical statistical inference is that of independence (Fotheringham, 2000:26). In other words, before running the program, there is a need to make sure all observations are unrelated to each other. However, as a matter of fact, “everything is related to everything else, but near things are more related than distant things” (Tobler, 1970). The concept of Tobler’s (1970) “first law of geography” is very similar to Odland’s explanation for spatial interaction: the movement of goods, people, or information over space means that events or circumstances at one place can affect conditions at other places if the places interact (1998:13). These movements or interactions among places usually vary with distance in systematic ways, namely positive autocorrelation, where similar values are clustered, and negative autocorrelation, where similar values are dispersed (Odland, 1998:24).

Therefore, the assumption of independence of the observations is questionable with spatial data. Indeed, the problems of ignoring spatial autocorrelation have been demonstrated by a number of geographers and statisticians. Ignoring the spatial dependence of observations may result in over- or under-estimation and misinterpretation of original datasets. If we assume that our observations are independent, then we assume zero spatial autocorrelation. If our data are positively spatially autocorrelated, then the standard error of the mean will be greater than if we
had assume independence. Conversely, if our data exhibit negative spatial autocorrelation, the standard error of the mean will be less than if we assume independence.

Goodchild (1986:5-6) has summarized the practical importance of testing spatial autocorrelation. First, as an index, it provides a type of information about a spatially distributed phenomenon; second, it might help to identify the variables accounting for partial and remaining variations; and third, as a measure of the process by which one place influences another, spatial autocorrelation analysis is often a necessary part of correct forecasting. One of the most famous alternatives for testing spatial autocorrelation, concerning similarity among attributes and similarity of locations, is Moran's index. The Moran index is positive when nearby areas tend to be similar in attributes, negative when they tend to be more dissimilar than one might expect, and approximately zero when attribute values are arranged randomly and independently in space (Goodchild, 1986:16).

Also, the selection of a spatial weighting function is the most important component in testing spatial correlation. A spatial weighting function is a set of rules that assign values or “weights” to every pair of locations in a study area, where the values of an autocorrelation statistic will depend on these weights as well as the data for the locations (Odland, 1998:29). Different definitions and methods of defining “neighbors” would result in either slight or considerable differences in the consequences. It is always wise to explore the dependence of the spatial data set prior to making statistical analyses or predicting interpolations.

For these reasons, rethinking the role of Geographic Information System (GIS) in the quantitative analysis of spatial data is quite important. The term “geographic information system” is not merely a piece of software for storing, querying, integrating, retrieving, displaying and modeling spatial data, but it is also intended to provide
quantitative locational analysis with a variety of statistical packages, such as SAS or SPSS. Likewise, in Xinhao Wang and Bruce Stauffer’s paper (2004:4), they point out that the most essential benefit of GIS is the development of a spatial data system for policy makers, planners and citizens to better understand the impact of proposed development. It is comprehensible because of its integrated system, simulation modeling, and computer visualization capabilities.

Furthermore, GIS is capable of examining spatial autocorrelation, that is, whether a variable exhibits a regular pattern over space in which its values at a set of locations depend on values of the same variable at other locations (Odland, 1998:7). Each technology contributes distinctive features to the system and provides the functions that allow users to examine the spatial relationships among entities. For example, the function of simulation modeling is capable of representing the dynamic relationships between cause and effect; the strength of visualization is to represent data in a way that may reveal patterns and relationships that are hard to detect by non-visual approaches such as texts and tables.

Moreover, a former study on multivariate statistical analysis in geography indicates that statistical analysis is far superior to any other research method when the aim is to make precise and unambiguous statements about relationships and patterns in sets of numbers (Johnston, 1978). Two types of procedure outlines are oriented to the two fundamental geographical questions (Johnston, 1978:1-9). The first question is whether there are relationships between phenomena in various locations. Assuming that both cause and effect are measured on either an interval or a ratio scale, it can be answered by the methods of correlation and regression, and of factor, principle components, and canonical correlation analysis. The second question, whether places are different in terms of the phenomena present there, can be gained through analysis of variance and discriminate analysis procedures when
assuming the causal variable is nominal and effect variable is interval or ratio.

As a comprehensive computerized tool, GIS helps us deeply understand the meaning of spatial information and how that information can more faithfully reflect the true nature of spatially distributed processes. Most essentially, GIS puts spatial statistics into action to measure the dependence among nearby values in a spatial distribution, test hypotheses about geographically distributed variables, and develop statistical models of spatial patterns.
CHAPTER III: MODEL CONSTRUCTION

As dependency on the automobile increases, parking becomes a critical problem, especially in compact cities with limited land. Most transportation developments require a great deal of additional space and construction is usually costly in time and in dollars. Even though new transportation developments provide for balancing the needs of travel, the issue of parking is frequently inevitable unless it has been well controlled in advance.

The cost of parking increases as the availability of developable land for parking decreases. There is a constant and intense competition between transportation and other land uses. However, a sustainable city must offer its population a suitable urban environment, employment, food, housing, and transportation without compromising the welfare of future population.

In order to effectively reduce the adverse impact of current transportation systems, it is essential to plan sustainable transportation systems for future and existing transportation developments as well as land use patterns. Park and ride facilities, offering one additional cost-efficient transport solution, have become more popular during the last few years. These facilities provide a car park, where people can transfer from car to bus or train. Travelers are encouraged to adopt mass transportation and the amount of commuter traffic is reduced to enter urban centers.

The goal of this chapter is to evaluate the implementation of park and ride facility planning as a strategy for sustainable transport management. Accordingly, the Minneapolis-St. Paul (Twin Cities) Metropolitan Area is selected as the study area and the introduction to its context is followed by the research hypothesis. Applied to the study area, the research methodology examines the spatial clusters and simulates a park and ride demand model by multivariate regression analyses, based on its context and usage (the number of spots used in the facilities).
The Context of the Twin Cities Metropolitan Area

The proposed methodology is applied to the Minneapolis-St. Paul (Twin Cities) Metropolitan Area (see FIGURE 3.1) as a case study area. The park and ride system in this region has grown steadily from its inception in the early 1970s. The Twin Cities Metropolitan Area has one of the nation's largest systems in terms of the number of park and ride facilities and total capacity (Lamothe, 2004:4-1). Currently, the Metropolitan Council, the metropolitan planning organization for the seven-county metropolitan area (Anoka County, Carver County, Dakota County, Hennepin County, Ramsey County, Scott County, and Washington County) works cooperatively with the Metro Transit, the regional transit provider, and the Minnesota Department of Transportation to plan, design, operate, and maintain the regional park and ride system (Lamothe, 2004:4-1).

FIGURE 3.1 Seven-County Twin Cities Metropolitan Area

Source: Minneapolis-St. Paul (Twin Cities) Metropolitan Council, 2004
According to the 2000 Travel Behavior Inventory (TBI), a comprehensive survey of travel in the Twin Cities Metropolitan Area conducted jointly by the Metropolitan Council and the Minnesota Department of Transportation, most trips (93%) were made by motor vehicles (10.3 trips per day per household). More than one half of the people (53%) making auto trips drove alone and more than three-fourths of the people (78%) drove alone for work. The 2000 TBI also indicates that the number of person trips per household generally increases with household income. The trip rate for households in the lowest income group (less than $5,000) is 6.2 trips (all modes) per day. By contrast, the trip rate for households in the highest income group ($150,000 or more) is 16.1 trips per day. The more vehicles a household has, the more motorized trips the household members tend to take. The average number of vehicles available to the household was 1.8. Person trips in motorized vehicles (including transit) ranged from an average of 3.2 trips per day for households with no vehicles to 14.7 trips per day for households with five or more vehicles.

Besides, traffic congestion was indicated in a Metro Residents Survey by 38% of the residents as the single most important problem facing the region (2003: 3-13). On average, the perceived commute time has increased to 27.5 minutes from 23.6 minutes in 2003. For those residents who identified transportation issues (including traffic congestion) as the single most important problem, 45% suggested improving and increasing mass transit, while another 32% suggested improving/increasing the road infrastructure (Metro Residents Survey, 2003: 23-24). To meet the area’s long-range transportation needs, residents think that resolving the transportation issues facing the region will require improving/increasing both road infrastructure and mass transit. As showed in TABLE 3.1, expanding the park and ride express bus program is very important to 38% of the residents, while other transportation programs are referred to as very important as well.
TABLE 3.1 Survey of Very Important Transportation Programs

<table>
<thead>
<tr>
<th>Transportation program</th>
<th>Percent of all residents indicating this as being Very Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Optimizing the capacity and safety of existing roads</td>
<td>64</td>
</tr>
<tr>
<td>2. Adding extra lanes to freeways</td>
<td>57</td>
</tr>
<tr>
<td>3. Developing a commuter/light-rail system</td>
<td>45</td>
</tr>
<tr>
<td>4. Expanding the park and ride/express bus program</td>
<td>38</td>
</tr>
<tr>
<td>5. Expanding the Metro Transit bus system</td>
<td>36</td>
</tr>
</tbody>
</table>

Source: Metro Residents Survey, 2003, Metropolitan Council, page 23

In fact, park and ride planning is not only a matter in the Twin Cities Metropolitan Area, but the proposed methodology could be implemented as a part of the transportation demand modeling in other places which have similar concerns. Congestion problems are predicted to accumulate in future years. For example, the Twin Cities transportation demand forecasting model was developed by Transplan in FY2004 and the Metropolitan Council re-calibrated the model based on the 2000 Travel Behavior Inventory (TBI). This indicates park and ride/express bus system programs are very important, and the integration with park and ride systems is still currently in the process.

**Research Hypothesis**

To integrate the Twin Cities Metropolitan Area’s park and ride systems and to model the demand for park and ride facilities, this research makes an attempt to better understand this by examining the efficiency of park and ride systems in terms of the number of used parking lots in each facility. By applying the proposed methodology, the objectives are to examine the spatial distribution of the of park and ride facility usage, and to identify explanatory variables for predicting park and ride facility usage.

The research hypothesis is that the successful park and ride facilities can be
predicted based upon factors such as serviced population, employment, traffic condition, median household income, vehicle availability, presence of proximity factors that favor the bus (such as the ease to fixed guideway features), and the accessibility to retail trade centers.

Therefore, the research question is how to size the provision of park and ride facilities, connected with choosing statistically significant independent variables for predicting park and ride usage in the Twin Cities Metropolitan Area. This is achieved by a classic multivariate regression analysis, which consists of park and ride facility usage as the dependent variable and other experimental independent variables.

However, the usage of park and ride facilities (the dependent variable in the demand forecasting model) must be tested for spatial autocorrelation. This indispensable procedure helps incorporate spatial association, while constructing a demand forecasting model for the park and ride facilities in the Twin Cities Metropolitan Area. Does the usage pattern exhibit complete spatial randomness, clustering, or regularity? While testing for spatial association, the null hypothesis is the spatial randomness in the usage of park and ride facilities. That is, the observed usage value at a park and ride facility does not depend on the values at neighboring locations. Conversely, the alternative hypothesis of spatial autocorrelation is the spatial dependence of the usage of park and ride facilities. In other words, if there is a positive spatial autocorrelation, like values tend to cluster in space, neighbors are similar. Conversely, a negative spatial autocorrelation means neighbors are dissimilar. The examination for testing spatial autocorrelation of park and ride usage is adopted to incorporate the results of best-fit model.

To decide whether or not the spatial dependence should be incorporated as a part of regression model while predicting the usage of park and ride systems, it is necessary to examine the spatial autocorrelation and the significance level of the
residuals of park and ride prediction model. The entire park and ride analytical process is presented quantitatively as well as geographically based on the conditions in the Twin Cities Metropolitan Area.

**Sampling and Zonal Structure**

The datasets for the park and ride demand modeling in this study are sourced from the Twin Cities MetroGIS, a collaborative organization representing local governments and other organizations established to foster sharing of geospatial data in the seven-county Minneapolis-St. Paul (Twin Cities) Metropolitan Area of Minnesota. The information in MetroGIS is primarily created by converting geographic data produced by Metro Commuter Services, Metropolitan Council and the U.S. Census Bureau. In this research, a series of databases for the forecasting model are constructed and organized into several themes of data (FIGURE 3.2). The list includes: (1) Transportation (park and ride lots, major highways, functional class roads, bus service and routes, bus stops, fixed guideways, and transit centers), (2) Planning and Development (FY2000 counties, FY2000 census tracts, and FY2000 transportation analysis zones), and (3) Demographics and Business (general demographic characteristics, selected economic characteristics, and major shopping centers). All of these data are then imported to the developed GIS geodatabase for park and ride demand modeling. The base year for this diagnostic modeling is FY2004, since it is the only time point that the park and ride facility usage is available.

**FIGURE 3.2 Development of Geodatabase**

- Geodatabase for Park and Ride Demand Modeling
  - Demographics and Business
  - Planning and Development
  - Transportation
1. Selecting Park and Ride Facilities for Modeling

The key dataset representing all existing and planned park and ride facilities in the Twin Cities Metropolitan Area is created by converting geographic data produced by the Metro Commuter Services and attaching information from the Metro Transit in FY2000. By March 2004, this database is updated with the latest additions, and consists of 183 geospatial data point toward the year of FY2005. Of the 183 data for park and ride facilities in the study area, 9 are planned for future, 35 are inactive, and only 139 facilities are active. The combined capacity of these parking lots is more than 20,000 parking spaces. The distribution of the built year vs. capacity is presented in FIGURE 3.3. The first peak of adopting park and ride facilities as an alternative of transportation strategies was for the years between FY1975 and FY1980. The second peak was between FY1990 and FY1995, with a relatively higher capacity. Recently, the trend is even more obvious, while designating high-capacity (greater than 400 lots) park and ride facilities in the study area.

FIGURE 3.3 Park and Ride Capacity vs. Built Year

Source: Minneapolis-St. Paul (Twin Cities) Metropolitan Council, 2004

For modeling the park and ride facility usage (the dependant variable) in the Twin Cities Metropolitan Area, only 139 park and ride facilities are eventually selected
because these facilities were active in a recent site visit. The park and ride dataset indicates associated information such as facility name, city location, type of facility, capacity of facility, used capacity (the number of used parking spots in a recent site visit), time period (future, active, or inactive), and the year the primary structure was initially built (MetroGIS, 2004). An assumption of this park and ride demand modeling process is that the usages of park and ride facilities in FY2004 are not affected by the built years and/or the design of park and ride facilities per se. Facility users would make their traveling decisions for logical reasons and other least-cost considerations, no matter which year their best choice of facility is built or how the facility is physically designed.

2. TAZ Definition

Traffic Analysis Zones (TAZs) are subdivisions of geographic areas delineated for land use and travel analysis purposes (Caliper, 2001). Typically, a TAZ system should follow U.S. Census geography (block, block group and tract) when possible, because it would make the use of census data easier and eliminate complex manipulation of the data. Besides, TAZs should contain land uses relatively homogenous in character (Caliper, 2001). Combinations of land use types within a TAZ should be avoided. Physical barriers such as railroad lines, rivers, and major roadways should make up the boundaries of the TAZs.

In this research, the basic zonal unit is “TAZ 2000” (see FIGURE 3.4), a modification developed jointly by the Minnesota Department of Transportation and the Metropolitan Council. Boundary lines are generally based on TIGER 2000 U.S. Census block boundaries projected to UTM, NAD83 Meters. Early 2004, this set of TAZs became the official TAZ system for travel demand modeling, socioeconomic
forecasts and community comprehensive plan development, in addition to its existing use for the 2000 Census Transportation Planning Package (CTPP).

FIGURE 3.4 Transportation Analysis Zones

Source: Minneapolis-St. Paul (Twin Cities) Metropolitan Council, 2004

The total number of TAZs in the Twin Cities Metropolitan Area is 1201, which is broken down into more detail than in the census tract level, specifically for transportation purposes (Metropolitan Council, 2004). Examples of variables in the "TAZ 2000" include population, the number of households, and the number and types of jobs within each TAZ boundary.

3. Thiessen Polygon Definition

Another zonal unit proposed by this research that defines the influence area of each park and ride facility in the Twin Cities Metropolitan Area is called the Thiessen Polygon (also known as Voronoi Tessellation, see FIGURE 3.5). These Thiessen
Polygons are generated around a set of points, i.e., park and ride point data, whose boundaries define the area closest to each point relative to all other points. They are mathematically defined by the perpendicular bisectors of the lines between all points and divide space such that each location is allocated to the nearest control point (ESRI, 2004).

FIGURE 3.5 Thiessen Polygons

Source: Minneapolis-St. Paul (Twin Cities) Metropolitan Council, 2004

Since 139 active park and ride facilities are selected from the Twin Cities transportation network system for this demand modeling process, 139 Thiessen Polygons are generated respectively as the influence areas for these selected park and ride facilities. Thus, the information for each park and ride facilities is transferred to its corresponding Thiessen Polygon so that we can determinate how much capacity, used capacity, and other associated data within each defined zonal area. This implies that the users would adopt the nearest facility rather than other facilities.
Data Aggregation and Centroids

Aggregation of the data is a critical part of the modeling process. The model uses the Thiessen Polygons to represent the influence areas of park and ride facilities. Given the information about park and ride facilities, each Thiessen Polygon has the same attributes that correspond to its facility. On the other hand, these Thiessen Polygons also propose a zonal boundary for predicting park and ride usage from the demand side. By overlaying the TAZs and Thiessen Polygons, it is possible to quantify potential factors that affect the usage of park and ride facilities and to compute the values of certain variables within Thiessen Polygon boundary (see FIGURE 3.6).

FIGURE 3.6 Overlaying of TAZs and Thiessen Polygons

Source: Minneapolis-St. Paul (Twin Cities) Metropolitan Council, 2004
Data aggregation from census tracts to TAZs, or from TAZs to Thiessen Polygons is essential for this park and ride demand model. These data aggregation procedures are based upon the assumption that the distribution of data are normally spread out within the census boundaries, so that measures can be computed proportionally, based on the proportion of the overlapped area. For example, the population and household data for each Thiessen Polygon is aggregated from TAZ data. In most cases, TAZs fit neatly into each Thiessen Polygon. In the few cases where they did not fit, a determination was made as to the appropriate Thiessen Polygon to include the shared ratio of population from TAZ, based on the overlapping proportion of the area within the TAZ.

In this section of the research paper discuss adoption of GIS as a tool and addresses some potential factors (independent variables) for this park and ride demand (dependant variable) modeling process. GIS is not only for displaying thematic maps, but also for analyzing spatial statistics for this facility planning. The following shows the results of data aggregation from 1201 TAZs to 139 park and ride Thiessen Polygons. The variables are divided into different categories and these analytical steps are documented as parts of the demand modeling processes.

1. Demographic Characteristics

To model the demand of park and ride facilities, 139 Thiessen Polygons are created to define influence areas for the 139 selected park and ride facilities. However, for the Twin Cities Metropolitan Area, a number of explanatory data related to the U.S. census demographic characteristics, such as population, number of households, employment, retail and non-retail employment, are initially stored in the transportation analysis zone level. Since park and ride facility usage data are available only for FY2004, these demographic data, respectively, for FY1990 and
FY2000 are projected into the year of 2004, assuming the trend between FY1990 and FY2000 remains the same without drastic change toward FY2004 (see TABLE 3.2).

**TABLE 3.2 Predictions of Demographic Explanatory Data by TAZs**

<table>
<thead>
<tr>
<th>TAZ</th>
<th>Population</th>
<th>Number of Household</th>
<th>Employment</th>
<th>Retail Employment</th>
<th>Non-Retail Employment</th>
</tr>
</thead>
<tbody>
<tr>
<td>FY1990</td>
<td>2288721</td>
<td>875833</td>
<td>1284265</td>
<td>231910</td>
<td>1052355</td>
</tr>
<tr>
<td>FY2000</td>
<td>2642056</td>
<td>1021454</td>
<td>1562833</td>
<td>171272</td>
<td>1391561</td>
</tr>
<tr>
<td>Difference/10</td>
<td>35333.5</td>
<td>14562.1</td>
<td>27856.8</td>
<td>-6063.8</td>
<td>33920.6</td>
</tr>
<tr>
<td>Expected_04</td>
<td>2783390</td>
<td>1079702</td>
<td>1674260</td>
<td>147017</td>
<td>1527243</td>
</tr>
<tr>
<td>Estimated_04</td>
<td>2754265</td>
<td>1079701</td>
<td>1674253</td>
<td>147031</td>
<td>1527233</td>
</tr>
</tbody>
</table>

Source: Minneapolis-St. Paul (Twin Cities) Metropolitan Council, 2004

The estimated values of all demographic characteristics for each TAZ in the year 2004 is basically derived from the values for each TAZ in FY2000 plus the proportional difference with its changing rate for 4 years, projected to FY2004.

Estimated Population = \([\text{POP}_{2000}] / 2642056 \times (35333.5) \times 4 + [\text{POP}_{2000}]\)  
Estimated Number of Households = \([\text{HH}_{2000}] / 1021454 \times (14562.1) \times 4 + [\text{HH}_{2000}]\)  
Estimated Employment = \([\text{EMP}_{2000}] / 1562833 \times (27856.8) \times 4 + [\text{EMP}_{2000}]\)  
Estimated Retail Employment = \([\text{RET}_{2000}] / 171272 \times (-6063.8) \times 4 + [\text{RET}_{2000}]\)  
Estimated Non-Retail Employment = \([\text{NRET}_{2000}] / 1391561 \times (33920.6) \times 4 + [\text{NRET}_{2000}]\)

After projecting the demographic variables into FY2004, these data are aggregated from TAZ level to Thiessen Polygons. TABLE 3.3 shows the summary of the demographic explanatory data, based on park and ride Thiessen Polygons.

**TABLE 3.3 Summary of Demographic Explanatory Data**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std.</th>
<th>Min</th>
<th>Max</th>
<th>S.E.</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUM_04_POP</td>
<td>139</td>
<td>19814.8557</td>
<td>23193.2852</td>
<td>1153</td>
<td>199218</td>
<td>1967.2288</td>
<td>13883</td>
</tr>
<tr>
<td>SUM_04_HH</td>
<td>139</td>
<td>7767.6330</td>
<td>9720.3427</td>
<td>487</td>
<td>82420</td>
<td>824.4687</td>
<td>5296</td>
</tr>
<tr>
<td>SUM_04_EMP</td>
<td>139</td>
<td>12044.9853</td>
<td>26983.1977</td>
<td>412</td>
<td>292980</td>
<td>2288.6850</td>
<td>6641</td>
</tr>
<tr>
<td>SUM_04_RET</td>
<td>139</td>
<td>1057.7770</td>
<td>1644.1588</td>
<td>14</td>
<td>14819</td>
<td>139.4557</td>
<td>608</td>
</tr>
<tr>
<td>SUM_04_NRE</td>
<td>139</td>
<td>10987.2874</td>
<td>25812.3654</td>
<td>382</td>
<td>281192</td>
<td>2189.3763</td>
<td>5945</td>
</tr>
</tbody>
</table>

Source: Minneapolis-St. Paul (Twin Cities) Metropolitan Council, 2004
The maps below (FIGURE 3.7 and 3.8) show estimated FY2004 population by TAZs and by park and ride Thiessen Polygons.

FIGURE 3.7 Estimated FY2004 Population by TAZs

FIGURE 3.8 Estimated FY2004 Population by Thiessen Polygons
The maps below (FIGURE 3.9 and 3.10) show estimated FY2004 Number of Household by TAZs and by park and ride Thiessen Polygons.

FIGURE 3.9 Estimated FY2004 Number of Household by TAZs

FIGURE 3.10 Estimated FY2004 Number of Household by Thiessen Polygons
The maps below (FIGURE 3.11 and 3.12) show estimated FY2004 Employment by TAZs and by park and ride Thiessen Polygons.

FIGURE 3.11 Estimated FY2004 Employment by TAZs

FIGURE 3.12 Estimated FY2004 Employment by Thiessen Polygons
The maps below (FIGURE 3.13 and 3.14) show estimated FY2004 Retail Employment by TAZs and by park and ride Thiessen Polygons.

FIGURE 3.13 Estimated FY2004 Retail Employment by TAZs

FIGURE 3.14 Estimated FY2004 Retail Employment by Thiessen Polygons
The maps below (FIGURE 3.15 and 3.16) show estimated FY2004 Non-Retail Employment by TAZs and by park and ride Thiessen Polygons.

FIGURE 3.15 Estimated FY2004 Non-Retail Employment by TAZs

FIGURE 3.16 Estimated FY2004 Non-Retail Employment by Thiessen Polygons
2. Economic Characteristics

Originally, the data for selected economic characteristics are stored at the census tract level (690 in total for the study area). To link with the demand modeling of park and ride facilities in the Twin Cities Metropolitan Area, median household income and vehicle ownership are also taken into account in the modeling process. The median household income might have a positive, negative or non-significant association with park and ride usage. The number of households without vehicle availability might result in higher or lower usage of park and ride facilities as well. Besides, the number of non-white people and the number of families under poverty level might also be sensitive to the park and ride usage. Therefore, these four economic data are aggregated into the park and ride Thiessen Polygon level.

The median household income (in dollars) for each Thiessen Polygon represents the average median income per household, derived from overlaying two zonal layers – census tracts and Thiessen Polygons, calculating the shared ratio of the number of households from census tracts, and computed values based on the overlapping proportion of the area within the Thiessen Polygons. The equation is as follows.

\[
\overline{IHMED}_k = \frac{IHMED_1 \times HH_1' + IHMED_2 \times HH_2' + IHMED_3 \times HH_3' + \ldots + IHMED_i \times HH_i'}{\sum_{i=1}^{n} HH_i'}
\]

where \( \overline{IHMED}_k \) is the average median household income in Thiessen Polygons \( k \),

\( IHMED_i \) is the median household income of the overlaying census tract \( i \), and

\( HH_i' \) is the number of households derived from the shared ratio of census tract \( i \).

Also, the number of households without vehicle availability may possibly be a part of the essential exploratory data. Because park and ride facilities are for travelers who drive to sites and then adopt public transit mode, it is assumed that the higher
number of households without vehicle availability results in a lower usage of park and ride facilities. That is, more households that do not own any vehicle would not utilize any parking space in the park and ride site. The data are organized from the initial census tract level to park and ride Thiessen Polygons, so that it could be related to usage in the influence area of park and ride facilities. TABLE 3.4 indicates the summary of these two economic explanatory data based on park and ride Thiessen Polygons. AVE_IHMED means the average median household income within each park and ride influence area, SUM_VEHN means number of households without owning a vehicle, NWH means the number of population are non-white people, and POVFAMS means the number of families under poverty level in the Thiessen Polygons.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std.</th>
<th>Min</th>
<th>Max</th>
<th>S.E.</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVE_IHMED</td>
<td>139</td>
<td>62367.0719</td>
<td>13022.4783</td>
<td>34187</td>
<td>94402</td>
<td>1108.5470</td>
<td>61161</td>
</tr>
<tr>
<td>SUM_VEHN</td>
<td>139</td>
<td>519.9072</td>
<td>1640.0761</td>
<td>1</td>
<td>15116</td>
<td>139.1094</td>
<td>130</td>
</tr>
<tr>
<td>NWH</td>
<td>139</td>
<td>290701.978</td>
<td>712667.441</td>
<td>4905</td>
<td>583111</td>
<td>60447.6634</td>
<td>88737</td>
</tr>
<tr>
<td>POVFAMS</td>
<td>139</td>
<td>53820.7554</td>
<td>134552.518</td>
<td>888</td>
<td>1037926</td>
<td>11412.5957</td>
<td>21444</td>
</tr>
</tbody>
</table>

Source: Minneapolis-St. Paul (Twin Cities) Metropolitan Council, 2004

The maps (FIGURE 3.17 and 3.18) show median household income by census tracts and by park and ride Thiessen Polygons.
FIGURE 3.17 Median Household Income by Census Tracts

FIGURE 3.18 Average Median Household Income by Thiessen Polygons
The maps (FIGURE 3.19 and 3.20) show number of households without vehicle availability by census tracts and by park and ride Thiessen Polygons.

FIGURE 3.19 Number of Households without Vehicle by Census Tracts

FIGURE 3.20 Number of Households without Vehicle by Thiessen Polygons
FIGURE 3.21 and 3.22 shows number of non-white people and number of families under poverty level in the Thiessen Polygons.

FIGURE 3.21 Number of non-white people by Thiessen Polygons

FIGURE 3.22 Number of Families under Poverty Level by Thiessen Polygons
3. Traffic Volume

Another possible factor related to transportation systems is traffic volume and its affect on the usage of park and ride facilities. Nearby traffic volume on highways could direct higher or lower usage of park and ride facilities. Released by the Office of Transportation Data and Analysis, Minnesota Department of Transportation, the annual average daily traffic (AADT) for FY2002 is adopted in this research, which assumes that AADT is sub-phenomenon of the changing demographics and economics. Traffic has changed, based on changing population and/or changing economic activities. Therefore, FY2002 AADT, the latest data available for traffic volume, is applied in the modeling process.

First, AADT density raster data throughout the entire Twin Cities Metropolitan Area are created, based on the FY2002 AADT on the highways. By means of zonal statistics, density raster values are transferred into 1201 TAZ level and 139 Thiessen Polygon level. Indicated in TABLE 3.5, AADT_Sum means the accumulated value for AADT density within the zones.

<table>
<thead>
<tr>
<th>AADT_Sum</th>
<th>N</th>
<th>Mean</th>
<th>Std.</th>
<th>Min</th>
<th>Max</th>
<th>S.E.</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>139</td>
<td>4614.77468</td>
<td>5363.50292</td>
<td>17.9066</td>
<td>25614.8602</td>
<td>454.92638</td>
<td>2831.3993</td>
</tr>
</tbody>
</table>

Source: Minneapolis-St. Paul (Twin Cities) Metropolitan Council, 2004

The map (FIGURE 3.23) shows the annual average daily traffic (AADT) on the highways and AADT density raster throughout the study area. The map (FIGURE 3.24) shows the annual average daily traffic (AADT) on the highways and accumulated AADT density by TAZs, followed by the second map (FIGURE 3.25), which shows accumulated AADT density by park and ride Thiessen Polygons.
FIGURE 3.23 Annual Average Daily Traffic (AADT) and Density Raster

FIGURE 3.24 Accumulated AADT Density by TAZs
4. Centroids and Distances

The proximity from the centroid of each park and ride influence area to its nearest parking facility could affect residents’ accessibility to each facility (see FIGURE 3.26). A set of centroids could be created by Geographic Information System (GIS) software. For example, centroids of park and ride Thiessen Polygons are closely placed geographically to the center of that influence area. These hubs are located, using GIS and transferred with their corresponding measured data so that data for each park and ride influence area are condensed into one location, a centroid node. Centroids of these park and ride Thiessen Polygons represent the origins where people live. By calculating the distance, the proximity can be measured as one exploratory factor, jointly resulting in different levels of facility usage.

Furthermore, the proximity from park and ride facilities to the destinations
should be taken into account simultaneously. Secondhand data derived from the Minnesota Metro Council indicate 331 major shopping centers within the Twin Cities Metropolitan Area. Without the destination survey from each park and ride facility, it is not easy to predict where people travel. An alternative developed by this research is to sum the distances from each park and ride facilities to all major shopping centers (see FIGURE 3.27).

This process results in a 139 by 331 distance matrix, by which the total distances from each facility can be calculated. Therefore, the relative proximity from park and ride facilities to all major shopping malls is measured and adopted in the demand model. FIGURE 3.28 shows the summary and distribution of park and ride proximity to major shopping malls (PR_SHP_SUM).

FIGURE 3.26 Park and Ride Proximity to Centroids of Thiessen Polygons

![Map showing the proximity of park and ride facilities to the centroids of Thiessen polygons.](image)
Moreover, the proximity of each park and ride facility to public transit might be very sensitive to park and ride usage. One consideration related to transit is bus routes. It is assumed that trips provided during weekdays and weekends could be
easily adjusted by transit agencies, based on desirable ridership so that only geospatial locations are taken into account. Besides, it also assumed that park and ride facilities might only be served by either express or suburban types of bus routes. So, only the distances between park and ride facilities and selected bus routes (express and suburban) are measured in the process (see FIGURE 3.29).

The other similar concern about transit is on fixed guideways, which include high occupied vehicle (HOV) lanes, bus lanes, bus ways, meter bypasses, online stations, shoulder lanes, and turnarounds. All of these are physically constructed as an advantage for transit oriented. Hence, the distances between each park and ride facility and its nearest segment of fixed guideway are also important to the usage modeling process (see FIGURE 3.30). Not only the distances are measured, based on the geospatial locations, but the nearest fixed guideway segment of park and ride facilities are identified.

FIGURE 3.29 Park and Ride Proximity to Express Route
In conclusion, proximity to centroid of park and ride Thiessen Polygons, to major shopping centers, to nearest express routes, and to nearest fixed guideways can be prospective explanatory data in the park and ride usage modeling process. The values are all aggregated into 139 park and ride Thiessen Polygons (TABLE 3.6).

<table>
<thead>
<tr>
<th>TABLE 3.6 Summary of Proximity Explanatory Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>DIS_Cen_PR</td>
</tr>
<tr>
<td>PR_Shp_Sum</td>
</tr>
<tr>
<td>DIS.PR_Exp</td>
</tr>
<tr>
<td>DIS.PR_Fix</td>
</tr>
</tbody>
</table>

Source: Minneapolis-St. Paul (Twin Cities) Metropolitan Council, 2004
However, every distance dataset does not exhibit like a normal distribution, and the values are highly skewed. The differences of mean and median are very large. For succeeding multivariate regression modeling process, it is necessary to transform distance if highly skewed distribution exists. To make distance values distribute more normally, logarithm is adopted for transformation. For example, the histograms (see FIGURE 3.31), created by ArcGIS Geostatistical Analyst, displays the distributions of distance from centroids of Thiessen Polygons to its park and ride facilities (DIS_CEN_PR), and the transformed distance by logarithm (LN_CEN_PR). The histogram for DIS_CEN_PR exhibit highly skewed to the right, compared with the histogram for LN_CEN_PR.

The coefficient of skewness is a measure of the symmetry of a distribution. For symmetric distributions, the coefficient of skewness is zero. If a distribution has a long right tail of large values, it is positively skewed, and if it has a long left tail of small values, it is negatively skewed. In FIGURE 3.31, the skewness of transformed distance (LN_CEN_PR) is relatively closer to zero than actual distance (DIS_CEN_PR) is.

Besides, the kurtosis is another indicator that provides a measure of how likely the distribution will produce outliers, based on the sized of the tailed of a distribution. The kurtosis of a normal distribution is 3. Distribution of transformed distance (LN_CEN_PR) has a kurtosis value relatively closer to 3 than actual distance (DIS_CEN_PR) does.

TABLE 3.7 Summarizes skewness and kurtosis of each proximity explanatory variable. This table verifies that all transformed distances represent more normalized distributions, so only transformed distances are adopted in the park and ride usage modeling process, instead of actual distance.
FIGURE 3.31 Comparison of Distribution Statistics for Proximity

TABLE 3.7 Summary of Transformed Proximity Explanatory Data

<table>
<thead>
<tr>
<th>Actual Distance</th>
<th>Transformed Distance (Ln)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Skewness</td>
</tr>
<tr>
<td>Cen_PR</td>
<td>1.4388</td>
</tr>
<tr>
<td>PR_Shp</td>
<td>1.5706</td>
</tr>
<tr>
<td>PR_Exp</td>
<td>2.9324</td>
</tr>
<tr>
<td>PR_Fixed</td>
<td>2.4096</td>
</tr>
</tbody>
</table>

TABLE 3.8 shows the summary of transformed proximity explanatory data.

TABLE 3.8 Summary of Transformed Proximity Explanatory Data

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std.</th>
<th>Min</th>
<th>Max</th>
<th>S.E.</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>LN_Cen_PR</td>
<td>139</td>
<td>7.0967</td>
<td>1.0311</td>
<td>4.0158</td>
<td>9.1076</td>
<td>0.0875</td>
<td>7.0708</td>
</tr>
<tr>
<td>LN_PR_Shp</td>
<td>139</td>
<td>15.9649</td>
<td>0.2498</td>
<td>15.5879</td>
<td>16.7386</td>
<td>0.0212</td>
<td>15.9268</td>
</tr>
<tr>
<td>LN_PR_Exp</td>
<td>139</td>
<td>7.8326</td>
<td>0.9801</td>
<td>5.0707</td>
<td>10.4719</td>
<td>0.0831</td>
<td>7.7140</td>
</tr>
<tr>
<td>LN_PR_Fix</td>
<td>139</td>
<td>7.4516</td>
<td>1.6227</td>
<td>2.4376</td>
<td>10.4119</td>
<td>0.1376</td>
<td>7.5269</td>
</tr>
</tbody>
</table>

Source: Minneapolis-St. Paul (Twin Cities) Metropolitan Council, 2004
CHAPTER IV: MULTIVARIATE SPATIAL REGRESSIVE ANALYSES

To identify “independent variables” for predicting park and ride facility usage within the Twin Cities Metropolitan Area, this part of the research intends to examine the significance of the spatial autocorrelation in the regression model simultaneously.

Multivariate Regression Analysis

The potential independent variables are summarized in TABLE 4.1

| TABLE 4.1 Summary of Potential Independent Variables |
|----------------------------------------|--------|--------|--------|--------|--------|--------|--------|
| Population \(^1\)                  | 139    | 1153   | 199218 | 19814.86 | 13883  | 23193.3114 | 1967.2310 |
| \(D_{HH}\) \(^2\)                  | 139    | 6.16   | 1571.25 | 306.68   | 267.08 | 261.6343 | 22.1915   |
| AADT \(^3\)                        | 139    | 17.91  | 25614.86| 4614.77  | 2831.40| 5363.5029| 454.9264  |
| \(iHMED\) \(^4\)                  | 139    | 34187.00| 94402.00| 62367.07 | 61161.00| 13069.5760| 1108.5470 |
| \(R_{VEHN}\) \(^5\)               | 139    | 0.20   | 18.34   | 3.77     | 2.58   | 3.4208   | 0.2901    |
| \(R_{EMP}\) \(^6\)                | 139    | 8.24   | 543.30  | 65.99    | 43.21  | 77.8754  | 6.6053    |
| \(R_{RET}\) \(^7\)                | 139    | 0.19   | 53.45   | 6.28     | 3.81   | 8.0021   | 0.6787    |
| \(R_{NRET}\) \(^8\)               | 139    | 7.79   | 500.60  | 59.58    | 40.09  | 71.7763  | 6.0880    |
| \(R_{NWH}\) \(^9\)                | 139    | 1.73   | 54.67   | 9.88     | 7.70   | 8.9874   | 0.7623    |
| \(R_{POVFA}\) \(^{10}\)           | 139    | 0.23   | 11.39   | 1.99     | 1.63   | 1.6669   | 0.1414    |
| Ln \(_{CEN}\) \(^{11}\)           | 139    | 15.59  | 16.74   | 7.83     | 15.93  | 1.6227   | 0.0212    |
| Ln \(_{EXP}\) \(^{12}\)           | 139    | 5.07   | 10.47   | 7.83     | 7.71   | 0.9801   | 0.0831    |
| Ln \(_{FIXED}\) \(^{13}\)         | 139    | 2.44   | 10.41   | 7.45     | 7.53   | 1.6227   | 0.1376    |
| Ln \(_{SHP}\) \(^{14}\)           | 139    | 15.59  | 16.74   | 15.96    | 15.93  | 0.2498   | 0.0212    |

\(^1\) Population: the number of population in the Thiessen Polygon (Sum_04_Pop);
\(^2\) \(D_{HH}\): the density of households in the Thiessen Polygons (\(km^2\));
\(^3\) AADT: the sum of annual average daily traffic in the Thiessen Polygon (Sum_AADT);
\(^4\) \(iHMED\): the average median household income in the Thiessen Polygon (Ave_IHMED);
\(^5\) \(R_{VEHN}\): the ratio of households without vehicle by number of households in the Thiessen Polygon (Sum_VEHN/Sum_04_Pop);
\(^6\) \(R_{EMP}\): the ratio of the number of employments by the population in the Thiessen Polygon (Sum_04_Emp/Sum_04_Pop);
\(^7\) \(R_{RET}\): the ratio of the number of employments in retail sector by the population in the Thiessen Polygon (Sum_04_Ret/Sum_04_Pop);
\(^8\) \(R_{Non-Retail}\): the ratio of the number of employments in non-retail sector by the population in the Thiessen Polygon (Sum_04_NRET/Sum_04_Pop);
\(^9\) \(R_{NWH}\): the ratio of the number of non-white people by the population in the Thiessen Polygon (NWHPOP/Sum_04_Pop);
\(^{10}\) \(R_{POVFA}\): the ratio of the number of families under the poverty level in the Thiessen Polygon (POVFA/HH);
\(^{11}\) Ln \(_{CEN}\): Transformed proximity from the centroid of the Thiessen Polygon to its park and ride facility (LN_CEN_PR);
\(^{12}\) Ln \(_{EXP}\): Transformed proximity from park and ride facility to its nearest express route segment (LN_PR_EXP);
\(^{13}\) Ln \(_{FIXED}\): Transformed proximity from park and ride facility to its nearest fixed guideway segment (LN_PR_FIX);
\(^{14}\) Ln \(_{SHP}\): Transformed total proximity from park and ride facility to all major shopping malls (LN_PR_SHP).
1. Correlation Matrix

A Pearson correlation matrix (Table 4.2) can determine the extent to which values of the two variables are linearly related to each other. Noticeably, "Usage" and "Capacity" are highly correlated to each other.

Table 4.2 Pearson Correlation Matrix of All Potential Independent Variables

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Usage</th>
<th>Capacity</th>
<th>Population</th>
<th>D_HH</th>
<th>AADT</th>
<th>IHMED</th>
<th>R_VEHN</th>
<th>R_EMP</th>
<th>R_RET</th>
<th>R_NRET</th>
<th>R_NWH</th>
<th>R_POVFA</th>
<th>Ln_CEN</th>
<th>Ln_EXP</th>
<th>Ln_FIXED</th>
<th>Ln_SHP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage</td>
<td>1.00</td>
<td>0.96</td>
<td>0.06</td>
<td>0.01</td>
<td>0.04</td>
<td>-0.05</td>
<td>-0.10</td>
<td>0.05</td>
<td>0.16</td>
<td>0.04</td>
<td>0.00</td>
<td>-0.02</td>
<td>-0.18</td>
<td>-0.11</td>
<td>-0.44</td>
<td>-0.10</td>
</tr>
<tr>
<td>Capacity</td>
<td>0.96</td>
<td>1.00</td>
<td>0.04</td>
<td>-0.02</td>
<td>0.05</td>
<td>-0.01</td>
<td>-0.14</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.16</td>
<td>-0.08</td>
<td>-0.40</td>
<td>-0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Population</td>
<td>0.06</td>
<td>0.04</td>
<td>1.00</td>
<td>0.54</td>
<td>0.46</td>
<td>-0.34</td>
<td>0.66</td>
<td>-0.06</td>
<td>-0.10</td>
<td>-0.05</td>
<td>0.46</td>
<td>0.38</td>
<td>0.22</td>
<td>-0.17</td>
<td>-0.18</td>
<td>-0.22</td>
</tr>
<tr>
<td>D_HH</td>
<td>0.01</td>
<td>-0.02</td>
<td>0.54</td>
<td>1.00</td>
<td>-0.02</td>
<td>-0.61</td>
<td>0.72</td>
<td>0.14</td>
<td>0.13</td>
<td>0.14</td>
<td>0.60</td>
<td>0.37</td>
<td>-0.34</td>
<td>-0.57</td>
<td>-0.48</td>
<td>-0.73</td>
</tr>
<tr>
<td>AADT</td>
<td>0.04</td>
<td>0.05</td>
<td>0.46</td>
<td>-0.02</td>
<td>1.00</td>
<td>-0.19</td>
<td>0.27</td>
<td>0.00</td>
<td>-0.04</td>
<td>0.01</td>
<td>0.08</td>
<td>0.12</td>
<td>0.33</td>
<td>0.21</td>
<td>0.12</td>
<td>0.24</td>
</tr>
<tr>
<td>IHMED</td>
<td>-0.05</td>
<td>-0.01</td>
<td>-0.34</td>
<td>-0.61</td>
<td>-0.19</td>
<td>1.00</td>
<td>-0.73</td>
<td>-0.17</td>
<td>-0.12</td>
<td>-0.18</td>
<td>-0.61</td>
<td>-0.56</td>
<td>0.15</td>
<td>0.29</td>
<td>0.31</td>
<td>0.38</td>
</tr>
<tr>
<td>R_VEHN</td>
<td>-0.10</td>
<td>-0.14</td>
<td>0.66</td>
<td>0.72</td>
<td>0.27</td>
<td>-0.73</td>
<td>1.00</td>
<td>0.17</td>
<td>0.10</td>
<td>0.17</td>
<td>0.76</td>
<td>0.72</td>
<td>-0.02</td>
<td>-0.32</td>
<td>-0.25</td>
<td>-0.45</td>
</tr>
<tr>
<td>R_EMP</td>
<td>0.05</td>
<td>0.00</td>
<td>-0.06</td>
<td>0.14</td>
<td>0.00</td>
<td>-0.17</td>
<td>0.17</td>
<td>1.00</td>
<td>0.81</td>
<td>1.00</td>
<td>0.18</td>
<td>0.02</td>
<td>-0.28</td>
<td>-0.38</td>
<td>-0.30</td>
<td>-0.39</td>
</tr>
<tr>
<td>R_RET</td>
<td>0.16</td>
<td>0.09</td>
<td>-0.10</td>
<td>0.13</td>
<td>-0.04</td>
<td>-0.12</td>
<td>0.10</td>
<td>0.81</td>
<td>1.00</td>
<td>0.75</td>
<td>0.09</td>
<td>-0.08</td>
<td>-0.32</td>
<td>-0.34</td>
<td>-0.35</td>
<td>-0.33</td>
</tr>
<tr>
<td>R_NRET</td>
<td>0.04</td>
<td>-0.01</td>
<td>0.05</td>
<td>0.14</td>
<td>0.01</td>
<td>-0.18</td>
<td>0.17</td>
<td>1.00</td>
<td>0.75</td>
<td>1.00</td>
<td>0.19</td>
<td>0.04</td>
<td>-0.26</td>
<td>-0.35</td>
<td>-0.29</td>
<td>-0.39</td>
</tr>
<tr>
<td>R_NWH</td>
<td>0.00</td>
<td>-0.03</td>
<td>0.46</td>
<td>0.60</td>
<td>0.08</td>
<td>-0.61</td>
<td>0.76</td>
<td>0.18</td>
<td>0.09</td>
<td>0.19</td>
<td>1.00</td>
<td>0.83</td>
<td>-0.10</td>
<td>-0.44</td>
<td>-0.39</td>
<td>-0.54</td>
</tr>
<tr>
<td>R_POVFA</td>
<td>-0.02</td>
<td>-0.04</td>
<td>0.38</td>
<td>0.37</td>
<td>0.12</td>
<td>-0.56</td>
<td>0.72</td>
<td>0.02</td>
<td>-0.08</td>
<td>0.04</td>
<td>0.83</td>
<td>1.00</td>
<td>-0.20</td>
<td>-0.15</td>
<td>-0.25</td>
<td>-0.25</td>
</tr>
<tr>
<td>Ln_CEN</td>
<td>-0.18</td>
<td>-0.16</td>
<td>0.22</td>
<td>-0.34</td>
<td>0.33</td>
<td>0.15</td>
<td>-0.02</td>
<td>-0.28</td>
<td>-0.32</td>
<td>-0.26</td>
<td>-0.10</td>
<td>0.00</td>
<td>1.00</td>
<td>0.36</td>
<td>0.40</td>
<td>0.48</td>
</tr>
<tr>
<td>Ln_EXP</td>
<td>-0.11</td>
<td>-0.08</td>
<td>-0.17</td>
<td>-0.57</td>
<td>0.21</td>
<td>0.29</td>
<td>-0.32</td>
<td>-0.35</td>
<td>-0.34</td>
<td>-0.35</td>
<td>-0.44</td>
<td>-0.20</td>
<td>0.36</td>
<td>1.00</td>
<td>0.52</td>
<td>0.74</td>
</tr>
<tr>
<td>Ln_FIXED</td>
<td>-0.44</td>
<td>-0.40</td>
<td>-0.18</td>
<td>-0.48</td>
<td>0.12</td>
<td>0.31</td>
<td>-0.25</td>
<td>-0.30</td>
<td>-0.35</td>
<td>-0.29</td>
<td>-0.39</td>
<td>-0.15</td>
<td>0.40</td>
<td>0.52</td>
<td>1.00</td>
<td>0.64</td>
</tr>
<tr>
<td>Ln_SHP</td>
<td>-0.10</td>
<td>-0.05</td>
<td>-0.22</td>
<td>-0.73</td>
<td>0.24</td>
<td>0.38</td>
<td>-0.45</td>
<td>-0.39</td>
<td>-0.33</td>
<td>-0.39</td>
<td>-0.54</td>
<td>-0.25</td>
<td>0.48</td>
<td>0.74</td>
<td>0.64</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Source: Minneapolis-St. Paul (Twin Cities) Metropolitan Council, 2004

Also, as indicated in the following BoxMap comparison by 139 park and ride Thiessen Polygons (Figure 4.1), the high capacity clusters are in close proximity to high usage clusters, and the low usage cluster are frequently due to the low capacity in the area. Not only the values of "Capacity" and "Usage" are quantitatively correlated, but they are spatially associated.

---

1 Usage: the number of parking lots used in the park and ride facility (Used_cap);
2 Capacity: the number of parking lots available in the park and ride Facility (Capacity);
As a consequence, "Capacity" is not an appropriate exploratory variable for predicting the dependent variable - "Usage" in the park and ride demand modeling process.

2. Stepwise Regression Analysis

First remove capacity from the potential independent variable list, due to the high correlation with dependent variable "Usage" (Used_Cap). What variables should be taken into account for successful park and ride system planning? To identify the statistically significant independent variables for predicting the park and ride usage in the Twin Cities Metropolitan Area, this research adopted a statistical approach "Stepwise Regression" to run a list of potential explanatory variables for a best fit model.

The idea of stepwise regression is to develop a best regression model in stages. The list of independent variables is repeatedly searched for variables which should be included in the model. The best explanatory variable is used first, then the second best, and so on. At each step in the stepwise process, the program (SPSS or JMP) must effectively fit a multiple regression model to the variables in the model in order to obtain their F-to-remove statistics, and it must effectively fit a separate regression model for each of the variables not in the model in order to obtain their F-to-enter statistics.
In this process, the input data is 139 park and ride Thiessen Polygons all over entire Twin Cities Metropolitan Area, each of which has all attributes in the table including "Usage" and other potential exploratory variables. While taking the "Usage" as the dependent variable, this research adopt "Mixed Direction" (forward and backward) stepwise regression, which includes the most significant term that satisfies Probability to Enter (0.25) and removes the least significant term satisfying Probability to Leave (0.25), to choose significant variables from the list. It continues removing terms until the remaining terms are significant and then it changes to the forward direction. The input and result are shows as follows:

- **Response**: "Usage" (USED_CAP)
- **Stepwise Regression Control Direction**: Mixed (Forward and Backward)

<table>
<thead>
<tr>
<th>Step</th>
<th>Parameter</th>
<th>Action</th>
<th>&quot;Sig Prob&quot;</th>
<th>Seq SS</th>
<th>RSquare</th>
<th>Cp</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LN_FIXE</td>
<td>Entered</td>
<td>0.0000</td>
<td>749915.1</td>
<td>0.1904</td>
<td>19.262</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>LN_SHP</td>
<td>Entered</td>
<td>0.0014</td>
<td>232122.9</td>
<td>0.2493</td>
<td>10.034</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>R_VEHN</td>
<td>Entered</td>
<td>0.0862</td>
<td>64020.25</td>
<td>0.2655</td>
<td>8.9377</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>POPULATION</td>
<td>Entered</td>
<td>0.0675</td>
<td>71536.46</td>
<td>0.2837</td>
<td>7.4776</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>IHMED</td>
<td>Entered</td>
<td>0.1028</td>
<td>56102.35</td>
<td>0.2980</td>
<td>6.7641</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>LN_CEN</td>
<td>Entered</td>
<td>0.1558</td>
<td>42047.66</td>
<td>0.3086</td>
<td>6.7303</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>D_HH</td>
<td>Entered</td>
<td>0.1482</td>
<td>43288.73</td>
<td>0.3196</td>
<td>6.5365</td>
<td>8</td>
</tr>
</tbody>
</table>
As a result, there are seven steps for selecting independent variables based on the significant probability. The first variable is picked up is the transformed proximity from park and ride facility to nearest fixed guideway segment (LN_FIXED), followed by second best variable: the transformed total proximity from each park and ride facility to all major shopping malls (LN_SHP). Similarly, the third, fourth, fifth, sixth and seventh best variables are chosen into the model.

Compared these seven models, both average household median income (IHMED) and the density of number of households (D_HH) have high coefficient with the third best independent variable – the ratio of number of household without vehicle ownership divided by total number of households in the Thiessen Polygon(R_VEHN). Therefore, R_VEHN is chosen in the model, rather than IHMED and D_HH. The following is the result of the second stepwise fit model with satisfactory variables.

- **Response**: "Usage" (USED_CAP)
- **Stepwise Regression Control Direction**: Mixed (Forward and Backward)

<table>
<thead>
<tr>
<th>Step History</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>
Finally, for predicting park and ride usage (Usage), only five variables are built into the classic multivariate regression model, i.e., the transformed distance from park and ride facility to the nearest fixed guideway (LN_FIXED), the transformed total distance from park and ride to all major shopping malls (LN_SHP), the ratio of the total number of households without vehicle by total households in the park and ride influence areas (R_VEHN), total population (Population), and the transformed distance from centroid of Thiessen Polygon to park and ride facility (LN_CEN).

In summary, this standard least square model can explain proximate 30% park and ride usage by using the five variables. The sum of squared residual is 2.77316e+006, and the standard error estimate of the regression (144.398), with adjustment for a loss in degrees of freedom (133). Also, F-statistic on the null hypothesis (11.185), and the associated probability (<0.0001) are summarized. The whole model is statistically significant at the 5.19216e-009 significance level.

The model equation is as:

\[
\text{Usage} = -2298.5310 + 0.0017 \times (\text{Population} - 13.5578 \times R_{VEHN \_HH}) - 60.9541 \times (\text{Ln \_Fixed}) - 22.8349 \times (\text{Ln \_Cen}) + 188.6492 \times (\text{Ln \_Shp})
\]

- **Actual by Predicted Plot**
Examination Spatial Autocorrelation

After stepwise regression and classic multivariate regression analysis, it is important to examine spatial autocorrelation of the park and ride usage data. If the statistically significant spatial association exists, it is necessary to test the spatial concern within the proposed classic regression model while developing a realistic park and ride demand model for the Twin Cities Metropolitan Area.

This section is to test the research question: "does the usage pattern exhibit complete spatial randomness, clustering, or regularity?" If so, does the spatial autocorrelation need to be one of the variables in the regression model? The procedure for exploratory spatial data analysis, related to spatial autocorrelation, is testified by the adoption of Geoda, a geospatial statistics software. The null hypothesis is the spatial randomness in the usage of park and ride facilities.

In Geoda, a shapefile representing 139 park and ride Thiessen Polygons are with the attribute table indicating dependant variable (Usage) and independent variables (LN_FIXED, LN_SHP, R_VEHN, Population, and LN_CEN). To testify the spatial autocorrelation, this research adopts the queen spatial weight matrix for defining spatial neighbors, meaning that every polygon area which touches the center polygon area is a neighbor either by a corner or a shared boundary. For instance, by definition, Thiessen polygon no. 1 has 6 neighbors, which are no. 135, 134, 88, 47, 120, and 9. The distribution of the weight matrix is as FIGURE 4.2.

FIGURE 4.2 Spatial Weight Matrix Property
1. Global Moran's I

From the Moran Scatterplot, Moran's I statistic for spatial autocorrelation are showed as a regression coefficient (see FIGURE 4.3, Moran's I=0.1049, and p=0.0184). On the vertical axis, a spatial lag is constructed as a weighted average by using the weights in the spatial weights matrix. On the horizontal axis, the value shows the "usage" at each location. Since a positive spatial autocorrelation (SA) is present as a whole, it implies that the observed "usage" at a location and its spatial lag will tend to be similar. Besides, the four quadrants in the Moran Scatterplot can separately indicate the presence of spatial heterogeneity and homogeneity. High-high usage clusters are in the first quadrant (positive SA), low-high usage clusters are in the second quadrant (negative SA), low-low usage clusters are in the third quadrant (positive SA), and high-low usage are in the fourth quadrant (negative SA).

FIGURE 4.3 Moran Scatterplot of Park and Ride Usage

This positive and statistically significant global spatial autocorrelation rejects the null hypothesis, and turn to the alternative hypothesis of the spatial dependence of
park and ride usage distribution in the Twin Cities Metropolitan Area. In other words, like usage values tend to cluster in space, neighbors are similar.

2. Local Clusters

In addition, local measures of spatial autocorrelation are implemented as LISA maps, which suggest outliers or spatial regimes (similar to the use of the Moran Scatterplot). In FIGURE 4.4, the high-high polygons in red (HH) mean high park and ride usage clusters; low-low polygons in blue (LL) mean low usage clusters; high-low polygons in pink (HL) mean the areas with high usage are surrounded by the areas with low usage; low-high polygons in purple (LH) mean the areas with low usage are surrounded by the areas with high usage. Markedly, the HH clusters in the north are near the boundary between Anoka County and Hennepin County and proximate to Interstate 94, and the other HH clusters in the south are within Dakota County, near Interstate 35W and 35E.

FIGURE 4.4 LISA Cluster Map

![LISA Cluster Map](image-url)
In FIGURE 4.5, the LISA map assesses the statistical significance of spatial clustering. The p values of the park and ride usage clustering range from 0.001 (in dark green) to 0.05 (in light green). The Thiessen Polygons in white means not statistically significant.

**FIGURE 4.5 LISA Significance Map**

**Spatial Autoregressive Analysis (SAR)**

Spatial autoregressive model (SAR) provides one possible framework to account for the remaining spatial autocorrelation and provides consistent parameter estimates. Theoretically, it helps to achieve better results by including the spatial context of each observation in the analysis.

Knowing the positive spatial autocorrelation ($I=0.1049$, $p=0.0184$) of park and ride usage exists in the Twin Cities Metropolitan Area, this section intends to review spatial regression diagnostics in Geoda, which allows testing the significance of the residuals resulted from the prediction model. If the residuals (the difference between
observed and predicted values) have spatial dependence, it is violated the rule of spatial randomness so that it is necessary to add in another variable indicating the spatial association in the classic multivariate regression model. If the residuals are without spatial dependence, the errors of the model prediction are reasonable.

In Geoda, once the spatial weights matrix is specified, the spatial regression option is enabled. As showed in FIGURE 4.6, the dependent variable (Usage) and the chosen independent variables (Ln_FIXED, LN_SHP, R_VEHN, Population, and LN_CEN) are according to the previous results of stepwise regression analysis.

**FIGURE 4.6 Regression Window in Geoda**

![Regression Window in Geoda](image)

The regression results give many of the standard diagnostics. The first section shows basic descriptive information, the goodness of fit measures, which are the same as the results of classic multivariate regression analysis. Additionally, spatial dependence test is reported as well.
REGRESSION SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set: 139_OLS
Dependent Variable: USAGE  Number of Observations: 139
Mean dependent var: 79.5899  Number of Variables: 6
S.D. dependent var: 168.344  Degrees of Freedom: 133
R-squared: 0.296016  F-statistic: 11.185
Adjusted R-squared: 0.269551  Prob(F-statistic): 5.19216e-009
Sum squared residual: 2.77316e+006  Log likelihood: -885.354
Sigma-square: 20850.8  Akaike info criterion: 1782.71
S.E. of regression: 144.398  Schwarz criterion: 1800.31
Sigma-square ML: 19950.8
S.E. of regression ML: 141.247

Variable  Coefficient  Std.Error  t-Statistic  Probability

CONSTANT  -2298.531  1140.789  -2.01486  0.0459351
LN_FIXE  -60.95412  10.12652  -6.019256  0.0000000
LN_SHP  188.6492  75.58929  2.495713  0.0137935
R_VEHN_HH  -13.55781  5.26905  -2.573103  0.0111755
POPULATION  0.001699245  0.000755053  2.250497  0.0260591
LN_CEN  -22.8349  14.97499  -1.524869  0.1296658

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 306.1714
TEST ON NORMALITY OF ERRORS
Jarque-Bera  2  1209.971  0.0000000
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
Breusch-Pagan test  5  139.5089  0.0000000
Koenker-Bassett test  5  18.30292  0.0025898
SPECIFICATION ROBUST TEST
White  20  29.75601  0.0739048

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX: Queen_139_OLS.GAL (row-standardized weights)
TEST  MI/DF  VALUE  PROB
Moran's I (error) -0.050336  -0.6248572  0.5320647
Lagrange Multiplier (lag)  1  0.0036519  0.9518126
Robust LM (lag)  1  4.3691370  0.0365955
Lagrange Multiplier (error)  1  0.9803524  0.3221118
Robust LM (error)  1  5.3458375  0.0207718
Lagrange Multiplier (SARMA)  2  5.3494894  0.0689244

COEFFICIENTS VARIANCE MATRIX

<table>
<thead>
<tr>
<th></th>
<th>CONSTANT</th>
<th>LN_FIXE</th>
<th>LN_SHP</th>
<th>R_VEHN_HH</th>
<th>POPULATION</th>
<th>LN_CEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1301400.072270</td>
<td>5677.204622</td>
<td>-85936.795629</td>
<td>-2279.448625</td>
<td>0.029840</td>
<td>5131.289479</td>
<td></td>
</tr>
<tr>
<td>5677.204622</td>
<td>102.546411</td>
<td>-392.156575</td>
<td>-6.494504</td>
<td>0.029840</td>
<td>-25.448579</td>
<td></td>
</tr>
<tr>
<td>-85936.795629</td>
<td>-392.156575</td>
<td>5713.740783</td>
<td>141.394783</td>
<td>-0.000921</td>
<td>-405.091412</td>
<td></td>
</tr>
<tr>
<td>-2279.448625</td>
<td>-6.494504</td>
<td>141.394783</td>
<td>27.762888</td>
<td>-0.002460</td>
<td>2.066075</td>
<td></td>
</tr>
<tr>
<td>0.029840</td>
<td>0.001235</td>
<td>-0.000921</td>
<td>-0.002460</td>
<td>0.000000</td>
<td>-0.003716</td>
<td></td>
</tr>
<tr>
<td>5131.289479</td>
<td>-25.448579</td>
<td>-405.091412</td>
<td>2.066075</td>
<td>-0.003716</td>
<td>224.250272</td>
<td></td>
</tr>
</tbody>
</table>
The overall fit of the model is good and the variables are highly significant. R square is approximate 0.296. The summary section also shows the tests of significance for each parameter. FIGURE 4.7 shows the actual values of park and ride usage versus the ordinary least square (OLS) predicted usage values.

FIGURE 4.7 Quantile Maps of Usage and OLS Predicted Usage Values

Besides, a close examination of regression residuals leads to insight, improvement of models, and further hypotheses. FIGURE 4.8 is a standard deviational map of the regression residuals, illustrating patterns of over- or under-prediction, as well as the magnitude of the residuals.

FIGURE 4.8 Standard Deviational Map of the OLS Residuals
Positive residuals in brown tones indicate the presence of under-prediction, and negative residuals in blue tones indicate model over-prediction. Is there any evidence of residual spatial autocorrelation in this model? To be assessed more rigorously, the residuals have to be tested for the presence of spatial dependence. According to the Moran's statistics (FIGURE 4.9) and the diagnostics for spatial dependence section, there is a negative spatial autocorrelation of residuals (l=-0.0503), and it is significant at the 0.5321 significance level, which means the high residuals tend to be next to low residuals but the spatial dependence is not statistically significant to affect the park and ride usage forecasting model.

In short, this spatial dependence test helps re-verify the reliability of the classic multivariate regression model. It is suggested that the regression assumption of normal and independently distributed residual is not violated, and there is no need to incorporate spatial dimension within this multivariate regression model for predicting successful park and ride usage in the Twin Cities Metropolitan Area.

FIGURE 4.9 Moran Scatterplot of Residuals
CHAPTER V: CONCLUSIONS AND RECOMMENDATIONS

Summary of Research

This research demonstrates the integrated uses of geographic information system (GIS) and statistical analyses for efficiently performing park and ride demand modeling. The methodology enhances the entire modeling process by facilitating the production of thematic maps to summarize data, and incorporating spatial dimension into classical regressive analyses. Proper application of this problem-solving methodology by an individual using appropriate data and analysis techniques could result in significant timesaving and logical evaluation in a visual manner that aids in the communication of data to decision makers and the public alike.

In summary, GIS was found to be helpful for validating model reliability of park and ride facility planning. The utility of GIS significantly reduces the time to calculate the measures between different zonal structures. This makes data more accessible and flexible to users for various analytical purposes. Given the geographical coordinate system, GIS is of the capability for estimating distance between pairs of points, polylines, or polygons. For instances, display of potential park and ride demand factors, such as population, ratio of employment, or distance to major shopping malls, can clearly be used to address the spatial distributions of measures.

In addition, an enhanced function of GIS integrated with regression modeling advances the park and ride modeling application while performing statistical analyses. With respect of spatial association, the examination of spatial dependence makes more rational and logical sense while modeling the demand of park and ride facilities. Outliers and clusters can be identified based on the quantitative measures and geographical location. The spatial dependence can be examined and applied to statistical modeling process. Therefore, the park and ride demand modeling process not only provides various output data that will help quantitatively visualize the
consequences and effects of “what if” questions in park and ride transportation facility policy, but re-think the model liability through spatial dependence test.

**Model Application**

Unlike traditional checklist-like format and guidelines, which may allow for the easy application and ready use of the site location criteria but have uncertainties and variances in cases, this research presents a repeatable methodology relative to geographical statistics for park and ride demand modeling. From the data aggregation, thematic map creation, spatial dependence examination, to spatial regressive analysis, this park and ride demand model aims to enhance the accuracy of the model with respect to the spatial distribution of actual park and ride usage observed in the facilities.

Moreover, since there is a lack of standardized approach to estimate this need, “Park and Ride Site Location Plan” is just released by Minnesota Metropolitan Council in May 2005. The plan also indicates to support three ongoing regional planning process: local comprehensive land use planning, transit support infrastructure planning, and federal transportation funding solicitation. Again, this research reflects the needs of the issues to model the demand and to maintain the growing park and ride system are expected to expend as challenges.

Instead of adopting the model referred to as park and ride demand model in the part of 2030 Transportation Policy Plan, this research argues different approach with the emphasis on the examination of spatial autocorrelation. When validating with the results to the new released Park and Ride Site Location Plan, this GIS-based statistical model corresponds to the results of unmet need areas in the Twin Cities Metropolitan Area. Some differences in “essential” and “preferred” geographic and site attributes are definitely valuable for future review. In addition, this modeling
process by this research has detail discussion in developing GIS database, understanding the relative immediacy of need in different areas around the region, and prioritizing the primary factors influencing the existing park and ride utilization. The process can be adopted by professionals or individuals for future expansion or changes of park and ride systems.

**Limitations and Recommendations**

*Park and Ride Sampling Dilemma*

During the process of this research, only the actual observed park and ride usage in FY2004 data is adopted as dependent variable to model the demand. This is due to a few abandon or new allocated park and ride facilities. The changes would result in different numbers of park and ride influence area by definition in discussion, and would thus make difficulties in sampling and processing. But, it is tested that the trend of park and ride usage is steadily growing over years as a whole.

*Unavailability of User Survey*

There is a lack of specific survey for park and ride users in the Twin Cities Metropolitan Area, such as age, occupation, origin-destination (O-D), travel purposes, main reasons, or concerns for improvement. With advanced qualitative information, this park and ride demand model could be refined in addition to quantitative analysis. Also, it is possible to overcome a few difficulties in simulating the travel impact and scenario test of future changes in park and ride facility planning.

*Mutual Influence of Modes*

The results of this park and ride demand modeling research could assist transportation planners in facility planning and guide the stakeholder to assess future investment. However, this demand modeling research is developed based on the FY2004 park and ride usage data, without the ability to integrating the mutual
influence between bus, light rail, vanpool, and other modes of transportation.

Service Area Refinement Process

A further theory or study to re-define the service area of park and ride facilities in the Twin Cities Metropolitan Area may help to refine the modeling process. The Thiessen Polygons of park and ride facilities are based on relative spatial locations, which can accurately compute the measures from Transportation Analysis Zones (TAZs) or census tracts, but regardless the size of each facility and the direction of travel flow.

Integration of Spatial Autocorrelation

The results of the park and ride demand modeling process provide an estimate of park and ride usage distribution. To better simulate, it is suggested to examine the spatial dependence and integrate to modeling process. Also, the spatial clusters and outliers could be addressed for existing and future park and ride facility planning by identifying unmet or overreached needs.

Future Research

Park and ride facility planning has been incorporated into transportation management and regional travel demand modeling. However, no specific research is conducted for the impact of park and ride faculties on the traffic reduction. If the travel O-D data can be recorded along with other travel characteristics in the regional travel demand modeling, the effect of park and ride facilities would be quantified and applied into traffic assignment process. As a result, impact of downstream traffic from outside commercial business districts can be also accomplished. Furthermore, with the ridership data of light rail transit, LRT systems can be better planed and integrated with park and ride systems and transportation networks.
REFERENCES

• Fotheringham, A. Stewart. 2000. Quantitative Geography Perspectives on Spatial Data Analysis.
• Minnesota Metropolitan Council. 2003. 2000 Travel Behavior Inventory (TBI).
• OCTA, Transportation Case Studies in GIS, U.S. Department of Transportation Federal Highway Administration.
• Office of Metropolitan Planning and Programs. 1998. Case Study 6: GIS for
Transit Planning at Orange County Transportation Authority. Washington, DC.

- Parsons Brinckerhoff Quade and Douglas, Inc. 1995. Park and Ride Demand Estimation Study: Final Report and User’s Manual, King County Department of Metropolitan Services (Metro). Seattle, WA.