MODELING DESTINATION CHOICE FOR SHOPPING

MODELING DESTINATION CHOICE FOR SHOPPING USING A GIS-BASED SPATIO-TEMPORAL FRAMEWORK: AN INVESTIGATION OF CHOICE SET GENERATION AND SCALE EFFECTS

By

SYLVIA YING HE, B.S.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Arts

McMaster University

MASTER OF ARTS (2006)
(School of Geography and Earth Sciences)

McMaster University
Hamilton, Ontario

TITLE: Modeling Destination Choice for Shopping Using a GIS-based Spatio-temporal Framework: An Investigation of Choice Set Generation and Scale Effects

AUTHOR: Sylvia Ying He, B.S. (Sun Yat-sen University)

SUPERVISOR: Dr. Darren M. Scott

NUMBER OF PAGES: xi, 103

Abstract

Incorrect specification of destination choice sets and changes in scale or unit definition can lead to biased parameter estimates and predictions as well as influence findings in statistical tests. Through an empirical study, this thesis evaluates bias as a result of unconstrained destination choice sets and scale effects. Given the deficiency of most current destination choice models, which is the lack of integration of spatio-temporal constraints in generating destination choice sets, the activity-based approach is proposed as a solution by taking into consideration both spatial and temporal constraints in the generation process. Standardized industrial classification (SIC) codes adjoined to shopping opportunities are used to facilitate the discrimination of different shopping types and the classification of shopping stores in order to better understand shopping behaviour.

Analytical results obtained by techniques sensitive to scale effects or zoning effects are likely to change as the aggregation level or area boundary varies. Traffic analysis zones (TAZ) as predefined analysis unit in transportation-related research may not be an optimal choice in the context of destination choice behaviour. Documenting the results on model estimations at different scales and zoning levels is important to investigate the modifiable areal unit problem (MAUP) effects and critically assess the reliability of the estimates. Sensitivity of parameter estimations and model goodness-of-fit between the TAZ system and 10 randomly generated grid systems show remarkable differences. Under a series of criteria, the best zoning system is recommended with certain conditions applied. Our results support the suspicion on the suitability of

predefined analysis units like TAZ and suggest grid systems could be a potential substitution.

The study area in this research spans seven counties of the Louisville MSA. Three primary data sources are used in our analysis: (1) a travel diary survey conducted in 2000 for seven counties of the Louisville MSA; (2) a 2002 Dynamap/Transportation 4.0 network produced by Geographic Data Technology Inc. (GDT); and (3) an urban opportunities data set for the Louisville MSA as geocoded from a database obtained from ReferenceUSA.

A time geography perspective, constrained destination choice set, discrimination of grocery and nongrocery shopping and MAUP effects on destination choice model characterize the contributions of this research to the literature.

Acknowledgements

This thesis has been supported financially by grant 0136818 from the National Science Foundation and grant 261850 from the Natural Sciences and Engineering Research Counsel of Canada, both of which were awarded to the supervisor.

Two years time at Mac is a phenomenal journey, during which I am most grateful for the guidance and encouragement from my supervisor Dr. Darren Scott. His passion in transportation and travel behavior study, his very organized files and his hard working attitude inspired me and kept me motivated. I also want to express my gratitude to Dr. Bruce Newbold and Dr. Antonio Páez for taking their time reviewing my thesis and giving me helpful suggestions, to Dr. Liaw for the useful research methods, to Dr. Pavlos Kanaroglou for his effort in creating such a friendly working environment for graduate students at CSpA, and to Dr. Magee for the interesting lectures.

The best place in the world is wherever your friends are. I like my office because of my officemates and friends: Dora, the buddy who gave me advice on the graduate life at Mac from my first day through the rest two years; Karen and Ruben whose writing and packaging skills helped take me to the land of sunshine, who shared the good coffee time weekdays and weekends; Hejun, Dimitris and Hanna who set a studious example for me; Fei and Kelly with whom I had relaxing extracurricular discussions.

Lastly, my heartfelt appreciation goes to my parents, who let me be independent, let me realize my interest and let me follow my own path. It is their unselfish love that leads me forward.

Table of Contents

List of Tables

List of Figures

Preface

This thesis is presented as a compendium of related manuscripts and consists of the following two chapters:

Chapter 2: An Analysis of Shopping Destination Choice Behavior Using a GIS-based Spatio-temporal Framework

Prepared for submission to *Geographical Analysis*

Chapter 3: Scale Effects on Constrained Destination Choice Models for Shopping

Prepared for submission to *Environment and Planning A*

While journal articles have been co-authored with the research supervisor, the content of each chapter was the sole responsibility of the thesis author. This includes establishing research objectives, reviewing literature, organizing and analyzing the data, computer programming, specifying and estimating models and interpreting results. The supervisor's contributions include data collection, suggestion of research topics and methods, critical evaluation of the manuscripts prior to journal submission, provision of the GIS-based potential path area (PPA) algorithm, editorial advice and discussion of the empirical results.

Chapter 1 Introduction

1.1 Activity-based Approach and Destination Choice Set Generation

The activity-based approach rose as a new development in travel behaviour analysis in the 1970s. The initial research drew heavily on time geography, planning theory and psychology (Jones et al. 1990). Jones et al. (1990) identified seven characteristics of the activity paradigm:

(i) Travel is treated as a derived demand.

(ii) Patterns of behaviour rather than discrete trips are addressed.

(iii) Interdependence and interactions of household members are emphasized.

(iv) Detailed timing, duration of activity and travel are addressed.

(v) Spatio-temporal and inter-personal constraints on travel and location choices are taken into account.

(vi) Interdependencies among events at different times, locations with different participants are considered.

(vii) Household and person categorization schemes based on their various activity demands and constraints are used.

In the last two decades, there has been significant development in the activity-based approach, which is partially credited to technological advancements in data collection. (e.g., starting in the late 1960s, with increased disaggregate data availability, development of econometric discrete choice models has been regarded as a great success in the realm of travel demand analysis and modeling [Pas 1990]). Innovations in computing, the internet and even GPS have been employed in the acquisition of dynamic and accurate activity data. The recording of activity/travel data has experienced an evolution: from the very traditional way of pencil and paper such as in the Portland travel diary survey (Cambridge Systematics 1996) to the laptop computer in the CHASE system (Doherty and Miller 2000), and more recently to the personal digital assistant (PDA) in the EX-ACT survey (Rindsfüser et al. 2003) and the precise location recording by GPS (Stopher 2004).

The largely enriched data have been used in interrelated activity/travel behaviour studies such as activity scheduling (Doherty and Axhausen 1999; Ettema et al. 1993; Ettema et al. 1996), intra-household interactions (Golob and McNally 1997; Scott and Kanaroglou 2002), geographic information systems (GIS) (Buliung and Kanaroglou 2004; Kwan 2000) and travel demand models (Miller et al. 2004; Shiftan and Suhrbier 2002). In terms of travel demand modeling, discrete choice models have been applied under different objectives. The logit model (e.g., Limanond 2005), ordered probit model (e.g., Scott and Kanaroglou 2002) and hazard model (e.g., Bhat et al. 2005) are among those frequently used models in activity-based modeling. With respect to destination/location choice modeling, the logit model (i.e., multinomial logit model) is conventionally adopted (e.g., Landau et al. 1982, Simma et al. 2002).

Prior to the activity paradigm, temporal constraints were ignored in destination choice modeling. Noted as one of the important features of the activity paradigm, by Jones et al. (1990), spatio-temporal constraints on location choice must be taken explicitly into account. Efforts made in this regard are shown by destination/location choice models with travel time included in the destination utility function (Kitamura and

Kermanshah 1984). However, the two-stage probability theory of Manski (1977) indicates that the probability of choosing an alternative from a universal choice set is the sum of the multiplication of the probability of choice set generation in the first stage and the probability of choosing an alternative conditional on the choice set in the second stage. It implies that spatio-temporal constraints should not only be included in the destination choice, but also in the generation of a destination choice set. An incorrect choice set (i.e., an assumption that a universal choice set is applied to every individual) is a form of model misspecification (Swait and Ben-Akiva 1987). It introduces bias in parameter estimates and in prediction (Williams and Ortuzar 1982; Landau et al. 1982; Thill 1992; Thill and Horowitz 1997) and results in erroneous interpretation of travel behaviour (Thill 1992; Pellegrini et al. 1997).

As a solution, the activity-based approach can be applied in the generation of a destination choice set within a spatio-temporal framework. The destination choice set for a given activity is indeed embedded in the individual's space-time prism, a concept introduced in time geography (Hägerstrand 1970), and can be extracted from it (Thill 1992). The potential path area (PPA), the projection of potential path space (PPS) on planar space, reflects how far an individual can reach given a time budget. Figure 1.1 provides an example of a PPA for shopping. Opportunities within the PPA are considered the destination choice set. A precise destination choice set relies heavily upon an algorithm capable of deriving an accurate PPA. Traffic network-based PPA algorithms have the virtues of linking discrete trips as a whole and taking account of link-specific traffic conditions (Kwan and Hong 1998; Scott 2006). Individual destination choice sets generated by such GIS-based approaches are characterised by the integration of real travel costs and individual time budgets in the choice set generation process.

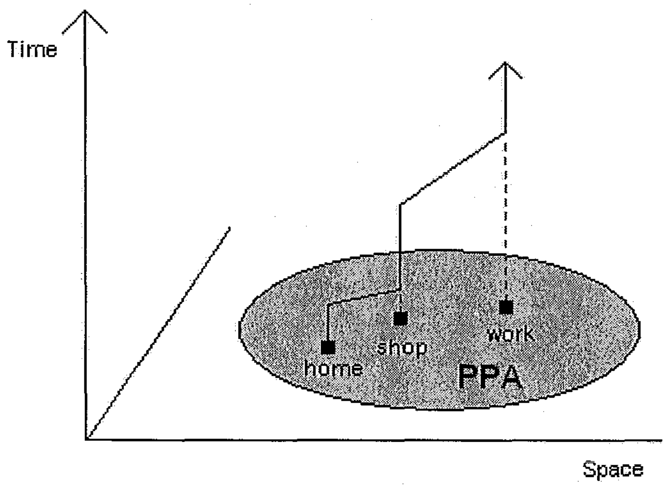


Figure 1.1: Example of an individual's potential path area (PPA) for shopping.

1.2 Research Objectives

Despite the stress by previous researchers that the incorrect specification of a choice set will lead to inevitable biased parameter estimates and predictions (Williams and Ortuzar 1982; Landau et al. 1982; Thill 1992; Thill and Horowitz 1997; Pellegrini et al 1997; Kwan and Hong 1998; Arentze and Timmermans 2005), the degree of inaccuracy resulting from an unconstrained choice set has not been systematically explored. In this vein, the first objective of this thesis is to explore the difference in the parameter estimations and predictions between unconstrained and constrained destination choice sets for shopping. Constrained destination choice sets in this research are generated via a geographic information system (GIS) using Scott's (2006) PPA algorithm.

Due to the high frequency compared to other activity types and the potential impact on business districts and transport planning, shopping behaviour is of particular interest to researchers in many fields such as business, marketing and transportation. Research of store choice has focused mainly on grocery shopping and there has been very little empirical research on nongrocery shopping (Fox et al. 2004). In an attempt to overcome this shortcoming, standardized industrial classification (SIC) codes adjoined to shopping opportunities are used to facilitate the discrimination of different shopping types (i.e., grocery, nongrocery) and the classification of shopping stores (e.g., department stores, retail bakeries, etc.). Use of SIC codes makes it possible to achieve our second objective, which is to study grocery and nongrocery shopping behaviour with regard to specific types of shopping opportunities.

Most data for transportation research are collected and disseminated based on traffic analysis zones (TAZs). A shopping destination choice model based on the zonal level would therefore be more meaningful in terms of its application in travel demand modeling. Moreover, store locations are often subject to change because of external (i.e., fire) and internal forces (i.e., store revenue). When thousands of stores are considered as possible destinations, the modeling task would be simplified if the stores are aggregated to zones. As a matter of fact, zones are often used when modeling activity destinations in activity-based research (Buliung 2004). In our research, we adopt a zonal based model. While a constrained choice set is generated to improve efficiency in parameter estimation and prediction, the underlying zoning system used to aggregate shopping opportunities becomes a concern since the extent of improvement in parameter estimations and prediction is not only dependent on a constrained choice set, but also on the zoning system for which shopping models are estimated. Grounded in social science (Gehlke and Biehl 1934, Robinson 1935), the modifiable areal unit problem (MAUP) is a potential source of error as a result of varied unit definition and area boundaries at different zonal levels. It triggers the third objective of this thesis, which is to evaluate model performance as well as the sensitivity of destination attractiveness variables (i.e., SIC variables) and temporal variables (i.e., travel cost and activity duration) to 11 zoning schemes, including the conventional TAZ system and 10 randomly generated grid systems (i.e., grid size of 1 km^2 to 10 km^2 in 1 km^2 intervals).

1.3 Thesis Contents

Including the introduction, this thesis consists of four chapters. The research makes use of three primary data sources: (1) a travel diary survey conducted in 2000 in seven counties of the Louisville KY-IN MSA, which contains household, personal and one-day travel information; (2) a 2002 Dynamap/Transportation 4.0 street network produced by Geographic Data Technology Inc. (GDT); and (3) an urban opportunities data set for the Louisville MSA as geocoded from a database obtained from ReferenceUSA.

Chapters 2 and 3 report on two possible inaccuracy in shopping destination choice modeling. With a description of the grocery, nongrocery and general shopping model specifications, Chapter 2 focuses on the comparison between unconstrained and constrained choice sets. Two objectives are achieved: first, the explicit incorporation of space-time constraints in destination choice sets; and second, the evaluation of both grocery and nongrocery shopping behaviour using SIC codes. From a time-geographic perspective, this chapter is the first attempt to incorporate a GIS-based space-time constrained choice set into a conventional destination choice model and compare estimates between constrained and unconstrained models. Results are presented with regard to the explanatory power and predictive ability of the models as well as parameter estimations.

Doubting the suitability of the TAZ system in shopping destination choice modeling, Chapter 3 focuses on a comparison of constrained model estimations between the TAZ system and 10 randomly generated grid systems. The objective of Chapter 3 is to evaluate the sensitivity of parameters and model performance as the zoning scheme changes. Shopping opportunities in individuals' constrained choice sets are aggregated to the TAZ system and the 10 grid systems respectively based on which shopping models are estimated. Scale effects are investigated through a series of assessments, including size of choice sets, parameter estimations and goodness-of-fit. In the summary, suitability of TAZ and grid systems is discussed through which the best zoning system is recommended under particular circumstances.

The thesis concludes in Chapter 4 with a review of the motivation and contributions of this research, leading to an open discussion of future studies.

1.4 References

Arentze T.A. and Timmermans H. 2005, An analysis of context and constraints-dependent shopping behaviour using qualitative decision principles, *Urban Studies*, vol. 42, pp. 435-448.

Bhat C.R., Srinivasan S., and Axhausen K.W. 2005, An analysis of multiple interepisode durations using a unifying multivariate hazard model, *Transportation Research Part B*, vol. 39, pp. 797-823.

Buliung R.N. 2004, *Data organization and the design of measures for exploration and analysis of household activity/travel behaviour,* Ph.D. Dissertation, McMaster University, Hamilton, ON.

Buliung R.N. and Kanaroglou P.S. 2004, On design and implementation of an object-relational spatial database for activity/travel behaviour research, *Journal of Geographical Systems*, vol. 6, pp. 237-262.

Cambridge Systematics 1996, *Data collection in the Portland, Oregon metropolitan area: case study*, U.S. Department of Transportation, DOT-T-97-09.

Doherty S.T. and Axhausen K.W. 1999, The development of a unified modeling framework for the household activity-travel scheduling process, in *Traffic and Mobility: Simulation-Economics-Environments*, Brilon W., Huber F., Schreckengerg M. and Wallentowitzpp H., eds., Springer, Berlin, pp. 35-56.

Doherty S.T. and Miller E.J. 2000, A computerized household activity scheduling survey, *Transportation*, vol. 27, pp. 75-97.

Ettema D., Borgers A., and Timmermans H. 1993, Simulation model of activity scheduling behaviour, *Transportation Research Record*, vol. 1413, pp. 1-11.

Ettema D., Borgers A., and Timmermans H. 1996, SMASH (simulation model of activity scheduling heuristics): some simulations, *Transportation Research Record*, vol. 1551, pp. 88-94.

Fox E.J., Montgomery A.L., and Lodish L.M. 2004, Consumer shopping and spending across retail formats, *Journal of Business*, vol. 77, pp. 25-60.

Gehlke C.E. and Biehl K. 1934, Certain effects of grouping upon the size of the correlation coefficient in census tract material, *Journal of the American Statistical Association*, vol. 29, pp. 169-170.

Golob T.F. and McNally M.G. 1997, A model of activity participation and travel interactions between household heads, *Transportation Research Part B*, vol. 31, pp. 177-194.

Hägerstrand T. 1970, What about people in regional science?, *Papers of the Regional Science Association*, vol. 24, pp. 7-21.

Jones P., Koppleman F., and Orfeuil J.-P. 1990, Activity analysis: state-of-the-art and future directions, in *Developments in Dynamic and Activity-Based Approaches to Travel Analysis*, Jones P., ed., Billing & Sons Ltd, Worcester, pp. 34-55.

Kitamura R. and Kermanshah M. 1984, Sequential model of interdependent activity and destination choices, *Transportation Research Record*, vol. 987, pp. 81-89.

Kwan M.-P. 2000, Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set, *Transportation Research Part C*, vol. 8, pp. 185-203.

Kwan M.-P. and Hong X.-D. 1998, Network-based constraints-oriented choice set formation using GIS, *Geographical Systems*, vol. 5, pp. 139-162.

Landau U., Prashker J.N., and Alpern B. 1982, Evaluation of activity constrained choice sets to shopping destination choice modeling, *Transportation Research Part A*, vol. 16, pp. 199-207.

Limanond T., Niemeier D.A., and Mokhtarian P.L. 2005, Specification of a tour-based neighborhood shopping model, *Transportation*, vol. 32, pp. 105-134.

Manski C.F. 1977, Structure of random utility models, *Theory and Decision*, vol. 8, pp. 229-254.

Miller E.J., Hunt J.D., Abraham J.E., and Salvini P.A 2004, Microsimulating urban systems, *Computers, Environment and Urban Systems*, vol. 28, pp. 9-44.

Pas E. 1990, Is travel demand analysis and modeling in the doldrums? in *Developments in Dynamic and Activity-Based Approaches to Travel Analysis*, Jones P., ed., Billing & Sons Ltd, Worcester, pp. 3-27.

Pellegrini P.A., Fotheringham A.S., and Lin G. 1997, An empirical evaluation of parameter sensitivity to choice set definition in shopping destination choice models, *Regional Science*, vol. 76, pp. 257-284.

Ringsfüser G., Muhlhans H., Doherty S.T., and Beckmann K.J. 2003, Tracing the planning and execution of activities and their attributes - design and application of a hand-held scheduling process survey, in *the 10th International Conference on Travel Behaviour Research*, Lucerne, Switzerland, August 10-15.

Robinson W.S. 1935, Ecological correlation and the behaviour of individuals, *American Sociological Review*, vol. 15, pp. 351-357.

Scott D.M. 2006, Constrained destination choice set generation: A comparison of GIS-based approaches, *85th Annual Meeting of the Transportation Research Board*: Compendium of Papers CD-Rom, Washington, D.C., January 22-26.

Scott D.M. and Kanaroglou P.S. 2002, An activity-episode generation model that captures interactions between household heads: development and empirical analysis, *Transportation Research Part B*, vol. 36, pp. 875-896.

Shiftan Y. and Suhrbier J. 2002, The analysis of travel and emission impacts of travel demand management strategies using activity-based models, *Transportation*, vol. 29, pp. 145-168.

Simma A., Schlich R., and Axhausen K.W. 2002, Destination choice modeling of leisure trips: the case of Switzerland, in *Monitoring and Management of Visitor Flows in Recreational and Protected Areas*, Arneberger A., Brandenburg C., and Muhar A., eds., Universität für Bodenkultur, Wien, pp. 150-158.

Stopher P.R. 2004, GPS, location, and household travel, in *Handbook of Transport Geography and Spatial Systems*, Hensher D.A., Button K.J., Haynes K.E., and Stopher P.R., eds., Elsevier Ltd, London, pp. 432-449.

Swait J.D. and Ben-Akiva M. 1987, Incorporating random constraints in discrete models of choice set generation, *Transportation Research Part B*, vol. 21, pp. 91-102.

Thill J.C. 1992, Choice set formation for destination choice modeling, *Progress in Human Geography*, vol. 16, pp. 361-382.

Thill J.C. and Horowitz J.L. 1997, Travel-time constraints on destination choice sets, *Geographical Analysis*, vol. 29, pp. 108-123.

Williams H.C.W.L. and Ortúzar J. de D. 1982, Behavioural theories of dispersion and the misspecification of travel demand models, *Transportation Research Part B*, vol. 16, pp. 167-219.

Chapter 2 An Analysis of Shopping Destination Choice Behaviour Using a GIS-based Spatio-temporal Framework

2.1 Introduction

Shopping trips account for a substantial proportion of total trips in the United States (Nelson and Niles 2000). The accurate prediction of shopping destination is important for travel demand forecasting. Beginning with early studies in shopping destination choice, accessibility (i.e., travel time) has been identified as a critical factor influencing choice and has been included in such models (Recker and Kostyniuk 1978; Koppleman and Hauser 1978; Landau et al. 1982; Kitamura and Kermanshah 1984; Timmermans 1996). However, we argue that for an accurate prediction of destination choice, accessibility should not only be included in a model as a determinant, but should also be incorporated in the choice set definition. Nonetheless, due to lack of information, most of the current studies on destination choice have not taken into account accessibility when defining the choice set. Rather, two assumptions are frequently made: one is to assume that an individual has access to a universal choice set, which contains all destinations in the entire study area (Timmermans et al. 1984; Fotheringham and Trew 1993; Thill and Horowitz 1991); another is to assume individuals living in the same neighborhood are faced with the same choice set. Depending on where their neighborhood is, the choice set has some degree of variation (Miller and O'Kelly 1983). A choice set based on either one of the above assumptions is imprecise.

An incorrect choice set is a form of model misspecification (Swait and Ben-Akiva 1987). The model assigns non-negative probabilities to all the alternatives in the universal choice set including those not in an individual's true choice set. As a result, it introduces biases in parameter estimation of the utility function and inaccurate prediction (Williams and Ortuzar 1982; Landau et al. 1982; Thill 1992; Thill and Horowitz 1997; Pellegrini et al. 1997; Kwan and Hong 1998; Arentze and Timmermans 2005) and subsequently leads to erroneous interpretation of human behaviour (Thill 1992; Pellegrini et al. 1997). Pellegrini et al. (1997), for example, adopted a baseline of 14 stores, and compared it with choice sets from 3 to 13 stores. A multinomial logit model was used to estimate a separate destination choice model for each choice subset for five market segments based on income and race. The parameters were found to be quite sensitive to the choice set specification, and the sensitivity was not homogeneous across all variables. In this regard, restrictions (i.e., time budget, socioeconomic status and cognitive perception) should be imposed to retrieve less arbitrary choice sets. Amidst various approaches to constrain a choice set, the space-time prism approach is attractive for its strong behavioural base (Thill 1992). Virtues of this approach are also discussed in other research (Kwan and Hong 1998; Arentze and Timmermans 2005; Scott 2006).

The contribution of this study lies in achieving two objectives. The primary objective of this paper is the explicit incorporation of space-time constraints in the derivation of destination choice sets. This approach in choice set generation is unique in that it takes into account the spatial distribution of shopping opportunities, realistic travel conditions, including link-specific travel cost, as well as an individual's time budget. In this research, we employ Scott's (2006) GIS-based algorithm for deriving network-based constrained destination choice sets. The constraint-oriented choice sets derived from the space-time prism algorithm can be implemented compatibly with the standard discrete choice models and maximum likelihood estimation (Thill 1992). Although such compatibility has been noted for a long time, there is hitherto no implementation of the space-time prism constrained choice set to the conventional destination choice model. This paper represents a first attempt to incorporate a GIS-based space-time constrained choice set into a conventional destination choice model and compare the model estimations between a constrained and an unconstrained model. The temporal effect in this paper has been carefully considered in the specification of the destination choice sets. The temporal variables are computed using two GIS software packages: ArcGIS and TransCAD.

The second objective of this paper is to evaluate both grocery and nongrocery shopping behaviour. Previous shopping destination choice studies have focused primarily on grocery, and there has been very little empirical research on nongrocery shopping (Fox et al. 2004). A possible reason for ignoring nongrocery shopping trips is that previous travel diary surveys do not contain enough detailed information regarding trip purpose and destination, which impedes modelers from distinguishing nongrocery shopping trips from grocery shopping trips. By tagging geocoded trips with their locations which have standard industrial classification (SIC) codes, discrimination of different types of shopping trips becomes possible. The breakdown of shopping opportunities to specific store types by SIC code provides a disaggregate way to study how shoppers behave in response to various types of shopping opportunities through different models.

When preparing the data, special consideration is given to two situations. First, since most disaggregate information on destination choice corresponds to shopping stores, many studies (especially in marketing, retailing and consumer services studies) focused on points as shopping destinations (Bell et al. 1998; Fox et al. 2004; Lloyd and Jennings 1978; Guy 1985; Richards and Ben-Akiva 1974; Timmermans 1981, 1996) rather than on zonal as destinations (Landau et al. 1982; Limanond et al. 2005). In transportation studies, however, most data are collected and disseminated based on traffic analysis zones (TAZs). A shopping destination choice model based on zones would therefore be more meaningful in terms of its application in travel demand forecasting. Moreover, store locations are often subject to change because of external (i.e., fire accident) and internal forces (i.e., store revenue). In an urban area, when thousands of stores are considered possible destinations, the modeling task would be simplified if the stores are aggregated to zones. As a matter of fact, zones are often used when modeling activity destinations in activity-based research (Buliung 2004). In our research, we adopt a zonal based model. The analysis uses 818 TAZs in seven counties of the Louisville KY-IN MSA as the alternatives for destination choice. The second consideration is, in the real world, people often undertake multipurpose trips (Ewing et al. 1994). Previous shopping experience and subsequent shopping choice could affect current shopping choice (Tardiff 1979; Miller and O'Kelly 1983; Tang et al. 2001; Arentze et al. 1993). Even for simple examples, the chain of conditional activity participation could be quite complicated (Barnard 1987). To avoid this issue, only single purpose trips are selected from the original travel survey data for our research.

The remainder of this paper is organized as follows. Section 2 reviews literature relating to choice set formation. Section 3 describes the study area and the process of choice set generation, as well as the research design. Section 4 presents empirical results from unconstrained and constrained models for three types of shopping—general, grocery, nongrocery. Coefficients and goodness-of-fit are compared. Conclusions are found in the final section, along with a summary of the contribution to the study of shopping destination choice behaviour.

2.2 Literature Review

2.2.1 Probabilistic Choice Rule and Constrained Choice Sets

From the analyst's perspective, the decision of choosing a shopping destination is a probabilistic issue. The two-stage probability theory of Manski (1977) has been widely referred to in previous studies of discrete choice modeling (Swait and Ben-Akiva 1987, Thill and Horowitz 1997, Pellegrini et al 1997). The probability of choosing an alternative j from a universal choice set $P(j)$ is the sum of the multiplication of the two parts of probability, namely, the probability of choice set generation $P(C)$ in the first stage and the probability of choosing an alternative conditional on the choice set $P(j|C)$ in the second stage.

$$P_n(j) = \sum_{C \in G} P_n(j|C) P_n(C) \tag{1}$$

where C is a random choice set of a universal choice set G. The probability of choosing an alternative would depend directly on the definition and generation of the choice set. In the following sub section, a number of geographic approaches in the time geography framework to delineate the choice set are presented. Different from conventional models where C is a random choice set of a universal choice set G, time geography takes into consideration individual activity scheduling and therefore the choice set C defined in this way is deterministic rather than random. Thus the above model is simplified as $P_n(j) = P_n(j|C)$ from a time geography perspective.

2.2.2 Time Geography Perspective

The concept of the space-time prism plays a key role in destination choice set specification was around for long time. Hägerstrand's (1970) time geography framework centers around the spatial and temporal constraints on the movement of individuals. Pred (1977) believed the most distinctive feature of this framework lies upon its ability to treat both individual and society as a whole and its major focus on "the various types of constraint and finitude which wall-in the action alternatives of individual." According to the ease to reschedule or relocate, activities can be divided into fixed and flexible activities that refer to mandatory and discretionary activities, respectively (Hägerstrand 1970; Landau et al 1982; Miller 2004, 2005; Arentze and Timmermans 2005). Between successive fixed activities, space-time prisms are determined. The size of the prism, potential path space (PPS) projected on a planar space is the potential path area (PPA). Opportunities within the PPA are considered the destination choice set.

The space-time prism as a means of measuring accessibility is not new. In the 1970s, various space-time measures were employed to better the representation of accessibility by taking account of individual activity scheduling (Lenntorp 1978, Burns

1979). Lenntorp (1978) used the number of paths in an activity schedule as a surrogate for accessibility, accounting for transportation characteristics like the location, operating hours of activities and a hypothetical activity schedule. Burns (1979) proposed a space-time prism with changes in velocities and time constraints of an individual, and the volume of the prism was calculated. The attempt to incorporate physical transport network conditions, non-uniform travel speed across the network and different travel modes in accessibility assessment was a milestone in the exploration of time geography.

Since the 1980s, approximations of the space-time prism have been incorporated in a number of destination choice models (Kitamura and Kermanshah 1984; Landau et al. 1982; Arentze and Timmermans 2005). Kitamura and Kermanshah (1984) developed a destination choice model and an activity choice model for four types of activities in the time geography framework and stressed temporal factors, specifically, time-of-day dependencies of activities and trips. They looked into the interdependency in multiple-sojourn trip chains and implied home location is of critical importance in the trips based on the fact that people sooner or later will return home after participating in out-of-home activities. The available time for out-of-home activity and travel was restricted by the individual time budget, which was the time period between current time and the time needed to return home.

To demonstrate how a time-constrained choice set can be incorporated in a destination choice model and how it improves the prediction accuracy, Landau et al. (1982) imposed a constrained choice set (an average of 27.9 zones compared to 35 zones) to a shopping destination choice model for all observations. Five categories of information were needed to obtain the constrained choice set, including 1) classification of shopping activities for workers and non-workers with regard to how the shopping trips are related to home/school; 2) activity scheduling constraints; 3) individual's time budget; 4) transportation limitations including travel mode and travel time; 5) store locations and hours. This information was used to calculate the "maximum possible net activity duration" (MPNAD) by subtracting travel time from an individual's time budget, taking account of store locations and hours. The authors argued that while there is a minimum amount of time an individual needs to spend in stores, this cannot be measured accurately, hence it was assumed to be zero. Thus, they determined the choice set as a set of alternatives where MPNAD is larger than zero. Once the choice set was determined, the probability of choosing an alternative was predicted using a random utility choice model. The constrained choice set model improved modestly the predictive accuracy of the traditional unconstrained destination choice model, which the authors believed could be improved by further restricting the choice sets. Despite the modest improvement, the activity constrained method had effectively avoided situations where a highly attractive yet distant destination was assigned a significant probability (Landau et al. 1982).

The space-time prism approach to define a choice set was combined with a decision tree approach by Arentze and Timmermans (2005) to simulate shopping destination choice behaviour. Similar to the approach used in Landau et al.'s (1982) study, the choice set is defined as all locations that are reachable within the space-time prism with stores' opening hours applied. Several assumptions were made in their study, such as the location and start-times of mandatory activities, and the position of the discretionary activities in the schedule. Once the choice set was defined, a decision-tree induction method classification and regression trees (CART) was applied to derive the best data-fitting decision-tree that was used in turn to handle a large number of potential variables. The results revealed that a large variation exists in the choiceset size across the total sample. Since the impact of constraints from the individual activity schedule on the shopping location choice was considerable, they highlighted the weakness of discrete choice models currently used for demand forecasting and analysis at store or shopping center levels and suggested that spatial shopping models used for analysis or prediction should be based on activity-constrained choice sets at an individual level (Arentze and Timmermans 2005).

Although the concept of the space-time prism was introduced in as early as the 1970s, it was not until the 1990s that a more spatially and temporally restrictive prism (or destination choice set) become operational. Since the 1990s, geographic information systems (GIS) have become more popular in a wide range of research and have demonstrated their suitability of being a tool to study space-time prism in terms of their great data storage and analysis capacity. With the arrival of the GIS paradigm, the time-geographic approach has become more operational, and it is easier to gain hands-on experience in destination choice set modeling (Thill 1992). Miller (1991) developed an algorithm for implementing a space-time prism in a GIS environment. He defined the basic space-time prism concepts and operational definitions such as travel times through network nodes, turn times and stop times. Based on previous work, the idea that travel times vary across the network and fluctuate with time was proposed. Assuming that the shortest path through the network would be selected by individuals, two network operations were carried out in his algorithm. First, all shortest paths from a particular origin up to a cumulative impedance limit along each path were calculated. Secondly, the shortest path from an arc to a particular destination up to a variable cumulative impedance, which in addition requires to test whether the path from an arc to the destination is achieved within its impedance limit, was calculated. Miller's algorithm is built on an assumption that the shortest path through the network would be selected by individuals.

Extending the work of Miller (1991), Kwan and Hong (1998) applied the concept of PPA to obtain the constrained destination choice set. The segment-specific travel speed represents the realistic travel environment which overcomes the limitation of mathematical and geometrical methods that assume uniform mobility and travel speed across the study area. Based on the traffic network, Kwan and Hong (1998) used GIS to derive the time constrained general urban opportunity choice set onto which an individual's cognitive map is imposed to obtain the narrowed choice set. Two important aspects of the cognitive environment of the individual were considered including the spatial knowledge of the various areas of the city and location preference. The authors argued that the places people selectively go to are not just dependent on their knowledge of the physical environment *per se*, but also their own preference to one place over another. It is also shown that the size of PPA derived through geometric methods may not be proportional to the number of opportunities due to the fact that urban opportunities are not evenly distributed across space (Kwan and Hong 1998).

Recently, Scott (2006) suggested a shortest path approach for estimating the potential path area. Instead of using the overlap of discrete representations of space defined by time intervals as used by Kwan and Hong (1998), Scott's algorithm is based on a continuous representation of space that is realized by a thorough search of links in the traffic network. More specifically, instead of identifying links from the intersection of service areas around origins and destinations, he selected a subset of links for possible inclusion in the PPA and then examined specific links in the subset to determine whether the links should be included in the PPA. By comparing the sizes of potential path areas, Scott's algorithm was able to determine a greater number of urban opportunities than that calculated by Kwan and Hong's (1998) algorithm for several time intervals, which evinced the weakness of their overlay approach in defining a choice set.

A limitation of Arentze and Timmermans' (2005) study is partially attributed to the applied CART system which recursively splits a set of explanatory variables into homogeneous partitions, including splitting certain types of continuous variables into discrete variables such as activity duration (short, medium and long) and the need for two assumed levels of maximum travel time (5 and 10 minutes) to define a choice set, which makes the destination choice set not a precisely time-constrained one. Kwan and Hong's (1998) algorithm to define the destination choice set is flawed for its underlying concept of a discrete time axis, which splits travel time into two parts, one for the origin and the other for the destination. The union of areas reachable within the allocated time is defined as one possible choice set. However, splitting travel time in a number of combinations and uniting all possible choice sets is not an exhaustive way to retrieve all the opportunities due to the fact that there are unlimited combinations of splitting travel time. In contrast, Scott's (2006) algorithm takes into account the origin and destination at the same time and retrieves one deterministic choice set by an exhaustive search of all possible links to be included in the PPA. He shows empirically the deficiency of Kwan and Hong's (1998) algorithm by enumerating the number of urban opportunities that their overlay method failed to capture.

The network-based generation of a restricted choice set is only the first step in the full formulation of a choice set generation model. The next stage would be to include a random utility function and a choice probability distribution (Kwan and Hong 1998). According to Thill (1992), the destination choice set for a given activity is actually embedded in the individual's space-time prism and can be derived straightforwardly by categorizing the destination in the prism selectively. The potential activity path can be elicited for each individual in a simulation scheme, and a deterministic choice set is accordingly obtained, which can be incorporated into a conventional choice model. Its main challenges lie in the substantial computational effort, the amount of detailed diary data and the behavioural soundness of the simulation scheme (Thill 1992). Because of these plausible challenges, although a number of GIS-based space-time prism approaches to delineate constrained choice sets have been developed, the delineated choice sets have not been incorporated into conventional choice models like the multinomial logit model to investigate the difference in model estimations and predictions between the constrained and unconstrained models. This gap in destination choice modeling triggers the motivations of our research.

2.3 Research Design and Data

2.3.1 Study Area and Data

The study area in this research spans seven counties of the Louisville MSA (Figure 2.1).

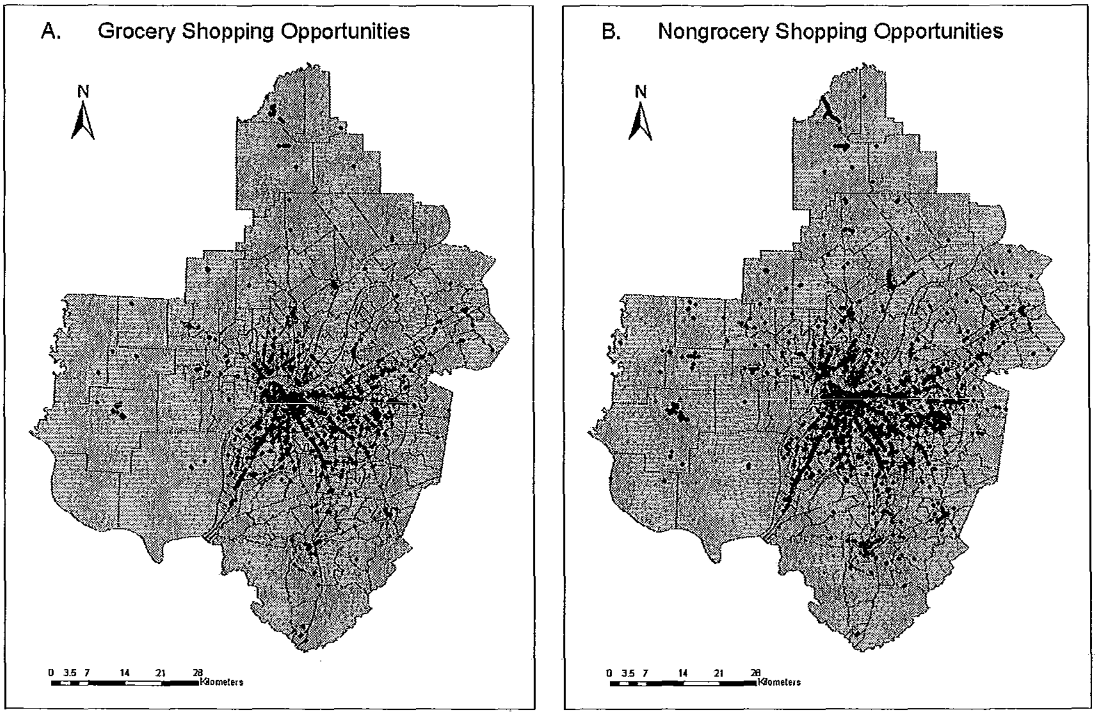


Figure 2.1: Distribution of (A) grocery and (B) nongrocery shopping opportunities in seven counties of the Louisville MSA, 2000.

Three primary data sources are used in our analysis:

(1) a travel diary survey conducted in 2000 for seven counties of the Louisville MSA, which contains household, personal and one-day travel information. Approximately 4600 households participated in the survey, where a total of 30,888 trips were conducted.

(2) a 2002 Dynamap/Transportation 4.0 network produced by Geographic Data Technology Inc. (GDT). The traffic network contains traffic network information, such as direction, length, speed and cost.

(3) an urban opportunities file for the Louisville MSA as geocoded from a database obtained from ReferenceUSA. The file contains 34,440 opportunities. Each opportunity is classified by SIC code.

2.3.2 PPA Algorithm

All the shopping opportunities are broken down into specific categories using the SIC code. The shopping opportunities information is then aggregated to the TAZs (Figure 2.2). We applied Scott's (2006) PPA algorithm to generate the network-based constrained destination choice set (network PPA) (Figure 2.3A). The algorithm selects a subset of links to make sure the origin and destination can be reached within a time budget, then it searches thoroughly the links from the subset to determine which links are to be included in the PPA. A program in ArcMap was revised to select the TAZs that intersect with the network PPAs (Figure 2.3B) in order to turn the network PPAs into TAZ based PPAs (Figure 2.3C). In this way, the polygon PPAs become a time-constrained choice set for shopping. The size of the choice set varies since the geographic location of the mandatory activity places before and after the shopping activity, the individual's time budget and the free-flow speed of links in the road network differ from each other.

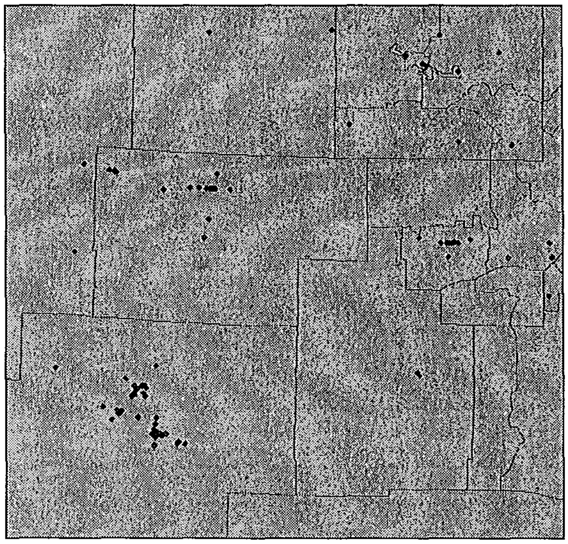


Figure 2.2: Shopping opportunities aggregation.

2.3.3 Choice Set Generation[1]

Choice sets are generated for three types of shopping: grocery, nongrocery and general shopping. An individual's choice set is formed by the actual chosen zone plus nine random non-chosen zones within the PPA. A random sampling program is written in MATLAB to generate the choice set. The sample size is 428 PPAs for grocery shopping, 328 PPAs for non-grocery shopping and 616 PPAs for all shopping[2]. For the unconstrained situation, the choices contain the actual chosen zone plus nine randomly selected zones that could be any nine of the 818 non-chosen TAZs. As for the constrained situation, the nine randomly selected zones could be any nine of the non-chosen zones in an individual's constrained choice set.

[1] The random sampling approach adopted requires that in a PPA there are at least ten zones containing appropriate shopping opportunities.

[2] As some of the shopping PPAs can be counted as both grocery and nongrocery, the sum of grocery (428) and nongrocery (328) PPAs is more than the number of general shopping PPAs (616).

A. Network PPA **B.** Network PPA intersected with TAZs **C.** Grid PPA based on TAZs

Figure 2.3: Derivation of a grid PPA from a network PPA.

2.3.4 Model and Specification

Disaggregate discrete choice models are based on discrete consumer choice observed in real markets and random utility theory. The best-known of them is the multinomial logit (MNL) model (McFadden 1974; Ben-Akiva and Lerman 1985; Barnard 1987; Timmermans et al. 1991). Many previous studies estimated destination choice parameters using the MNL (Recker and Kostyniuk 1978; Kitamura and Kermanshah 1984; Koppleman and Hauser 1978; Pellegrini et al. 1997; Miller and O'Kelly 1983; Timmermans et al. 1984). Its popularity resides in the fact that econometric specification of the model is obtained directly from the utility maximization framework (McFadden 1974). Even when there is a large choice set, the logit model can still be estimated consistently by using smaller choice sets that include the actually chosen alternative and randomly selected alternatives from the full choice set under the irrelevant independent alternatives (IIA) property assumption (McFadden 1978). As applied in transport research, the utility of taking part in an activity at a location depends on the attractiveness of that location, the travel time and activity time (Ettema and Timmermans 2004; Miller 1999).

The utility of destination j for individual n is specified as:

$$U_{jn} = V_{jn} + \varepsilon_{jn}$$
$$V_{jn} = \sum_{k} \beta_k X_{kj} \qquad \forall j \in C_n,\ \forall C_n \in C \tag{2}$$

where X_{kj} is a matrix of the destination attributes and temporal factors, β_k is a vector of coefficients of the zonal characteristics and temporal factors for each zone, and ε is the disturbance and it is assumed independent. C_n is the constrained choice set for individual *n*. An individual will choose destination j if and only if

$$U_{jn} \geq U_{ln} \qquad \forall j, l \in C_n \tag{3}$$

The probability that an individual chooses destination j conditional on the choice set is:

$$P_n(j|C_n) = \frac{e^{V_{jn}}}{\sum_{l \in C_n} e^{V_{ln}}} \tag{4}$$

Two types of destination attributes are considered for the shopping destination choice models developed for this research: number of shopping opportunities (Table 2.1) and temporal factors. This study decomposes shopping activities into two categories: grocery, and non-grocery. Thirty one dummies corresponding to four-digit SIC codes are created to help classify the shopping opportunities into grocery, grocery/nongrocery, and nongrocery. A store variety index was created to indicate the range of stores in each zone. It was defined as the ratio of the number of store types in a particular zone to the total number of defined store types (i.e., 31):

Store diversity = number of types of the 31 stores in the zone / 31. (5)

The temporal factors include: generated travel time from the first mandatory activity zone to the shopping zone (Time1) and from the shopping zone to the second mandatory activity zone (Time2)[3], and generated activity duration for each zone (Activity duration) is obtained by subtracting the difference between a shopper's actual travel time and the sum of two parts of generated travel time from a shopper's actual activity duration:

---

[3] The generation of travel time variables is mainly for the computation of the activity duration variable. Although another set of models using travel time instead of activity duration is also estimated, it results in similar findings. Given that the activity duration variable accounts for an individual's time budget and makes more sense in activity-based research, models using activity duration are presented in this paper.

Table 2.1: Definition of the four-digit SIC code and corresponding explanatory variables

Shopping category	SIC code	Description of destination attractiveness variables	Frequency	Number of stores
Grocery	5411	Grocery stores	281	463
	5431	Fruit and vegetable markets	0	29
	5441	Candy, nut, and confectionery stores	1	23
	5461	Retail bakeries	1	121
Grocery/nongrocery	5301	Wal-Mart	38	7
	5401	Meijer	32	5
	5912	Drug stores and proprietary stores	49	150
	5311	Department stores	73	107
	5331	Variety stores	5	26
Nongrocery	5211	Lumber and other building materials dealers	13	157
	5251	Hardware stores	3	53
	5261	Retail nurseries, lawn and garden supply stores	1	92
	5511	Motor vehicle dealers (new and used)	6	321
	5531	Auto and home supply stores	8	238
	5621	Women's' clothing stores	2	119
	5661	Shoe stores	1	99
	5712	Furniture stores	1	186
	5731	Radio, television, and consumer electronics stores	6	79
	5735	Record and pre-recorded tape stores	1	91
	5932	Used merchandise stores	5	263
	5941	Sporting goods stores and bicycle shops	2	152
	5942	Book stores	5	50
	5943	Stationery stores	2	37
	5944	Jewellery stores	3	118
	5945	Hobby, toy, and game shops	5	80
	5949	Sewing, needlework, and piece goods stores	1	17
	5992	Florists	1	118
	5993	Tobacco stores and stands	1	33
	5995	Optical goods stores	2	53
	5999	Miscellaneous retail stores, not elsewhere classified	3	374
	6502	Shopping malls and centers	64	19
Total			616	3680

Note: The grocery/nongrocery shopping category refers to those shopping opportunities where shoppers can conduct both grocery and nongrocery shopping activities. Shoppers can conduct grocery shopping in some department stores (i.e., Wal-mart, K-Mart, SAM's club). The separation of Wal-Mart from department stores and Meijer from grocery stores is because of their relative influential role in affecting shoppers' destination choice.

$$Activity\ duration = Actual\ activity\ duration - ((Time\ 1 + Time\ 2) - actual\ travel\ time \quad (6)$$

One important component of the variables in the model is zonal level travel time. In TransCAD, the centroid of the each TAZ was generated and travel time from centroid to centroid was calculated. However, the software assumes that the inner zonal time is zero, which is unreasonable. To obtain the inner zonal travel time, a conventional method for transforming the irregular shapes of TAZs into circles of the corresponding area was adopted. The travel time is obtained by dividing the radius of the circle by the average speed in the zone. The average speed is calculated by:

$$Speed_j = \frac{\sum_i Length_i}{\sum_i Cost_i} \quad (7)$$

where j denotes the number of zones (818 in total) and i indicates the number of links within each zone. A program was written in ArcMap to get the summation of *Length* and *Cost*. The inner zonal time was then updated in the travel time matrix from TransCAD.

Table 2.2 lists the definition of independent variables aside from the destination attractiveness variables (i.e., Table 2.1).

Table 2.2: Definition of explanatory variables (excluding SIC code variables)

Variable	Definition
Store diversity	Store diversity index in the zone
Activity duration	Maximum shopping activity duration in the zone (in minute)
Middle age	1 if shopper is 35-64, 0 otherwise
Elderly	1 if shopper is 65+, 0 otherwise
Middle income	1 if shopper's household income is \$20,000-\$80,000, 0 otherwise
High income	1 if shopper's household income is over \$80,000, 0 otherwise
Refuse	1 if shopper refuses to provide income information, 0 otherwise
Employed	1 if shopper is employed (including part time), 0 unemployed
Female	1 if shopper is female, 0 male
Licensed driver	1 if shopper is a licensed driver, 0 otherwise

Three models (Models 1, 2 and 3) are estimated, each of which is based on a different combination of explanatory variables (Table 2.3). Only significant parameters are included in the models. The econometric software package LIMDEP (Greene 2002) is used in all the discrete choice model calibrations. Adjusted ρ^2 ($\overline{\rho}^2$) is calculated. The model estimations and predictions based on unconstrained and constrained choice sets are compared and reported in the next section. Since the space-time constraints are incorporated in the constrained choice set, making it remarkably different from the unconstrained choice set, it is expected that the parameter estimations and model goodness-of-fit based on constrained choice sets would be considerably different from those based on unconstrained choice sets.

Table 2.3: Model type and corresponding independent variables

Model	Independent variables
1. Destination attractiveness	Number of specific stores in zone
2. Activity duration	Number of specific stores in zone and activity duration
3. Interaction	Number of specific stores in zone, activity duration and interactions

Three types of PPAs are defined according to Table 2.1. If an individual's actual shopping destination is a grocery store, his/her PPA is defined as a grocery PPA. Nongrocery PPAs and grocery/nongrocerys PPA are defined likewise. Three types of shopping are modeled in this paper. Each type of shopping includes the corresponding PPAs. For instance, grocery, grocery/nongrocery PPAs are locations where grocery shopping can be conducted, therefore they are included in the grocery shopping model.

2.4 Results

Table 2.4 shows that the number of alternative in shoppers' constrained choice sets varies with the average number being 284. The large variance indicates that shoppers are faced by a substantially different choice set. If there was no restriction imposed on the choice set and the whole area was assumed accessible, then there would be 818 alternatives in every shopper's choice set, which would be far from the actual size of choice sets (i.e., an average of 284).

Table 2.4: Descriptive statistics concerning the number of alternatives in the constrained choice sets

Number of alternatives in shoppers' choice set	Frequency
0-50	107
51-100	79
101-200	109
201-300	82
301-400	59
401-500	42
501-600	46
601-700	44
701-800	35
801-818	13
Total	616

n	*Mean*	*Min*	*Max*
616	284	10	818

2.4.1 Explanatory Power of the Models

Table 2.5 shows the improvement in $\bar{\rho}^2$ in unconstrained and constrained models. There are three model specifications for each shopping type. As defined in Table 2.3, Model 1 includes only the number of specific stores in a zone, Model 2 considers the travel time and individual's time budget and contains the activity duration variable, and Model 3 controls interactions between the destination characteristics and individual's socio-demographic characteristics to test our belief that people with different socioeconomic characteristics will weigh destination attractiveness from various angles. The breakdown of the total number of stores into the number of specific type of stores helps fit the data. In the constrained situation, Model 1 can describe 9.2% of the data in the grocery model and 19.3% in the nongrocery model. With the temporal factors added, there is a marked improvement in Model 2 in terms of the considerable increase in $\bar{\rho}^2$. The improvement in Model 2 ranges from 36.6% to 54.2% in the unconstrained models while it ranges from 12.6% to 15.1% in the constrained model. The inclusion of interactions (Model 3) improves the model fit from Model 2 modestly.

Table 2.5: $\bar{\rho}^2$ improvement in unconstrained and constrained models

	Unconstrained			*Constrained*		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
General						
n=616						
$L(0)$	-1418.3924	-1418.3924	-1418.3924	-1418.3924	-1418.3924	-1418.3924
$L(\beta)$	-1166.4406	-584.0230	-578.5906	-1156.5900	-977.7558	-958.6616
ρ^2	0.1776	0.5883	0.5921	0.1846	0.3107	0.3241
$\bar{\rho}^2$	0.1713	0.5833	0.5857	0.1754	0.3015	0.3079
Grocery						
n=428						
$L(0)$	-985.5064	-985.5064	-985.5064	-985.5064	-985.5064	-985.5064
$L(\beta)$	-914.0445	-377.1674	-374.1836	-889.3140	-738.7136	-721.6376
ρ^2	0.0725	0.6173	0.6203	0.0976	0.2504	0.2678
$\bar{\rho}^2$	0.0695	0.6112	0.6122	0.0915	0.2423	0.2546
Nongrocery						
n=328						
$L(0)$	-755.2479	-755.2479	-755.2479	-755.2479	-755.2479	-755.2479
$L(\beta)$	-600.0907	-323.0748	-325.2422	-603.6291	-504.2484	-497.7970
ρ^2	0.2054	0.5722	0.5694	0.2008	0.3323	0.3409
$\bar{\rho}^2$	0.1975	0.5630	0.5574	0.1928	0.3217	0.3276
Shopping type	Improvement in $\bar{\rho}^2$			Improvement in $\bar{\rho}^2$		
general	0.0000	0.4120	0.0024	0.0000	0.1261	0.0064
grocery	0.0000	0.5417	0.0010	0.0000	0.1508	0.0123
nongrocery	0.0000	0.3655	-0.0056	0.0000	0.1289	0.0059

By and large, the table shows that the $\bar{\rho}^2$ of the unconstrained models are similar to the constrained model without temporal factors (Model 1), but much higher than the constrained model with temporal factors included (Models 2 and 3). An explanation of the difference in $\bar{\rho}^2$ between the unconstrained and constrained models as well as other measurements of goodness-of-fit can be found in Appendix A. (See section 2.7).

2.4.2 Predictive Ability

Table 2.6 shows that the constrained models are capable of providing a more accurate prediction of shopping destination zone than their unconstrained counterparts.

Table 2.6: Comparison of unconstrained and constrained shopping models via percentage right

	General shopping			*Grocery shopping*			*Nongrocery shopping*		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Unconstrained									
n	616	616	616	428	428	428	328	328	328
Hits	16	19	17	5	13	11	13	16	16
Percentage Right	2.5974	3.0844	2.7597	1.1682	3.0374	2.5701	3.9634	4.8780	4.8780
Expected percentage right	1.0461	1.7978	1.7851	0.2617	1.5507	1.5036	1.3987	2.7734	2.8618
Constrained									
n	616	616	616	428	428	428	328	328	328
Hits	56	74	84	37	52	57	40	46	51
Percentage Right	9.0909	12.0130	13.6364	8.6449	12.1495	13.3178	12.1951	14.0244	15.5488
Expected percentage right	3.8283	5.6805	6.7628	2.7541	4.6978	5.1152	3.9863	6.6681	6.7006
Improvement (Constrained minus unconstrained)									
Hits	40	55	67	32	39	46	27	30	35
Percentage Right	6.4935	8.9286	10.8767	7.4767	9.1121	10.7477	8.2317	9.1464	10.6708
Expected percentage right	2.7822	3.8827	4.9777	2.4924	3.1471	3.6116	2.5876	3.8947	3.8388

The number of correctly predicted zones (hits) and percentage right are higher in the constrained models. For example, in the three constrained general shopping models, the percentage right ranges from 9.09 to 13.64. In contrast, the percentage right ranges from 2.60 to 3.08 in the unconstrained general shopping models, implying lower predictive ability. The result is not unexpected as there are fewer alternatives in a constrained choice set, which gives the actual chosen alternative a higher probability of being chosen than in an unconstrained situation.

2.4.3 Parameter Estimations

Tables 2.7, 2.8 and 2.9 summarize unconstrained and constrained models for general, grocery and nongrocery shopping, respectively. When comparing the constrained models and unconstrained models, the former ones are more efficient in the sense that the standard errors of the estimated parameters in models for all three shopping types are smaller [4]. Notable differences between the coefficients in the unconstrained and constrained models exist. Since the constrained choice set is closer to the realistic choice set and the coefficients are associated with smaller standard errors, the following shopping behaviour analysis is focused mainly on the results from the constrained models. The coefficient of each type of store reveals the contribution of each specific category to the destination utility. Most types of stores increase the utility functions while a few categories of stores reduce the destination attractiveness. For instance, Table 2.8 shows that candy stores and bakery stores are not attractive in grocery shopping. Table 2.9 shows that in nongrocery shopping, used merchandise stores, auto stores and stationery stores play a negative role in the utility function. The large magnitudes of Wal-Mart and Meijer demonstrate their influential roles in a shopping destination zone. The coefficients for the activity duration are highly significant and positive in all types of shopping, indicating that the more time available for shopping, the greater the propensity that an

[4] Standard errors of the estimates are not reported in this paper. They can be computed using the reported coefficients and t-statistics.

individual is going to choose the destination. This reflects of the increased value that shoppers place on time (Kracklauer et al. 2001).

Models including interaction terms (Model 3) reveal interesting findings. The difference in socio-demographic characteristics suggests varied perspectives towards the destination attractiveness. In general shopping (Table 2.7) employment status, age, household income and possession of a driver's license affect the choice of a shopping destination. An employed shopper finds candy stores more attractive and they care more about the time spent in the stores than the unemployed. On the other hand, activity duration is less important for the high income shoppers and licensed drivers showing that whether a shopper is from a high income household or is a licensed driver can change greatly his or her view of time that could be spent in stores. The high income individuals are also less likely to shop in variety stores where the quality of the goods is relatively poor. Comparatively, Meijer attracts the middle income families by its mass-oriented goods with reasonable prices. Radio stores which sell electronic merchandise are surprisingly less attractive for middle income families than for low income families. A possible reason for this would be that the low-income households spend more time on lower cost/free entertainment such as radio and TV. For the elderly, they favor retail bakeries more than do young people. On the other hand, young people enjoy shopping in shopping malls and centers more than the middle age and the elderly do.

For grocery shopping (Table 2.8), some of the shopping behaviours are similar to the findings for general shopping such as the favorites on candy stores for the employed and on bakeries for elderly shoppers. In addition, household income exhibits a monetary

Table 2.7: General shopping models

Variables	*Unconstrained* Model 1		Model 2		Model 3		*Constrained* Model 1		Model 2		Model 3	
	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Number of Opportunities												
Grocery stores	0.2091	5.1397	0.4602	7.7117	0.4578	7.6450	0.2402	5.7794	0.1770	3.8396	0.1931	4.0842
Candy stores							-0.8087	-2.9907	-0.6092	-2.4401	-1.2310	-3.7962
Bakeries			0.3294	3.0030	0.3352	3.0564	-0.3528	-3.4039	-0.4967	-4.5268	-0.5810	-4.5475
Meijer	2.3911	9.2670	3.0817	6.7986	3.0333	6.7711	2.3777	9.1979	2.2345	8.0641	2.8005	6.8369
Drugstores	0.2066	2.6433					0.1843	2.3636				
Department stores	0.2530	3.8954					0.2435	3.5739	0.2278	2.9734	0.2702	3.3605
Used merchandise stores							0.0984	2.5082				
Variety stores			1.2081	4.7822	1.1622	4.5985			-1.0022	-4.1584	-0.9142	-3.6925
Nurseries			1.3568	8.6768	0.9934	4.1149	0.3337	2.8214	0.3186	2.3783	0.2935	2.1373
Auto stores	-0.1481	-2.7428										
Radio stores	0.3138	3.8774					0.2974	3.8474	0.2849	3.2466	0.5626	4.6438
Record tape stores									-0.2803	-2.7950	-0.2648	-2.5861
Stationery stores	-1.1093	-4.5835	-0.5784	-2.2374	-0.5370	-2.1046	-0.9177	-3.6536	-1.1807	-4.3211	-1.2895	-4.4386
Florists							0.1986	2.0780				
Miscellaneous stores	0.1826	4.6099					0.1772	3.9144				
Shopping malls	0.4339	2.8750					0.9168	5.3828	0.6592	3.7194	2.2227	6.1255
Store diversity									5.6645	8.4975	5.5410	8.0816
Temporal Factor												
Activity duration			0.1920	26.7893	0.1966	24.8947			0.1193	16.5776	0.2816	7.3869
Socio-demographic Characteristics												
Candy stores x Employed											1.0091	3.1456
Bakeries x Elderly											0.4020	1.9989
Meijer x Middle income											-1.1428	-2.0434
Variety x High income											-2.2242	-2.0997
Nurseries x Female					0.6041	2.0167						
Radio stores x Middle Income											-0.5054	-3.6907
Shopping malls x Middle age											-1.6498	-4.3864
Shopping malls x Elderly											-2.2666	-5.0240
Activity duration x Employed											0.0347	2.2925
Activity duration x Licensed driver											-0.1680	-4.2761
Activity duration x High income					-0.0380	-2.2617					-0.0484	-2.0685

Variables	*Unconstrained* Model 1	Model 2	Model 3	*Constrained* Model 1	Model 2	Model 3
SUMMARY STATISTICS						
n	616	616	616	616	616	616
$L(0)$	-1418.3924	-1418.3924	-1418.3924	-1418.3924	-1418.3924	-1418.3924
$L(\beta)$	-1166.4406	-584.0230	-578.5906	-1156.5900	-977.7558	-958.6616
ρ^2	0.1776	0.5883	0.5921	0.1846	0.3107	0.3241
$\bar{\rho}^2$	0.1713	0.5833	0.5857	0.1754	0.3015	0.3079

Table 2.8: Grocery shopping models

Variables	*Unconstrained* Model 1 Coeff.	t-ratio	Model 2 Coeff.	t-ratio	Model 3 Coeff.	t-ratio	*Constrained* Model 1 Coeff.	t-ratio	Model 2 Coeff.	t-ratio	Model 3 Coeff.	t-ratio
Number of Opportunities												
Grocery stores	0.2067	5.2478	0.3292	4.5327	0.3334	4.6044	0.2751	5.9403	0.2690	5.2935	0.2291	4.0981
Candy stores									-0.5406	-2.0296	-1.8210	-4.0148
Bakeries			-0.4416	-2.8856	-0.3790	-2.4519	-0.2220	-2.2010	-0.2935	-2.6211	-0.4829	-3.6351
Wal-Mart							0.9141	2.7844	1.3651	3.9548	1.5780	4.4314
Meijer	1.7232	7.6780	1.2097	2.6397	1.2691	2.6917	2.1349	9.1851	2.0255	7.8138	2.1519	8.1407
Drugstores			0.2897	2.2519	0.3271	2.5152	0.2117	2.3832				
Department stores	0.3504	7.0141	0.7401	7.1080	0.5828	4.9834	0.2788	4.8436	0.2532	3.1621	0.3700	4.1741
Store diversity									2.1560	3.6032	2.0412	3.3832
Temporal Factor												
Activity duration			0.2332	22.8684	0.1959	14.5249			0.1348	14.8426	0.1165	9.8227
Socio-demographic Characteristics												
Grocery stores x High income											0.3018	2.1997
Candy stores x Employed											1.8649	3.8446
Bakeries x Elderly											0.6549	2.8039
Department stores x High income											-0.4751	-2.4141
Department stores x Elderly					0.6583	2.8084						
Activity x Employed					0.0773	3.8513						
Activity x Middle income											0.0375	2.0743
SUMMARY STATISTICS												
n	428		428		428		428		428		428	
$L(0)$	-985.5064		-985.5064		-985.5064		-985.5064		-985.5064		-985.5064	
$L(\beta)$	-914.0445		-377.1674		-374.1836		-889.3140		-738.7136		-721.6376	
ρ^2	0.0725		0.6173		0.6203		0.0976		0.2504		0.2678	
$\bar{\rho}^2$	0.0695		0.6112		0.6122		0.0915		0.2423		0.2546	

Table 2.9: Nongrocery shopping models

Variables	*Unconstrained* **Model 1**		**Model 2**		**Model 3**		*Constrained* **Model 1**		**Model 2**		**Model 3**	
	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Number of Opportunities												
Meijer			4.3549	7.3537	4.5642	7.2961			2.7383	8.2057	2.7079	8.1005
Drugstores	0.3293	3.2305					0.3755	3.7132	0.3117	2.8892	0.3249	3.0016
Department stores							0.2524	3.2356				
Used merchandise stores									-0.1420	-2.1824	-0.1377	-2.1156
Hardware stores	-0.5675	-2.2100										
Motor vehicle dealers							0.1128	2.6042				
Auto stores									-0.1576	-2.1632	-0.1912	-2.4877
Radio stores	0.6122	5.8886					0.2515	2.2814				
Sporting goods stores			0.3710	2.8969	2.4208	2.5544						
Book stores	0.3819	2.6441	1.3896	9.5260	1.4457	9.8757						
Stationery stores			-0.8499	-2.2184	-0.8492	-2.1980			-1.0536	-3.0508	-1.0607	-3.0481
Florists	0.3084	2.5223	0.4240	2.4412	0.4530	2.5706	0.4331	3.4354				
Tobacco stores			0.6092	2.2593	0.6419	2.3643						
Shopping malls	1.1899	7.8128					1.0114	5.2235	0.4302	2.4494	0.4676	2.6573
Store diversity									6.1241	8.4798	5.9670	8.2239
Temporal Factor												
Activity duration			0.1667	17.6647	0.1786	17.2843			0.0814	9.8646	0.0643	6.2826
Socio-demographic Characteristics												
Auto stores x High income											0.3204	1.9710
Sporting goods stores x Licensed driver					-2.0962	-2.2022						
Activity x Middle income											0.0456	2.6472
Activity x High income					-0.0487	-2.0671						
SUMMARY STATISTICS												
n	328		328		328		328		328		328	
$L(0)$	-755.2479		-755.2479		-755.2479		-755.2479		-755.2479		-755.2479	
$L(\beta)$	-600.0907		-323.0748		-325.2422		-603.6291		-504.2484		-497.7970	
ρ^2	0.2054		0.5722		0.5694		0.2008		0.3323		0.3409	
$\overline{\rho}^2$	0.1975		0.5630		0.5574		0.1928		0.3217		0.3276	

impact on the individuals' choice of different stores. High-income people favor grocery stores, but they do not like grocery shopping at department stores. Nowadays, at least in terms of grocery shopping, department stores such as K-Mart and Wal-Mart have attracted low income families by lowering the commodity price. Middle income households seem to be more pressed for time than low-income households do in that the former group sees more importance in the activity time spent at a destination. For nongrocery shopping (Table 2.9), the different opinions on stores are again attributed to income inequality. Although auto stores are unattractive for low-income families, they are actually favored by wealthy people. Similar to what is found in the grocery shopping model, middle income shoppers are more likely to choose a destination zone if they could spend more time shopping in stores.

2.5 Conclusion

Discussed in previous destination choice studies, an unconstrained choice set formation will lead to bias in parameter estimations and inaccurate forecasts. This paper investigates shoppers' destination choice behaviour by delineating a space-time constrained choice set. The choice set is unique individually in the sense that one individual's activity schedule is different from another's and shopping opportunities are distributed unevenly across the study area. Our empirical results show, destination choice modeling based on the space-time constrained choice set leads to considerable differences in parameter estimates and probability predictions from that based on a universal choice set.

We improve upon existing shopping destination choice studies by (a) defining a behaviourally sound destination choice set using a GIS-based algorithm (Scott 2006) and (b) separating grocery shopping from nongrocery shopping. The incorporation of space-time constraints in defining destination choice sets is activity based. It gives substantial consideration to an individual's activity scheduling and distribution of specific store types. It is the adoption of the spatio-temporal framework that differentiates our work from previous studies in shopping destination choice behaviour and characterizes the contribution of this research.

The primary contribution lies in that the GIS-based constrained choice set takes into account an individual's activity scheduling, leading to an individual-specific space-time constrained destination choice set. Model estimations and predictions based on the constrained choice set show significant differences from the unconstrained choice set. A comparison between unconstrained and constrained models indicates that:

1. Determinants for unconstrained models could be quite different from the constrained models. It suggests that some categories of stores or activities undertaken at these locations could be less sensitive to the constraint of time. They could fit the data very well without the presence of an individual's time budget. When time budget is incorporated to derive time-constrained destination choice sets, some of these variables lost their descriptive capability while some are replaced by other variables that are more capable of describing data in the time constrained situation.

2. By comparing the magnitude of parameters in the unconstrained and constrained models, those in the unconstrained models are found to be larger than or the same as those in the constrained models. It is implied that if a shopper's choice set is overestimated, so will the effects of the explanatory variables. In addition, most of the standard errors of the unconstrained estimates are larger than those of the constrained estimates. This suggests that the estimators in constrained models are more efficient and tend to be more consistent.
3. Percentage right is an indicator of the predictive ability of models. Comparatively, constrained models are able to predict destinations better than unconstrained models. This proves the advantage of the constrained model from a predictive point of view. This empirical evaluation of constrained models deepens our faith in the importance of a well-specified constrained choice set.

The secondary contribution lies in the exploration of grocery and nongrocery shopping behaviour by breaking down shopping opportunities by SIC code. Classification of stores provides a magnified view into what types of shopping trips shoppers have conducted. By including the interaction terms in the models, we can exam the perspectives that the different groups have on the shopping destination characteristics. There is more confidence in the behaviours revealed in the constrained models because of the better specified choice set that the models are based on. Unconstrained models reveal certain behaviours that are not revealed by the constrained models. Some of them are consistent with our expectation, such as less consideration to activity duration for a licensed driver and for the unemployed. But some of the behaviours are suspicious or even counterintuitive such as the higher popularity of sporting goods stores for passengers and the higher popularity of department stores for elderly. These findings could be a result of the mis-specified unconstrained choice set. On the other hand, results exclusively from the unconstrained models could also be true. These results do not appear in the constrained model because the destination characteristic parameter forming the interactions not significant in the constrained model. Further investigation is needed to determine which interactions revealed exclusively by the unconstrained models are reflections of the true shopping behaviour.

Travel-time constrained destination choice modeling is of particular importance to market analysts, location analysts and transportation planners who predict the share in alternatives by discrete choice models (Thill and Horowitz 1997). In the future, more constraints can be imposed on the universal choice set to define a more restrictive and realistic choice set. This would help specify an even more sound and correct model in the context of shopping destination choice.

2.6 References

Arentze T.A. and Timmermans H. 2005, An analysis of context and constraints-dependent shopping behaviour using qualitative decision principles, *Urban Studies*, vol. 42, pp. 435-448.

Arentze T., Borgers A., and Timmermans H. 1993, A model of multi-purpose shopping trip behaviour, *Regional Science*, vol. 72, pp. 239-256.

Barnard P.O. 1987, Modeling shopping destination choice behaviour using the basic multinomial logit model and some of its extensions, *Transportation Reviews*, vol. 7, pp. 17-51.

Bell D.R., Ho T.-H., and Tang C.S. 1998, Determining where to shop: fixed and variable costs of shopping, *Journal of Marketing Research*, vol. 35, pp. 352-369.

Ben-Akiva M. and Lerman S.R. 1985, *Discrete choice analysis: theory and application to travel demand,* The MIT Press, Cambridge.

Buliung R.N. 2004, *Data organization and the design of measures for exploration and analysis of household activity/travel behaviour,* Ph.D. Dissertation, McMaster University, Hamilton, ON.

Burns L.D. 1979, *Transportation, temporal, and spatial components of accessibility,* D.C. Heath and Company, Lexington.

Ettema D. and Timmermans H. 2004, Social exclusion and space-time accessibility under conditions of uncertain travel times: theory and numerical simulations, in *Annual meetng of Association of American Geographers*, Philadelphia, March 14-19.

Ewing R., Haliyur P., and Page G.W. 1994, Getting around a traditional city, a suburban planned unit development, and everything in between, *Transportation Research Record*, vol. 1466, pp. 53-62.

Fotheringham A.S. and Trew S. 1993, Chain image and store choice modeling: the effects of income and race, *Environment and Planning A*, vol. 25, pp. 179-196.

Fox E.J., Montgomery A.L., and Lodish L.M. 2004, Consumer shopping and spending across retail formats, *Journal of Business*, vol. 77, pp. 25-60.

Greene W.H. 2002, *LIMDEP Version 8.0* Econometric Software, Inc., Plainview, NY

Guy C. 1985, The food and grocery shopping behaviour of disadvantaged consumers: some results from the Cardiff consumer panel, *Transactions of the Institute of British Geographers*, vol. 10, pp. 181-190.

Hägerstrand T. 1970, What about people in regional science?, *Papers of the Regional Science Association*, vol. 24, pp. 7-21.

Kitamura R. and Kermanshah M. 1984, Sequential model of interdependent activity and destination choices, *Transportation Research Record*, vol. 987, pp. 81-89.

Koppleman F.S. and Hauser J.R. 1978, Destination choice behaviour for non-grocery-shopping trips, *Transportation Research Record*, vol. 673, pp. 157-165.

Kracklauer A., Passenhaim O., and Seifert D. 2001, Mutual customer approach: how industry and trade are executing collaborative customer relationship management, *International Journal of Retail and Distribution Management*, vol. 29, pp. 515-519.

Kwan M.-P. and Hong X.-D. 1998, Network-based constraints-oriented choice set formation using GIS, *Geographical Systems*, vol. 5, pp. 139-162.

Landau U., Prashker J.N., and Alpern B. 1982, Evaluation of activity constrained choice sets to shopping destination choice modeling, *Transportation Research Part A*, vol. 16, pp. 199-207.

Lenntorp B. 1978, A time-geographic simulation model of individual activity programmes, in *Human Activity and Time Geography*, Carlstein T., Parkes D., and Thrift N., eds., Edward Arnold, London, pp. 163-180.

Limanond T., Niemeier D.A., and Mokhtarian P.L. 2005, Specification of a tour-based neighborhood shopping model, *Transportation*, vol. 32, pp. 105-134.

Lloyd R. and Jennings D. 1978, Shopping behaviour and income: comparisons in an urban environment, *Economic Geography*, vol. 54, pp. 157-167.

Manski C.F. 1977, Structure of random utility models, *Theory and Decision*, vol. 8, pp. 229-254.

McFadden D. 1978, Modeling the choice of residential location, in *Spatial Interaction Theory and Planning Models*, Karquist A., Lundqvist L., Snickars F., and Weibull J., eds., Elsevier, Amsterdam, pp. 75-96.

McFadden D. 1974, Conditional logit analysis of qualitative choice behaviour, in *Frontiers in Econometrics*, Zarembka P., ed., Academic Press, New York, pp. 105-142.

Miller E.J. and O'Kelly M.E. 1983, Estimating shopping destination choice models from travel diary data, *The Professional Geographer*, vol. 35, pp. 440-449.

Miller H.J. 1991, Modeling accessibility using space-time prism concepts within geographical information systems, *Geographical Information Systems*, vol. 5, pp. 287-301.

Miller H.J. 1999, Measuring space-time accessibility benefits within transportation networks: basic theory and computational methods, *Geographical Analysis*, vol. 31, pp. 187-212.

Miller H.J. 2004, Activities in space and time, in *Transport Geography and Spatial Systems*, Stopher P., Button K., Haynes K. and Hensher D., eds., Pergamon/Elsevier Science, London.

Miller H.J. 2005, A measurement theory for time geography, *Geographical Analysis*, vol. 37, pp. 17-45.

Nelson D. and Niles J. 2000, Observations on the causes of nonwork travel growth, in *79th Annual Meeting of the Transportation Research Board: Compendium of Papers CD-ROM*, Washington D.C., January 9-13.

Pellegrini P.A., Fotheringham A.S., and Lin G. 1997, An empirical evaluation of parameter sensitivity to choice set definition in shopping destination choice models, *Regional Science*, vol. 76, pp. 257-284.

Pred A. 1977, The choreography of existence: comments on hagerstrand's time-geography and its usefulness, *Economic Geography*, vol. 53, pp. 207-221.

Recker W. and Kostyniuk L.P. 1978, Factors influencing destination choice for the urban grocery shopping trip, *Transportation*, vol. 7, pp. 19-33.

Richards M.G. and Ben-Akiva M. 1974, A simultaneous destination and mode choice model for shopping trips, *Transportation*, vol. 3, pp. 343-356.

Scott D.M. 2006, Constrained destination choice set generation: A comparison of GIS-based approaches, in *85th Annual Meeting of the Transportation Research Board: Compendium of Papers CD-ROM*, Washington D.C., January 9-13.

Swait J.D. and Ben-Akiva M. 1987, Incorporating random constraints in discrete models of choice set generation, *Transportation Research Part B*, vol. 21, pp. 91-102.

Tang C.S, Bell D.R., and Ho T.-H. 2001, Store choice and shopping behaviour: how price format works, *California Management Review*, vol. 43, pp. 56-74.

Tardiff T.J. 1980, Definition of alternatives and representation of dynamic behaviour in spatial choice models, *Transportation Research Record* no. 723, pp. 25-30.

Thill J.C. 1992, Choice set formation for destination choice modeling, *Progress in Human Geography*, vol. 16, pp. 361-382.

Thill J.C. and Horowitz J.L. 1991, Estimating a destination-choice model from a choice-based sample with limited information, *Geographical Analysis*, vol. 23, pp. 298-315.

Thill J.C. and Horowitz J.L. 1997, Travel-time constraints on destination choice sets, *Geographical Analysis*, vol. 29, pp. 108-123.

Timmermans H. 1981, Multiattribute shopping models and ridge regression analysis, *Environment and Planning A*, vol. 13, pp. 43-56.

Timmermans H. 1996, A stated choice model of sequential mode and destination choice behaviour for shopping trips, *Environment and Planning A*, vol. 28, pp. 173-184.

Timmermans H., Borgers A., and Waerden P. van der 1991, Mother logit analysis of substitution effects in consumer shopping destination choice, *Journal of Business Research*, vol. 23, pp. 311-323.

Timmermans H., van der Haijden R., and Westerveld H. 1984, Decision making between multiattribute choice alternatives: a model of spatial shopping behaviour using conjoint measurements, *Environment and Planning A*, vol. 16, pp. 377-387.

Williams H.C.W.L. and Ortúzar J. de D. 1982, Behavioural theories of dispersion and the misspecification of travel demand models, *Transportation Research Part B*, vol. 16, pp. 167-219.

2.7 Appendix A

2.7.1 Measure of goodness-of-fit and predictive ability

The likelihood function is defined as:

$$L(\beta_1, \beta_2, \ldots, \beta_k) = \prod_{n=1}^{N} \prod_{j \in C_n} P_n(j)^{y_{jn}} \tag{1}$$

Where $y_{jn} = \begin{cases} 1 & \text{if individual } n \text{ chose alternative } j \\ 0 & \text{otherwise} \end{cases}$

The parameters can be obtained by maximizing the log-likelihood function:

$$\ln L = \sum_{n=1}^{N} \sum_{j \in C_n} y_{jn} \ln P_n(j) \tag{2}$$

Specifically, $\ln L(0)$ is the value of the log likelihood function when all the parameters are zero. It is the most naïve model, which is a model where each alternative has an equal share of being chosen. $\ln L(\hat{\beta})$ is the value of the maximized log likelihood function.

2.7.2 ρ^2

ρ^2 and $\bar{\rho}^2$ are two measures of goodness of fit. ρ^2 is the explained portion of the model and it is defined as:

$$\rho^2 = 1 - \frac{\ln L(\hat{\beta})}{\ln L(0)} \tag{3}$$

While $\bar{\rho}^2$ is adjusted by the number of parameters, K in the model. It is denoted as:

$$\bar{\rho}^2 = 1 - \frac{\ln L(\hat{\beta}) - K}{\ln L(0)} \tag{4}$$

The odds ratio does not depend on other alternatives:

$$\ln\left[\frac{P_n(j)}{P_n(l)}\right] = x_n'(\beta_j - \beta_l) \tag{5}$$

In the unconstrained case, the individual choice set is assumed the same as the universal choice. Therefore, $P_n(j|C_n) = P_n(j|C)$

In the constrained case, each individual's choice set C_n differs from each other's. Expectedly, $P_n(j|C_n) \geq P_n(j|C)$, given $C_n \in C$. It can be shown that the ρ^2 of a constrained model is higher than an unconstrained model as follows:

$$\ln L = \sum_{n=1}^{N}\sum_{j\in C_n} y_{jn} \ln P_n(j) = \sum_{n=1}^{N}\sum_{j\in C_n} y_{jn}(V_{jn} - \ln\sum_{l\in C_n} e^{V_{ln}}) \tag{6}$$

In the above equation, the difference of log likelihood between a constrained model and an unconstrained model lies in the term $\sum_{l\in C_n} e^{V_{ln}}$.

For convenience to compare the constrained and unconstrained situations, we define $C_n \in C$, $M \in \overline{C \cap C_n}$ $(M = C - C_n)$. This term in the unconstrained case would become

$$\sum_{l\in C_n} e^{V_{ln}} + \sum_{m\in M} e^{V_{mn}}$$

$$\ln L(\hat{\beta}_*) = \sum_{n=1}^{N}\sum_{j\in C_n} y_{jn} \ln P_n(j) = \sum_{n=1}^{N}\sum_{j\in C_n} y_{jn}[V_{jn} - \ln(\sum_{l\in C_n} e^{V_{ln}} + \sum_{m\in M} e^{V_{mn}})] \tag{7}$$

where $\ln L(\hat{\beta}_*)$ is the log likelihood function in an unconstrained situation.

Since $\sum_{m\in M} e^{V_{mn}} > 0$, it follows that:

$$\ln L(\hat{\beta}) > \ln L(\hat{\beta}_*)$$

$$1 - \frac{\ln L(\hat{\beta})}{\ln(0)} > 1 - \frac{\ln L(\hat{\beta}_*)}{\ln(0)}$$

Which shows that ρ^2 of a constrained model is higher than one of an unconstrained model.

When there is a large choice set and MNL models are based on random sampling, which is the case in our study, ρ^2 of a constrained model would be lower than that of an unconstrained model. The term $\sum_{l \in C_n} e^{V_{ln}} + \sum_{m \in M} e^{V_{mn}}$ in a non-sampling model would become $\sum_{q \in C} e^{V_{qn}}$ in the sampling model. In the case of time-constrained destination choice models, the travel time would reduce the utility of the destination $\beta_t < 0$. The travel time of the constrained alternatives is less than that of the unconstrained alternatives, $t_l < t_q$, therefore, $t_l' \beta_t > t_q' \beta_t^*$.

Provided other factors in the utility function do not vary much between the constrained and unconstrained situation,

$$X_{ln}\beta > X_{qn}\beta^*$$
$$\sum_{l \in C_n} e^{V_{ln}} > \sum_{q \in C} e^{V_{qn}}$$

In this case, ρ^2 of a constrained model is lower than that of an unconstrained model. In other words, the temporal effect of shopping destination has been absorbed by the space-time prism. When using the space-time prism to constrain the destination choice set, it is a sort of "advance consumption" of the temporal effect in alternatives. Therefore, the temporal factors appear to be of less significance in constrained models.

However, it is not necessarily the case in the models without the temporal effect such as the models with SIC only. When travel time or activity duration is not controlled, the condition $\sum_{l \in C_n} e^{V_{ln}} > \sum_{q \in C} e^{V_{qn}}$ does not necessarily hold. Hence ρ^2 of a constrained model could be higher or lower than that of an unconstrained model.

2.7.3 Percentage right

Percentage right ("% right") is used to examine the predictive ability improvement of a constrained choice set is in the full sample. Comparison between the percentage right of a constrained and unconstrained model could prove the predictive improvement of constrained models. Percentage right is defined as:

$$\frac{100}{N} \sum_{n=1}^{N} \hat{y}_n$$

where $\hat{y}_i$ is one if the highest predicted probability corresponds to the actual chosen alternative and zero otherwise. This measurement, however, is less sensitive to the low probabilities of chosen alternatives and it is not a good way to preserve the desirable property of replicating the probabilities of the alternatives:

$$\frac{1}{N} \sum_{n=1}^{N} P_n(j) = \frac{1}{N} \sum_{n=1}^{N} y_{jn} \tag{8}$$

which can be derived by setting the first derivatives of the log likelihood function with respect to all coefficients equal to zero (Ben-Akiva and Lerman 1985). A better measure of % right is

$$\frac{100}{N} \sum_{n=0}^{N} \sum_{j=0}^{J} P_n(j)^{y_{jn}} \tag{9}$$

Chapter 3 Scale Effects on Constrained Destination Choice Models for Shopping

3.1 Introduction

Predefined zoning systems such as traffic analysis zones (TAZs), census tracts and enumeration areas are familiar to professionals in the transportation field. Nevertheless, they should not be taken for granted as they might not be suitable as units for analysis for three reasons: first, sizes of the zones vary considerably from the urban core to the city outskirts, which results in a large standard deviation in the distribution of the data. Second, their boundaries are not updated to coincide with rapid economic development and dramatic changes in the physical landscape and social environment (i.e., Ding [1998] called for an adjustment in TAZs).The resultant traffic flow estimations and transportation system evaluation would not be precise (Ding 1998). Three, most people in day-to-day life are not directly dealing with these "basic units." People perceive and understand the environment based on their knowledge, behaviour and many other factors immeasurable and/or unknown to analysts. The unit of their "cognitive map" (Kitchin 1994) is not likely to overlap with the pre-defined basic unit (e.g., perceptions of a neighborhood, Coulton et al. 2001; Guo and Bhat 2004).

While conventional models in transport studies rely on data aggregated using TAZs, the predefined TAZs are criticized as they are not adjusted over time (Ding 1998). To improve the TAZ zoning, researchers have designed and implemented algorithms for an optimal TAZ zoning system according to a number of criteria (O'Neal 1991; Ding 1998). The adoption of TAZ is associated with the modifiable areal unit problem

(MAUP). MAUP occurs when data are collected and analyzed at a zonal level (i.e., TAZs). It consists of a scale effect and a zoning effect (Páez and Scott 2004; Wong 2004). Analytical results obtained by techniques sensitive to scale effects or zoning effects are likely to change as the aggregation level or area boundary varies (Openshaw 1978; Openshaw and Taylor 1981; Fotheringham and Wong 1991; Zhang and Kukadia 2005).

Previous studies have examined the aggregation effects and zoning effects of data in a transportation context. Researchers have questioned the suitability of the pre-defined TAZ and the reliability of analysis based on TAZ (Putman and Chung 1989; O'Neill 1991; Ding 1998; Horner and Murray 2002; Zhang and Kukadia 2005). Grids are among those that have been suggested as alternative zoning systems (Zhang and Kukadia 2005). Fotheringham and Rogerson (1993) pointed out that documenting the results on model estimations at different scales and zoning levels is important to assess MAUP effects and critically evaluate the reliability of statistical estimates. Hitherto, the majority of MAUP research that has been carried out in transportation has focused on the scale and zoning effects of the gravity model and spatial interaction models (Putman and Chung 1989), leaving the logit model an area worthy of further exploration (Fotheringham and Wong 1991; Zhang and Kukadia 2005).

On the other hand, there are ample studies on shopping destination choice. An accurate prediction of shopping destination is important for travel demand forecasting. Many previous studies (especially in marketing, retailing and consumer services) focused on point shopping destinations (e.g., Bell et al. 1998; Fox et al. 2004; Lloyd and Jennings 1978; Guy 1985; Timmermans 1981, 1996) rather than on zonal shopping destinations (e.g., Landau et al. 1982; Limanond et al. 2005). In transportation studies, however, most data are collected based on traffic analysis zones (TAZs). A shopping destination choice model based on a zonal scheme would be more meaningful in terms of its application in travel demand forecasting. Moreover, the location of a store is often subject to change because of external (i.e., fire accident) and internal forces (i.e., store revenue). When thousands of stores are considered as possible destinations, prediction would be more accurate if the stores are aggregated to zones (i.e., TAZ, grids). As a matter of fact, zones are often used when modeling activity destinations in activity-based research (Buliung 2004). In this sense, we adopt a zonal based model in our research.

Most of the current shopping behaviour studies are confined to general and grocery shopping behaviour (Miller and O'Kelly 1983; Park et al. 1989; Kim and Park 1997) and there has been very little empirical research on nongrocery shopping (Fox et al. 2004). Unavailability of standardized industrial classification (SIC) information for shopping opportunities imposes difficulty in distinguishing grocery from nongrocery shopping. Another limitation of conventional shopping models is the widely adopted assumption of a universal destination choice set for every individual (Timmermans et al. 1984; Fotheringham and Trew 1993). Since the impact of constraints from an individual's activity schedule on the shopping location choice is considerable, spatial shopping models used for analysis or prediction should be based on activity-constrained choice sets at an individual level (Arentze and Timmermans 2005).

In this paper, our objective is to investigate the scale effect on constraints-oriented shopping destination choice. The innovations introduced in this paper are twofold: a) the retrieval of a space-time constrained destination choice set and b) the breaking down of shopping opportunity categories using SIC code to model grocery and nongrocery shopping separately. We apply the concept of the space-time prism and derive a potential path area (PPA), which is the planer area that could be reached given a time constraint and two consecutive fixed activities, for each individual using the TAZ conventional zoning system and 10 randomly generated grid systems (i.e., from cell size of 1 km^2 to 10 km^2 in 1 km^2 increments). This space-time constrained destination choice set is no doubt more realistic than the assumption that all shopping opportunities are accessible for each individual. Based on the constrained choice set, we model three types of shopping behaviour: grocery, nongrocery and grocery & nongrocery (general) shopping. The use of SIC codes provides an enlarged view as to which specific types of shopping opportunities are sensitive to the MAUP given an individual's spatio-temporal constraints. Parameter estimations and goodness-of-fit of 10 grid models are compared against the TAZ model. We also use GIS and propose a new travel time-based point-by-point matching measurement of model prediction. It has been noted that GIS is capable of depicting the spatial aspect of goodness-of-fit of a model. It helps provide an accurate sensitivity to predictions when spatial data are involved (Fotheringham and Rogerson 1993).

The remainder of this study is structured as follows. We start by reviewing studies of the modifiable areal unit problem in Section 2, followed by a description of the study area, PPA algorithm and choice set generation, as well as the model specification in Section 3. The empirical results of parameter estimations and model goodness-of-fit in the 11 zoning systems are reported and compared in Section 4. We conclude in Section 5 by summarizing the findings of this research, with a recommendation for the 1 km^2 grid system with regard to the spatial aspect of goodness-of-fit of a model.

3.2 Literature Review

3.2.1 Modifiable Areal Unit Problem

Shopping destinations can be defined as either points or zones. Ideally, researchers expect that travel forecasts would be consistent regardless of any zoning system (Barnard 1987). However, at the zonal level, the modifiable areal unit problem (MAUP) is a potential source of error as a result of varied unit definition and area boundaries at different aggregation schemes. Different scale levels and zoning configurations are known to affect statistical tests and parameter estimations. In fact, early works have shown that a substantial change in correlation coefficients occurs at different data aggregation levels (Gehlke and Biehl 1934; Robinson 1935; Openshaw and Taylor 1979). Empirical studies in disciplines such as geography, sociology, ecology and criminology have shown that changes in scale or unit definition could influence statistical indices (i.e., mean, standard deviation, correlations) and parameter estimations (Openshaw 1984; Fotheringham and Wong 1991; Amrhein 1995; Plante et al. 2004; Ratcliffe 2005).

The MAUP is grounded in social science where a parallel term "ecological fallacy" refers to the potential erroneous inference of individual behaviour based on aggregated spatial data (Gehlke and Biehl 1934; Robinson 1935). Gehlke and Biehl (1934) noticed the positive relation between correlation coefficients and level of aggregation of census tracts. MAUP consists of a scale (or aggregation) effect and a zoning effect. Scale effect refers to inconsistency resulting from different resolutions (Páez and Scott 2004; Wong 2004). When aggregating spatial units, the scale of data analysis changes and spatial resolution decreases due to the replacement of finer with coarser units. Zoning effect refers to inconsistency resulting from different zone partitioning (Páez and Scott 2004; Wong 2004). There could be infinite possible zoning schemes for a finite number of zones. The choice of scale and zoning depends on the study goals, and there is no universal solution (Fotheringham et al. 2000).

Analytical results obtained by techniques sensitive to scale effects or zoning effects are likely to change as the aggregation level or area boundary varies (Cliff and Haggett 1970; Openshaw 1977a, 1977b, 1978; Openshaw and Taylor 1981; Fotheringham and Wong 1991; Fotheringham and Rogerson 1993). Openshaw (1977a, 1977b, 1978) has conducted extensive research on the problem of spatial data aggregation and optimal zoning design. His studies show that the parameter estimates and goodness-of-fit of spatial interaction models can depend heavily on the properties of the specific zoning system. It was suspected that zones of irregular shapes may be at a disadvantage compared with regular polygons yet there was no empirical evidence to support it (Openshaw 1977a, 1978). Instead of attempting to model the effects of scale and aggregation, Openshaw (1977b) designed a hierarchical heuristic approach and implemented it in an automatic zoning procedure in order to select a set of zones that optimized an objective function that was used to measure the performance of a model. He suggested more attention should be paid to the way in which zoning systems are designed in the context of spatial data aggregation.

Other research concerning the development of an optimal zoning system include Cliff and Haggett (1970) in the area of regional zoning design, and O'Neill (1991) and Ding (1998) who studied zoning design with respect to the TAZ. Cliff and Hagget (1970) explored a combinatorial procedure to aggregate sub-areas into regions. As there are numerous ways of alternative groupings, they proposed a measure of efficiency (i.e., a joint measure of an aggregation index and an information index) in the regional grouping process and attempted to find the optimal value of it.

The TAZ is the basic unit in most transport studies. The size of the TAZ varies among metropolitan areas yet it is similar to census tracts or smaller geographic units such as enumeration districts or block groups (Horner and Murray 2002). Rapid changes of physical landscapes, economics and travel patterns in an urban area call for TAZ adjustment so as to describe and evaluate the transportation system precisely (Ding 1998). To minimize systematic errors in estimation, O'Neill (1991) and Ding (1998) attempted to design an optimal TAZ zoning system using GIS. Ding (1998) developed a TAZ design algorithm that considered eight common criteria. These criteria are individually desirable but cannot be absolutely compatible (i.e., uniqueness and homogeneity). It was found that certain estimates such as total vehicle miles increased as the number of TAZs increased. When there were more than 50 TAZs, all major transportation service indices became stable. Although it is seemingly self-explanatory that a smaller spatial unit associated with greater flexibility in data aggregation is more likely to result in smaller errors, Ding could not find evidence to support the argument.

Acknowledging the existence of MAUP, Openshaw and Taylor (1981) proposed a need for a scale/zoning-independent framework to analyze spatial data. In the last two decades, a number of application-specific solutions to MAUP have been derived (Tobler 1989; Holt et al. 1996). In an attempt to investigate the aggregation effect on the regression coefficient, Holt et al. (1996) developed a model which shows that the regression coefficient is dependant on sample size in each area in the sample, intra-area correlation, intra-area cross correlation and the correlation coefficient. They believed that in order to evaluate the aggregation effect, say from the individual level to a higher level (i.e., census districts, postal code areas), concentration should be on identifying the critical factors influencing the intra-area cross correlation coefficient and its variation patterns with varied sets of area. Beardwood and Kirby (1975) showed that the predictions of trip distribution using a gravity model from a coarser zoning system can be made consistent with the ones from the original zoning system and *vice versa*, by way of illustrating the gravity model's excludability property (data can be excluded without affecting the predictions) and compressibility property (data can be aggregated to larger zones without affecting the predictions) under certain conditions. However, a true MAUP effects independent framework is only achieved when using individual-level data (Horner and Murray 2002). Yet the reality is that the focus of geographical analysis is often concerned with entities above the individual level (Holt et al. 1996) and the majority of social data are only available at certain aggregation levels (Amrhein 1995).

A number of the MAUP studies suggest that parameters are sensitive to changes in aggregation. Estimations and model performance in linear regression models (Fotheringham and Wong 1991) and spatial interaction models of the gravity type (Openshaw 1977a, Putman and Chung 1989) have been explored. By calibrating a linear regression model predicting the mean family income by four independent variables, Fotheringham and Wong (1991) found noticeable differences in the sensitivity of the parameter estimates. Two of the parameter estimates (i.e., the percentage of population who are homeowners and the percentage of population who are black) in their model exhibited far less sensitivity to changes in scale than the other two estimates (i.e., the percentage of the population that is blue-collar and the percentage of population that is age over 65 years). The reason regarding the insensitivity of the former two parameter estimates remains unknown.

To systematically study the aggregation effect on a multivariate spatial interaction model, Putman and Chung (1989) adopted one random aggregation (RA) and four systematic aggregations (equal number of basic spatial units per zone, equal total area, equal total population and equal number of low-income households per zone) procedures to combine the basic spatial units into zones, and compared the sensitivity of parameters and goodness-of-fit based on these five systems. They concluded that the systematic aggregation method yielded better goodness-of-fit and more reliable parameters than the random method. The equal area method yielded the best goodness-of-fit among the five aggregation methods. Most parameters in one aggregation method are significantly different from the other methods. In addition, descriptive statistics of attractiveness variables were used to explore the possible causes for the remarkable response in parameters to different zonal aggregations. They found that the reduction in parameter variance was associated with a reduction of model fit and that smaller standard deviation of the attractiveness variables caused smaller standard deviation of the parameter estimates. No clear relationship was shown between the parameter reliability and the goodness-of-fit of the model.

The popular multinomial logit model has attracted attention in recent transportation studies (Guo and Bhat 2004; Zhang and Kukadia 2005). In the context of residential location choice, Guo and Bhat (2004) questioned the assumption that the TAZ and other administratively defined areal units are coterminous with an individual's perceived neighbourhood. They mimicked individual's perception of a neighbourhood by treating it as hierarchical residential groupings (i.e., census block, block group, census tract and county). They argued that a multi-scale logit model (MSL), where spatial neighbourhood attributes observed at multiple levels were incorporated in the utility function, was more advantageous than the conventional single level models and therefore MSL could be a solution to the scale effect in residential location choice studies. Zhang and Kukadia (2005) examined the scale effect and the zonal effect by modeling travel mode choice based on three conventional pre-defined boundaries (i.e. census block, block group and TAZ) and five grid sizes (i.e. 1/16-, 1/4-, 1/2-, 1-, 2-miles). T-statistics of estimated coefficients within each of the two data aggregation schemes were computed. Their results showed that while coefficient estimates based on pre-defined areal units were unstable, estimates based on the grids changed more systematically and produced more tractable and stable results. They recommended the grid system with resolution of 0.5 mile among the eight discussed zoning schemes.

With the availability of greater computing power, the MAUP continues to attract attention for studying spatial problems (Zhang and Kukadia 2005; Ratcliffe 2005; Lembo et al. 2006). GIS makes it possible to carry out extensive empirical experiments so that researchers are able to explore the underlying relationship of large amounts of data and no longer need to rely solely on statistical theory (Fotheringham and Rogerson 1993). In the next section, we will present the research design and preparation of data using GIS in order to investigate the aggregation effect on constrained shopping destination choice models.

3.3 Research Design and Data

3.3.1 Study Area and Data

The study area in this research spans seven counties of the Louisville, MSA. Three primary data sources are used in our analysis:

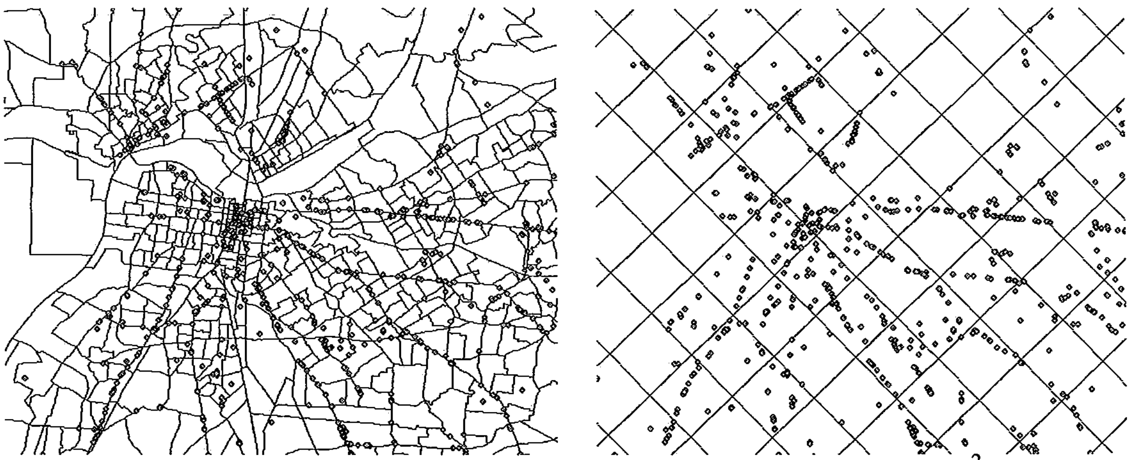
(1) a travel diary survey conducted in 2000 for seven counties of the Louisville MSA, which contains household, personal and one-day travel information. Approximately 4600 households participated in the survey, where a total of 30,888 trips were conducted.

(2) a 2002 Dynamap/Transportation 4.0 network produced by Geographic Data Technology Inc. (GDT). The traffic network contains traffic network information, such as direction, length, speed and cost.

(3) an urban opportunities file for the Louisville MSA as geocoded from a database obtained from ReferenceUSA. The file contains 34,440 opportunities. Each opportunity is classified by SIC code.

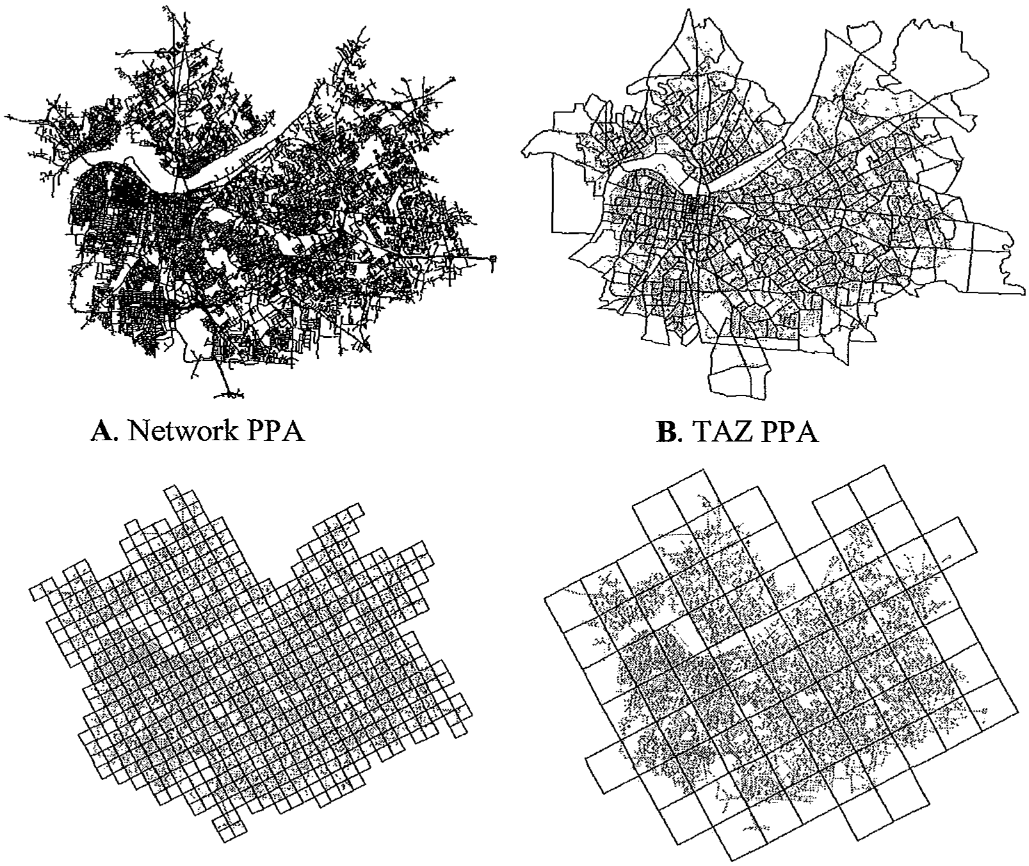
3.3.2 PPA Algorithm

With the initial data sources, all the shopping opportunities are broken down into specific categories using SIC codes. The shopping opportunities information is then aggregated to the TAZs and 10 randomly generated grid systems (from cell size 1 km^2 to 10 km^2 in 1 km^2 increments) (Figure 3.1). We applied Scott's (2006) algorithm to generate the network-based constrained destination choice sets (network PPAs) (Figure 3.2A). The algorithm selects a subset of links to make sure the origin and destination can be reached within an individual's time budget, then it searches thoroughly the links from the subset and determines which links are to be included in the PPA. A program in ArcMap was revised to select the TAZs and grids that intersect with the network PPAs in



A. TAZ **B.** grid of 10 km^2

Figure 3.1: Shopping opportunities aggregation based on (A) TAZ and (B) grid of 10 km^2.



A. Network PPA

B. TAZ PPA

C. Grid PPA of 1 km^2

D. Grid PPA of 10 km^2

Figure 3.2: Network PPA and grid PPA (on TAZs, grids of 1 km^2 and 10 km^2).

order to turn the network PPAs into TAZ/grid-based PPAs (Figures 3.2B, C and D). In this way, the polygon PPAs become time-constrained choice sets for shopping. Although in the real world people often undertake multipurpose trips (Ewing et al. 1994), for simplicity's sake, only single purpose shopping trips/PPAs are selected from the original travel survey. The size of the choice set varies since the geographic locations of the mandatory activities before and after the shopping activity, the individual's time budget and the speed associated with each link in the road network are different from each other.

3.3.3 Model Specification

Disaggregate discrete choice models are based on discrete consumer choice observed in real markets and random utility theory. The best-known of them is the multinomial logit (MNL) model (McFadden 1974, Ben-Akiva and Lerman 1985). Several previous studies estimated shopping destination choice parameters using the multinomial logit model (Pellegrini et al. 1997; Miller and O'Kelly 1983; Timmermans et al. 1984). As applied in transport research, the utility of taking part in an activity at a location depends on the attractiveness of that location, the travel time and the activity time (Ettema and Timmermans 2004; Miller 1999).

The utility of destination j for individual n is specified as:

$$U_{jn} = V_{jn} + \varepsilon_{jn}$$
$$V_{jn} = \sum_{k} \beta_k X_{kj} \qquad \forall j \in C_n,\ \forall C_n \in C \tag{1}$$

where X_{kj} is a matrix of the destination attributes and temporal factors, β_k is a vector of coefficients of the zonal characteristics and temporal factors for each zone, and ε is the disturbance and it is assumed independent. C_n is the constrained choice set for individual n. An individual will choose destination j if and only if

$$U_{jn} \geq U_{ln} \qquad \forall j, l \in C_n \tag{2}$$

The probability that an individual chooses destination j conditional on the choice set is:

$$P_n(j|C_n) = \frac{e^{V_{jn}}}{\sum_{l \in C_n} e^{V_{ln}}} \tag{3}$$

Two types of destination attributes are considered for the shopping destination choice models developed for this research: number of shopping opportunities (Table 3.1) and temporal factors. This study decomposes shopping activities into two categories: grocery, and non-grocery. Thirty one dummy variables corresponding to four-digit SIC codes are created to help classify the shopping opportunities into grocery, grocery/nongrocery, and nongrocery. A store variety index is created to indicate the range of stores in the each zone. It is defined as the ratio of the number of store types in a particular zone to the total number of defined store types (i.e., 31):

Store diversity = number of types of the 31 stores in the zone / 31. (4)

The temporal factors include: generated travel time from the first mandatory activity zone to the shopping zone (Time1), generated travel time from the shopping zone to the second mandatory activity zone (Time2)[5], and generated activity duration for each zone (Activity) is obtained by subtracting the difference between a shopper's actual travel time and the sum of two parts of generated travel time from a shopper's actual activity duration:

Activity duration =Actual activity duration - ((Time 1+ Time 2) - actual travel time) (5)

One important component of the variables in the model is zonal level travel time. In TransCAD, the centroid of the each TAZ and grid was generated and travel time from centroid to centroid was calculated. However the software assumes the inner zonal time to be zero, which unreasonable. To obtain the inner zonal travel time, a conventional method for transforming the grids and the irregular shapes of TAZs into circles of the corresponding area was adopted. The travel time is obtained by dividing the radius of

[5] The generation of travel time variables is mainly for the computation of the activity duration variable. Although another set of models using travel time instead of activity duration is also estimated, it results in similar findings. Given that the activity duration variable accounts for an individual's time budget and makes more sense in activity-based research, models using activity duration are presented in this paper.

Table 3.1: Definition of the four-digit SIC code and corresponding explanatory variables

Shopping category	SIC code	Description of destination attractiveness variables	Frequency	Number of stores
Grocery	5411	Grocery stores	281	463
	5431	Fruit and vegetable markets	0	29
	5441	Candy, nut, and confectionery stores	1	23
	5461	Retail bakeries	1	121
Grocery/nongrocery	5301	Wal-Mart	38	7
	5401	Meijer	32	5
	5912	Drug stores and proprietary stores	49	150
	5311	Department stores	73	107
	5331	Variety stores	5	26
Nongrocery	5211	Lumber and other building materials dealers	13	157
	5251	Hardware stores	3	53
	5261	Retail nurseries, lawn and garden supply stores	1	92
	5511	Motor vehicle dealers (new and used)	6	321
	5531	Auto and home supply stores	8	238
	5621	Women's' clothing stores	2	119
	5661	Shoe stores	1	99
	5712	Furniture stores	1	186
	5731	Radio, television, and consumer electronics stores	6	79
	5735	Record and pre-recorded tape stores	1	91
	5932	Used merchandise stores	5	263
	5941	Sporting goods stores and bicycle shops	2	152
	5942	Book stores	5	50
	5943	Stationery stores	2	37
	5944	Jewellery stores	3	118
	5945	Hobby, toy, and game shops	5	80
	5949	Sewing, needlework, and piece goods stores	1	17
	5992	Florists	1	118
	5993	Tobacco stores and stands	1	33
	5995	Optical goods stores	2	53
	5999	Miscellaneous retail stores, not elsewhere classified	3	374
	6502	Shopping malls and centers	64	19
Total			616	3680

Note: The grocery/nongrocery shopping category refers to those shopping opportunities where shoppers can conduct both grocery and nongrocery shopping activities. Shoppers can conduct grocery shopping in some department stores (i.e., Wal-mart, K-Mart, SAM's club). The separation of Wal-Mart from department stores and Meijer from grocery stores is because of their relative influential role in affecting shoppers' destination choice.

the circle by the average speed in the zone. The average speed is calculated by:

$$Speed_j = \frac{\sum_i Length_i}{\sum_i Cost_i} \tag{6}$$

where j denotes the number of zones and i indicates the number of links within each zone. A program was written in ArcMap to get the summation of *Length* and *Cost*. The inner zonal time was then updated in the travel time matrix from TransCAD.

Three types of PPAs are defined. If an individual's actual shopping destination is a grocery store, his/her PPA is defined as a grocery PPA. Nongrocery PPAs and grocery/nongrocery PPAs are defined likewise. Table 3.2 shows the three shopping models in this paper. Each type of models includes the corresponding PPAs. For instance, grocery, grocery/nongrocery PPAs are locations where grocery shopping can be conducted, therefore they are included in the grocery shopping model (Model 2).

Table 3.2: Model definition

Model	**Shopping type**	**PPAs included**	**Sample size**
1	General	Grocery, grocery/nongrocery, nongrocery	481
2	Grocery	Grocery, grocery/nongrocery	295
3	Nongrocery	Nongrocery, grocery/nongrocery	264

3.3.4 Choice Set Generation

There are a large number of alternatives in the constrained choice sets. We adopted a straightforward way of drawing a subset, which is comprised of the actual chosen zone plus nine random non-chosen zones within the PPA. This sampling approach requires that in a PPA there are at least ten zones which contain appropriate shopping opportunities. Figure 3.3 displays two types of invalid PPAs that have been removed. In the first PPA (A), there are less than 10 zones; in the second PPA (B), there are more than ten zones, but less than ten zones contain appropriate shopping opportunities.

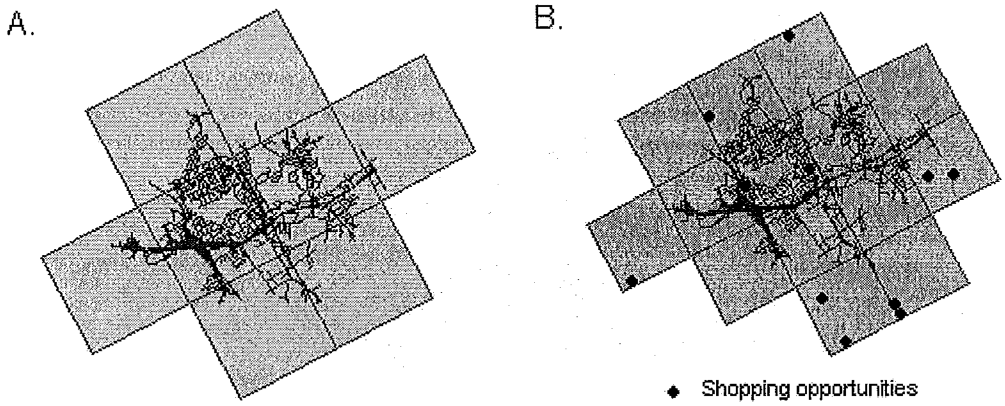


Figure 3.3: Invalid PPAs.

Only the samples (PPAs) that are valid in all zoning systems are selected and modeled. In total, there are 295 grocery shopping PPAs, 264 nongrocery shopping PPAs and 481 general shopping PPAs, which are used in all modeling (Table 3.3).

Table 3.3: Changes in valid samples across different zoning systems

Zoning system	**Number of PPAs**		
	General	*Grocery*	*Nongrocery*
TAZ	616	428	328
1 km^2	620	401	342
2 km^2	638	447	354
3 km^2	637	451	348
4 km^2	617	454	340
Zoning system	**Number of PPAs**		
	General		*General*
5 km^2	615	447	339
6 km^2	605	430	341
7 km^2	599	434	336
8 km^2	589	429	331
9 km^2	590	431	326
10 km^2	581	430	328
Mean	610	435	338
In common	481	295	264

Since TAZs have been used widely in current transportation research and model estimations based on TAZ are relatively more comparable to related studies in the field than other zoning systems, we use the model calibrated from the TAZ zoning system as a benchmark against which the modeling results from the 10 grid zoning systems are compared. To better tell the difference in estimated coefficients, the same set of explanatory variables used in the TAZ model are also used in the models for the 10 grid zoning systems. The comparison focuses on parameter coefficients, standard errors and different measures of goodness-of-fit.

3.4 Results

3.4.1 Constrained Choice Sets

As shown in Figure 3.2, an individual's constrained choice set varies across the 11 zoning systems. Table 3.4 shows that for general shopping, which includes all PPAs, the average size of the choice set decreases as the grid size increases, from 563 destination zones in the 1 km^2 grid to 73 zones as the grid size increases to 10 km^2. The ratio of zones in an individual's choice set to zones in the universal choice set steadily increases with the mean proportion going up from 9.9% to 11.7%. All the TAZ zones are accessible in the largest PPA (100%), but not all the zones in the grid systems (88.2% to 95.5%). The reason for the lower percentage of accessibility in grid systems is that not the entire study area is covered with a traffic network, especially in suburban and rural areas, making some of the grids inaccessible when the grids are superimposed onto the traffic network.

Table 3.4: Statistics of constrained destination choices across different zoning systems

System	Universal choice set	Constrained choice set				Ratio of individual to universal choice set		
		Mean	Std.	Min	Max	Mean	Min	Max
TAZ	818	338	232	24	818	0.413	0.029	1.000
1 km^2	5713	563	649	44	5039	0.099	0.008	0.882
2 km^2	2916	302	345	29	2693	0.104	0.010	0.924
3 km^2	1965	210	237	23	1847	0.107	0.012	0.940
4 km^2	1493	163	180	20	1407	0.109	0.013	0.942
5 km^2	1207	134	147	16	1140	0.111	0.013	0.945
6 km^2	1016	114	124	13	956	0.112	0.013	0.941
7 km^2	872	99	107	11	828	0.114	0.013	0.950
8 km^2	767	89	95	11	727	0.116	0.014	0.948
9 km^2	687	80	86	11	656	0.117	0.016	0.955
10 km^2	625	73	77	10	591	0.117	0.016	0.946

3.4.2 Parameter Estimations

Estimation of the three models reveals the sensitivity of variables to variations in the zoning systems (Table 3.5-3.7). For example, Figure 3.4 shows that the estimation of attractiveness variables characterizing TAZs are quite different from those characterizing the grid systems in Model 1 (general shopping). Lines are more clustered in the grid systems, which demonstrate that the coefficient estimates across the zoning systems do not vary considerably, except for Meijer, variety stores and stationery stores. The most outspread line is found for Meijer (0.4635 to 3.0197). The coefficients for variety stores and stationery stores also differ substantially with respect to the TAZ model. The number of variety stores showed a negative influence on the destination choice in the TAZ model while it shows a positive effect for most of the grid models with the exception of the 1 km^2 and 4 km^2 grid models. Most of the attractiveness variables fail to evince the same sign of estimates in all zoning systems, except for candy stores, Meijer and department stores. Relative to other parameters, the estimated coefficient of grocery stores decreases

Table 3.5: General shopping models in 11 zoning systems (Model 1)

	TAZ		1 km^2		2 km^2		3 km^2		4 km^2		5 km^2	
	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio
Grocery stores	0.1493	2.6762	0.2487	4.5399	0.0590	1.3166	0.0620	1.4592	0.0034	0.0980	0.0348	1.1489
Candy stores	-0.7281	-2.4985	-0.3069	-1.1289	-0.4049	-1.9373	-0.4480	-2.5322	-0.7874	-4.6379	-0.2945	-2.0367
Bakeries	-0.4476	-3.5507	-0.2619	-2.4946	-0.2731	-3.0240	-0.2284	-2.9539	-0.4331	-5.7550	-0.2578	-3.6566
Meijer	2.5113	7.3117	3.0197	7.5636	1.8279	5.1367	0.8148	2.2476	1.0098	3.3648	0.7485	2.2083
Department stores	0.1784	1.9935	0.2527	2.8649	0.0981	1.6551	0.2959	4.1958	0.3015	5.1796	0.1871	3.9281
Variety stores	-1.1246	-4.0302	-0.0676	-0.3444	0.1463	0.8860	0.2634	1.7007	-0.0108	-0.0874	0.1045	0.8749
Motor vehicle dealers	-0.2155	-3.5798	-0.0183	-0.3437	-0.0283	-0.9372	0.0068	0.2288	-0.0253	-1.0954	-0.0695	-2.9631
Furniture stores	0.2127	3.1275	-0.0090	-0.1446	-0.0083	-0.1805	-0.1308	-2.4482	-0.0748	-1.8802	0.0521	1.3471
Record tape stores	-0.3825	-3.2132	-0.1237	-1.1931	-0.1938	-2.3969	-0.2682	-3.2149	0.0277	0.4718	-0.1156	-1.9267
Stationery stores	-1.7015	-4.9076	-0.6677	-3.4260	-0.5875	-3.3927	-0.3124	-2.6585	0.0693	0.6442	-0.3229	-3.7056
Tobacco stores	-0.4554	-2.0568	0.2558	1.2631	-0.1208	-0.6407	-0.1615	-0.8909	0.0159	0.1158	-0.0481	-0.3022
Shopping malls	0.6296	2.9484	0.4308	2.0964	0.4795	2.8353	0.7777	4.8992	0.3137	2.0691	0.1815	1.1474
Store diversity	9.3849	10.4696	5.1504	7.0294	6.5070	10.1717	5.6909	8.5929	5.6030	8.6869	6.1212	8.8296
Activity duration	0.1349	16.0858	0.1291	16.1286	0.1238	15.6346	0.1504	17.8947	0.1123	15.4293	0.1400	17.5669
Summary statistics												
n	481		481		481		481		481		481	
$L(0)$	-1107.5434		-1107.5434		-1107.5434		-1107.5434		-1107.5434		-1107.5434	
$L(\beta)$	-686.5655		-732.8951		-694.6324		-673.0818		-710.3162		-661.3919	
ρ^2	0.3801		0.3383		0.3728		0.3923		0.3587		0.4028	
$\bar{\rho}^2$	0.3675		0.3256		0.3602		0.3796		0.3460		0.3902	

Table 3.5: General shopping models in 11 zoning systems (Model 1) (continued)

	6 km²		**7 km²**		**8 km²**		**9 km²**		**10 km²**	
	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio
Grocery stores	-0.0310	-1.1715	-0.0222	-0.9600	0.0707	3.2276	0.0012	0.0559	-0.0110	-0.4618
Candy stores	-0.4672	-3.4691	-0.4030	-3.7469	-0.2107	-1.4678	-0.1859	-1.6042	-0.2302	-1.9734
Bakeries	-0.1021	-1.9270	-0.0145	-0.2440	-0.2528	-4.0249	-0.0974	-1.5353	0.0093	0.1868
Meijer	0.4635	1.4506	1.3225	5.7097	0.8015	2.9740	1.1778	4.6734	1.9912	7.5569
Department stores	0.1738	3.4901	0.2444	5.0389	0.1180	3.0315	0.0812	2.0666	0.1376	3.2122
Variety stores	0.1435	1.1194	0.1614	1.3066	-0.0929	-0.7095	0.1683	1.3233	0.3855	3.0609
Motor vehicle dealers	-0.0349	-1.6817	-0.0897	-4.0780	-0.0158	-0.8687	-0.0656	-3.1292	-0.1031	-4.6321
Furniture stores	-0.0956	-2.7323	-0.0170	-0.4735	-0.0239	-0.7703	0.0335	0.8469	0.0078	0.2421
Record tape stores	-0.2681	-5.2935	-0.1551	-2.7331	-0.0316	-0.5744	-0.0286	-0.6832	-0.0985	-2.4984
Stationery stores	0.0163	0.1873	-0.1135	-1.4524	-0.0104	-0.1668	-0.0079	-0.1126	-0.0175	-0.2873
Tobacco stores	-0.0449	-0.2856	-0.1679	-1.3585	0.2378	1.6998	-0.3458	-2.4139	-0.6024	-3.6846
Shopping malls	0.0347	0.3016	-0.0392	-0.3810	0.5212	4.2832	0.1811	1.5032	-0.2635	-2.4308
Store diversity	7.0214	12.3319	6.0429	9.2753	3.2087	5.7284	5.4115	8.0617	6.6634	10.2497
Activity duration	0.1290	16.5625	0.1180	15.6638	0.1258	16.9646	0.1245	16.7261	0.1452	18.6095
Summary statistics										
n	481		481		481		481		481	
$L(0)$	-1107.5434		-1107.5434		-1107.5434		-1107.5434		-1107.5434	
$L(\beta)$	-670.4548		-692.5623		-709.2841		-691.5624		-671.2398	
ρ^2	0.3947		0.3747		0.3596		0.3756		0.3939	
$\bar{\rho}^2$	0.3820		0.3620		0.3469		0.3629		0.3813	

Table 3.6: Grocery shopping models in 11 zoning systems (Model 2)

	TAZ		1 km²		2 km²		3 km²		4 km²		5 km²	
	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio
Grocery stores	0.3713	5.9526	0.2006	3.4161	0.1772	3.3064	0.1374	3.0145	0.1404	3.5429	0.0848	2.5850
Meijer	2.3582	6.8629	2.5796	7.4865	2.2325	6.0909	1.9036	6.2280	0.9405	2.8108	2.0338	6.5061
Department stores	0.1709	1.9593	0.0925	1.1312	0.1404	2.2116	-0.0163	-0.2738	0.0972	1.7173	0.0434	0.9283
Variety stores	-0.8145	-2.2498	0.1810	0.8522	0.6525	3.7931	0.4026	2.3029	0.4144	2.9381	0.6917	5.1019
Store diversity	3.6061	5.1709	3.4863	5.2208	2.7781	5.3508	2.7978	5.5712	1.6443	3.3060	1.3525	2.6219
Activity duration	0.1613	14.4210	0.1721	14.5696	0.1546	13.7385	0.1546	14.2051	0.1551	14.3647	0.1549	15.1764
Summary statistics												
n	295		295		295		295		295		295	
$L(0)$	-679.2626		-679.2626		-679.2626		-679.2626		-679.2626		-679.2626	
$L(\beta)$	-459.5411		-465.9981		-435.8751		-447.5576		-455.0374		-444.0867	
ρ^2	0.3235		0.3140		0.3583		0.3411		0.3301		0.3462	
$\bar{\rho}^2$	0.3146		0.3051		0.3495		0.3323		0.3213		0.3374	

	6 km²		7 km²		8 km²		9 km²		10 km²	
	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio
Grocery stores	0.0253	0.9901	0.0381	1.4210	0.0458	2.0922	0.0421	2.2249	0.0119	0.6521
Meijer	0.5236	1.7345	1.4401	5.9103	0.6389	2.1842	2.0147	7.9005	2.0008	8.1215
Department stores	0.0555	1.0651	0.0263	0.5328	0.0300	0.8962	0.0404	1.0808	0.0000	-0.0010
Variety stores	0.5805	4.1443	0.5623	4.5281	0.4794	3.7145	0.5274	4.0223	0.4041	3.0237
Store diversity	1.9090	4.2198	1.7684	3.3936	0.9298	1.8713	1.4005	2.5828	2.5305	4.7952
Activity duration	0.1451	14.1473	0.1585	14.9015	0.1661	15.4973	0.1678	15.4822	0.1440	14.7485
Summary statistics										
n	295		295		295		295		295	
$L(0)$	-679.2626		-679.2626		-679.2626		-679.2626		-679.2626	
$L(\beta)$	-468.9302		-463.7892		-456.2913		-421.1592		-447.8673	
ρ^2	0.3097		0.3172		0.3283		0.3800		0.3407	
$\bar{\rho}^2$	0.3008		0.3084		0.3194		0.3711		0.3318	

Table 3.7: Nongrocery shopping models in 11 zoning systems (Model 3)

	TAZ		1 km^2		2 km^2		3 km^2		4 km^2		5 km^2	
	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio
Meijer	3.1811	7.4794	2.8544	7.2459	2.5762	6.1895	1.4363	3.9281	1.4141	4.3383	1.5276	4.5307
Drugstores	0.3479	2.5783	0.1878	1.8695	0.0500	0.6029	0.1787	2.3979	0.0470	0.5880	0.1220	1.7506
Department stores	0.3581	4.0613	0.3669	4.3256	0.1564	2.5361	0.3042	5.1470	0.2770	5.0950	0.1550	3.4332
Sewing stores	-0.7536	-2.0222	0.5727	2.2350	-0.2831	-1.3805	0.0675	0.3579	-0.5664	-2.5457	0.1890	1.2184
Store diversity	4.8886	5.7980	2.6547	3.3600	4.8088	7.8120	2.7945	4.2442	3.8846	6.6155	3.9776	6.8688
Activity duration	0.0912	9.4525	0.0942	10.0230	0.0922	10.0081	0.0948	10.8164	0.0837	9.5552	0.0959	10.4854
Summary statistics												
n	264		264		264		264		264		264	
$L(0)$	-607.8825		-607.8825		-607.8825		-607.8825		-607.8825		-607.8825	
$L(\beta)$	-376.3219		-411.0581		-385.1614		-388.1344		-385.5640		-351.0889	
ρ^2	0.3809		0.3238		0.3664		0.3615		0.3657		0.4224	
$\bar{\rho}^2$	0.3711		0.3139		0.3565		0.3516		0.3559		0.4126	

	6 km^2		7 km^2		8 km^2		9 km^2		10 km^2	
	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio
Meijer	0.7759	2.2562	2.0051	7.4946	1.4633	5.2458	1.6608	5.9301	1.9157	6.7236
Drugstores	-0.0497	-1.0404	-0.0566	-1.0842	-0.0473	-0.7791	0.0447	0.9511	-0.0090	-0.1936
Department stores	0.1437	2.6996	0.3239	6.3834	0.1924	5.9430	0.1479	4.2648	0.2091	5.2859
Sewing stores	-0.0478	-0.3113	0.0922	0.5168	-0.3130	-1.8298	0.1988	1.5019	0.0841	0.6354
Store diversity	4.6211	8.3571	4.3519	6.8949	3.4188	6.1940	3.3457	5.3752	3.6737	5.9951
Activity duration	0.0958	10.2098	0.1094	11.3850	0.0917	10.6066	0.1000	10.8470	0.1106	11.7132
Summary statistics										
n	264		264		264		264		264	
$L(0)$	-607.8825		-607.8825		-607.8825		-607.8825		-607.8825	
$L(\beta)$	-354.0311		-335.1089		-372.5486		-354.6251		-350.1163	
ρ^2	0.4176		0.4487		0.3871		0.4166		0.4240	
$\bar{\rho}^2$	0.4077		0.4389		0.3773		0.4068		0.4142	

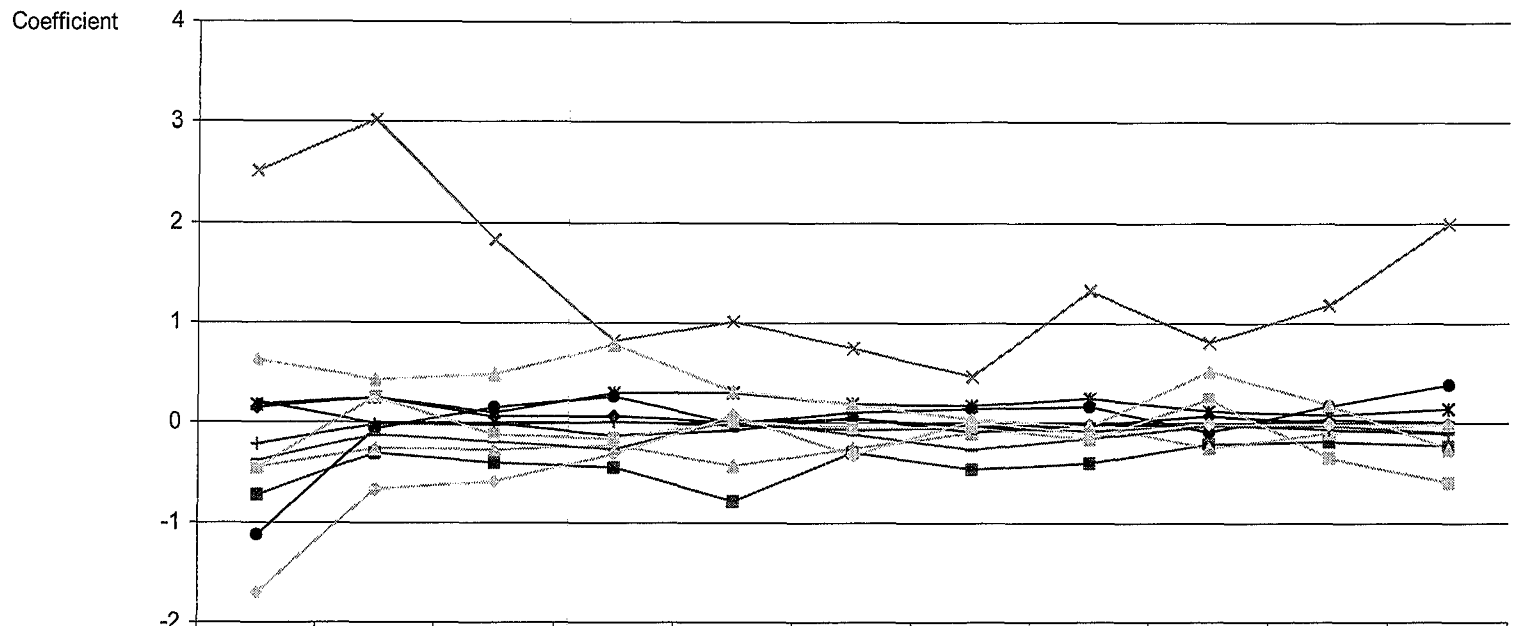
systematically with grid aggregation whilst the magnitudes of these coefficients shrink as the grid size increases. In other words, their marginal effects decrease at lower resolutions.

Also for Model 1, Figure 3.5 shows that the standard errors of the coefficients tend to be smaller at lower resolutions. In the TAZ zoning system, the standard error of each attractiveness variable is nearly always higher than other grid zoning systems except at Meijer where the standard error of 1 km^2 grid is larger than the TAZ. When comparing Figure 3.4 and Figure 3.5, variation of standard error also shows a similar trend to variation of the coefficient. Parameters that have larger variation in coefficients across the zoning systems also appear to have larger variation in standard errors, such as Meijer, variety stores, stationery stores and candy stores. Similar trends are found in Model 2 (grocery shopping) and Model 3 (nongrocery shopping) (Figure 3.6-3.9).

Figure 3.10 shows that activity duration is relatively resistant to the aggregation effect with no obvious trend in Model 1. Activity duration and store diversity index are highly significant in the three shopping models in all 11 zoning systems. From Figure 3.11, magnitudes of the store diversity index in the grid zoning systems are smaller than the TAZ model, with the lowest value in 8 km^2, which illustrates its smaller marginal effect at higher aggregation level.

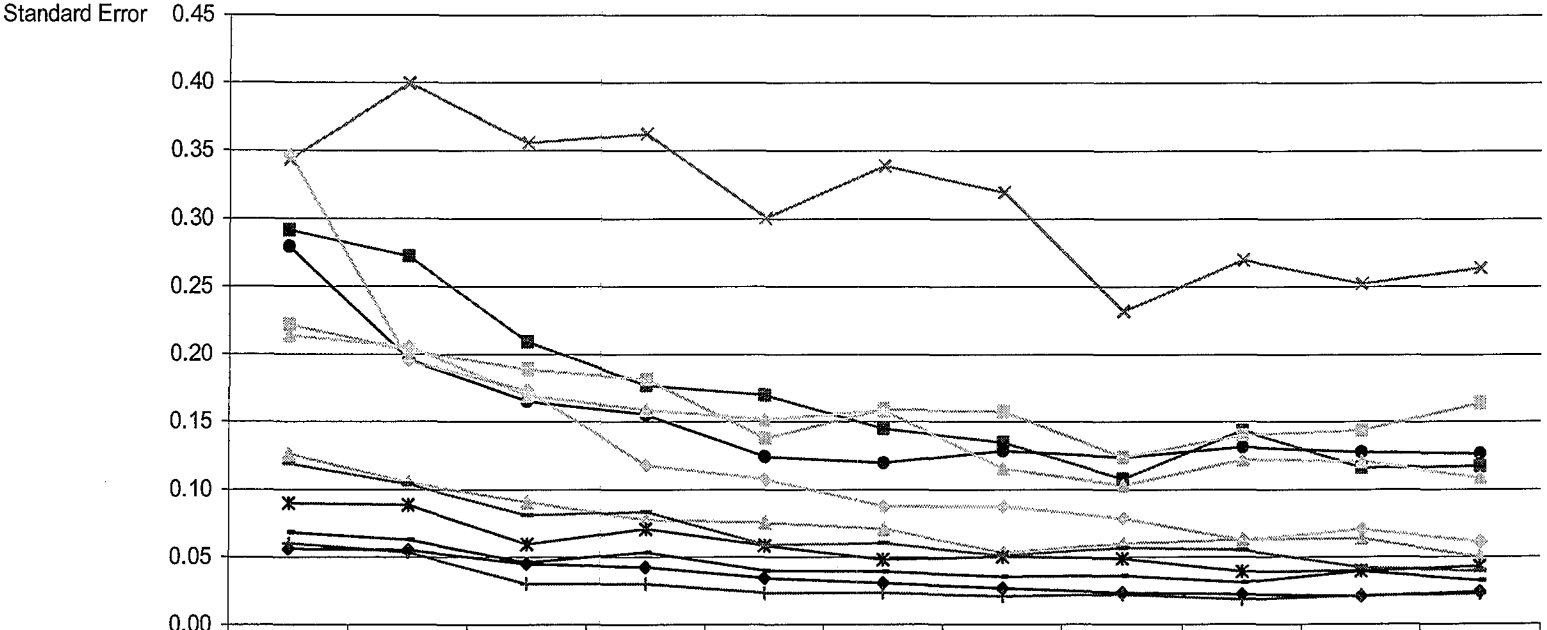
Using the estimates from the TAZ model and grid models, t-statistics are calculated to examine the difference between estimated coefficients.

$$t = \frac{\hat{\beta}_{grid} - \hat{\beta}_{TAZ}}{std.error(\hat{\beta}_{grid} - \hat{\beta}_{TAZ})} \tag{7}$$



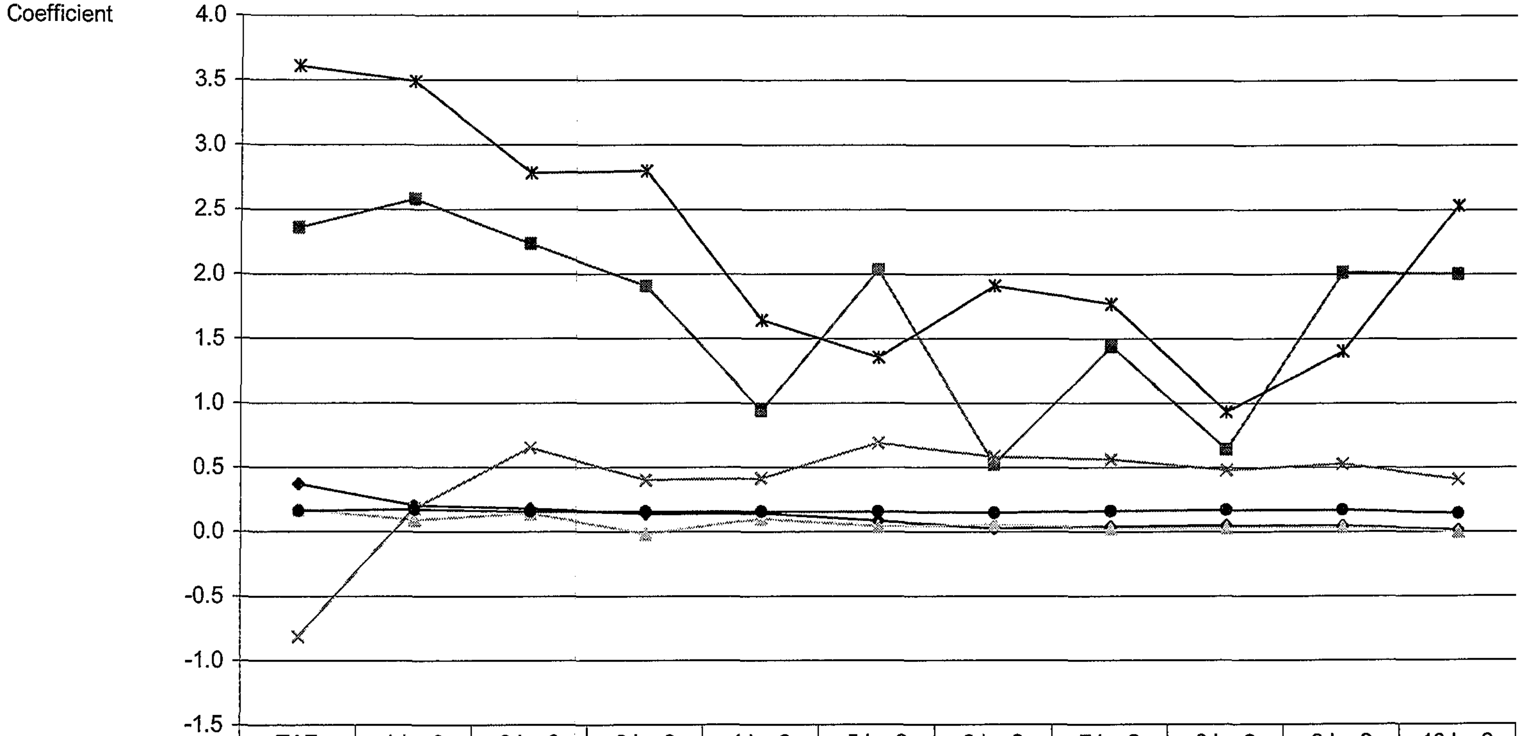
	TAZ	1 km2	2 km2	3 km2	4 km2	5 km2	6 km2	7 km2	8 km2	9 km2	10 km2
Grocery stores	0.1493	0.2487	0.0590	0.0620	0.0034	0.0348	-0.0310	-0.0222	0.0707	0.0012	-0.0110
Candy stores	-0.7281	-0.3069	-0.4049	-0.4480	-0.7874	-0.2945	-0.4672	-0.4030	-0.2107	-0.1859	-0.2302
Bakeries	-0.4476	-0.2619	-0.2731	-0.2284	-0.4331	-0.2578	-0.1021	-0.0145	-0.2528	-0.0974	0.0093
Meijer	2.5113	3.0197	1.8279	0.8148	1.0098	0.7485	0.4635	1.3225	0.8015	1.1778	1.9912
Department stores	0.1784	0.2527	0.0981	0.2959	0.3015	0.1871	0.1738	0.2444	0.1180	0.0812	0.1376
Variety stores	-1.1246	-0.0676	0.1463	0.2634	-0.0108	0.1045	0.1435	0.1614	-0.0929	0.1683	0.3855
Motor vehicle dealers	-0.2155	-0.0183	-0.0283	0.0068	-0.0253	-0.0695	-0.0349	-0.0897	-0.0158	-0.0656	-0.1031
Furniture stores	0.2127	-0.0090	-0.0083	-0.1308	-0.0748	0.0521	-0.0956	-0.0170	-0.0239	0.0335	0.0078
Record tape stores	-0.3825	-0.1237	-0.1938	-0.2682	0.0277	-0.1156	-0.2681	-0.1551	-0.0316	-0.0286	-0.0985
Stationery stores	-1.7015	-0.6677	-0.5875	-0.3124	0.0693	-0.3229	0.0163	-0.1135	-0.0104	-0.0079	-0.0175
Tobacco stores	-0.4554	0.2558	-0.1208	-0.1615	0.0159	-0.0481	-0.0449	-0.1679	0.2378	-0.3458	-0.6024
Shopping malls	0.6296	0.4308	0.4795	0.7777	0.3137	0.1815	0.0347	-0.0392	0.5212	0.1811	-0.2635

Figure 3.4: Effects of store numbers among the 11 zoning systems (Model 1).



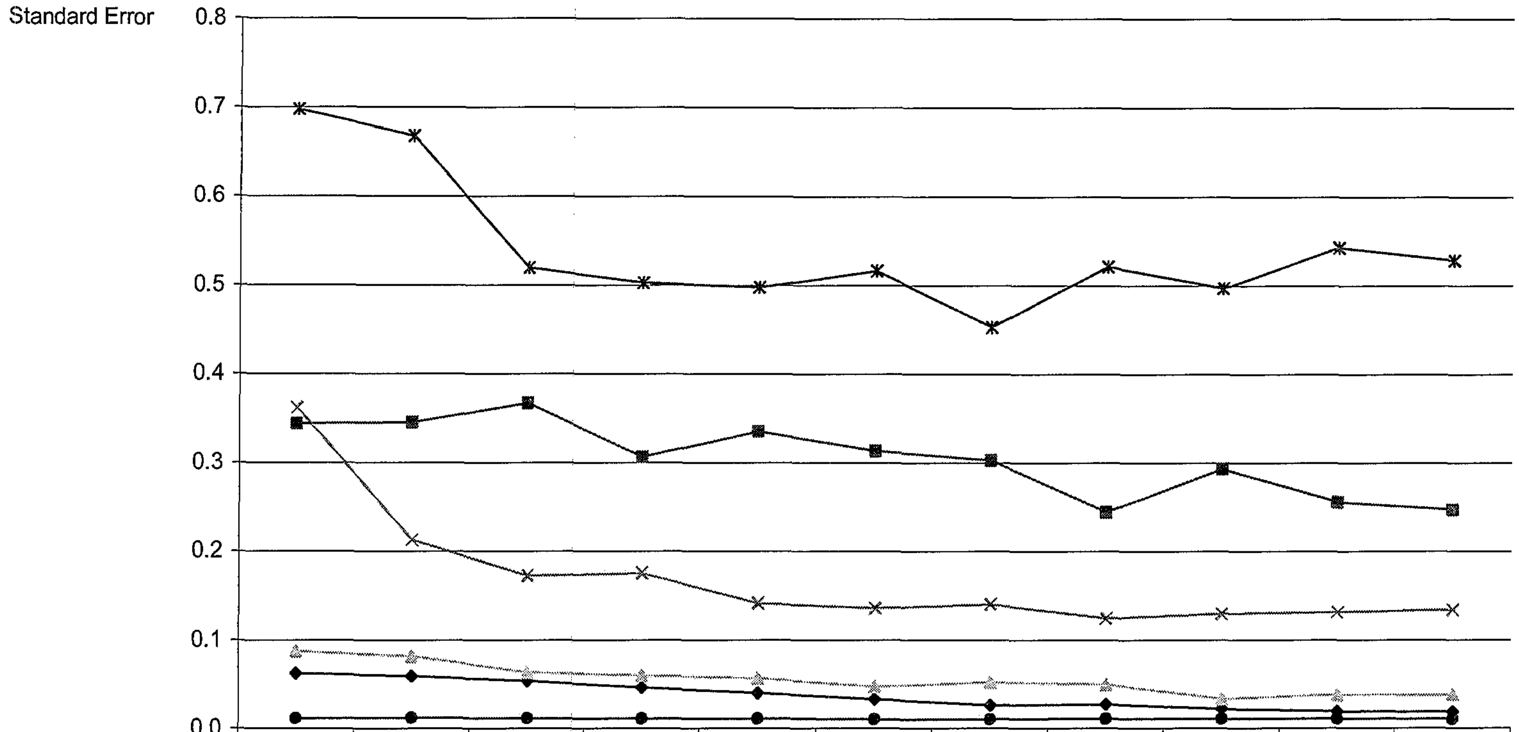
	TAZ	1 km2	2 km2	3 km2	4 km2	5 km2	6 km2	7 km2	8 km2	9 km2	10 km2
Grocery stores	0.0558	0.0548	0.0448	0.0425	0.0343	0.0303	0.0265	0.0231	0.0219	0.0208	0.0238
Candy stores	0.2914	0.2719	0.2090	0.1769	0.1698	0.1446	0.1347	0.1075	0.1435	0.1159	0.1167
Bakeries	0.1260	0.1050	0.0903	0.0773	0.0753	0.0705	0.0530	0.0595	0.0628	0.0635	0.0496
Meijer	0.3435	0.3992	0.3558	0.3625	0.3001	0.3389	0.3195	0.2316	0.2695	0.2520	0.2635
Department stores	0.0895	0.0882	0.0592	0.0705	0.0582	0.0476	0.0498	0.0485	0.0389	0.0393	0.0428
Variety stores	0.2790	0.1963	0.1651	0.1549	0.1239	0.1194	0.1282	0.1235	0.1309	0.1272	0.1259
Motor vehicle dealers	0.0602	0.0534	0.0302	0.0299	0.0231	0.0235	0.0207	0.0220	0.0182	0.0209	0.0223
Furniture stores	0.0680	0.0624	0.0458	0.0534	0.0398	0.0387	0.0350	0.0360	0.0311	0.0395	0.0320
Record tape stores	0.1190	0.1036	0.0808	0.0834	0.0586	0.0600	0.0506	0.0568	0.0550	0.0419	0.0394
Stationery stores	0.3467	0.1949	0.1732	0.1175	0.1075	0.0871	0.0871	0.0781	0.0621	0.0701	0.0609
Tobacco stores	0.2214	0.2025	0.1886	0.1813	0.1373	0.1591	0.1572	0.1236	0.1399	0.1432	0.1635
Shopping malls	0.2136	0.2055	0.1691	0.1587	0.1516	0.1582	0.1149	0.1029	0.1217	0.1205	0.1084

Figure 3.5: Standard error of store numbers among 11 zoning systems (Model 1).



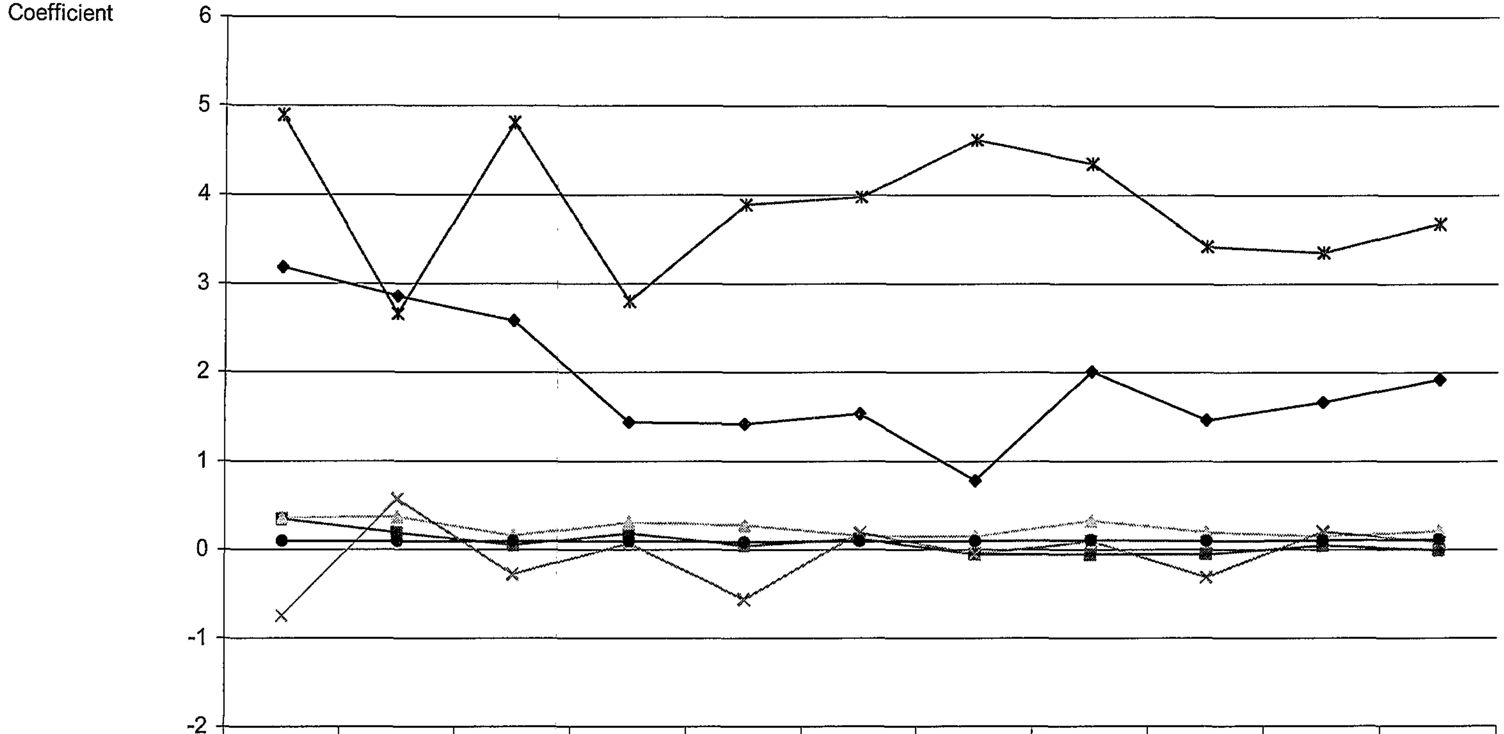
	TAZ	1 km2	2 km2	3 km2	4 km2	5 km2	6 km2	7 km2	8 km2	9 km2	10 km2
Grocery stores	0.3713	0.2006	0.1772	0.1374	0.1404	0.0848	0.0253	0.0381	0.0458	0.0421	0.0119
Meijer	2.3582	2.5796	2.2325	1.9036	0.9405	2.0338	0.5236	1.4401	0.6389	2.0147	2.0008
Department stores	0.1709	0.0925	0.1404	-0.0163	0.0972	0.0434	0.0555	0.0263	0.0300	0.0404	0.0000
Variety stores	-0.8145	0.1810	0.6525	0.4026	0.4144	0.6917	0.5805	0.5623	0.4794	0.5274	0.4041
Store diversity	3.6061	3.4863	2.7781	2.7978	1.6443	1.3525	1.9090	1.7684	0.9298	1.4005	2.5305
Activity duration	0.1613	0.1721	0.1546	0.1546	0.1551	0.1549	0.1451	0.1585	0.1661	0.1678	0.1440

Figure 3.6: Effects of explanatory variables among the 11 zoning systems (Model 2).



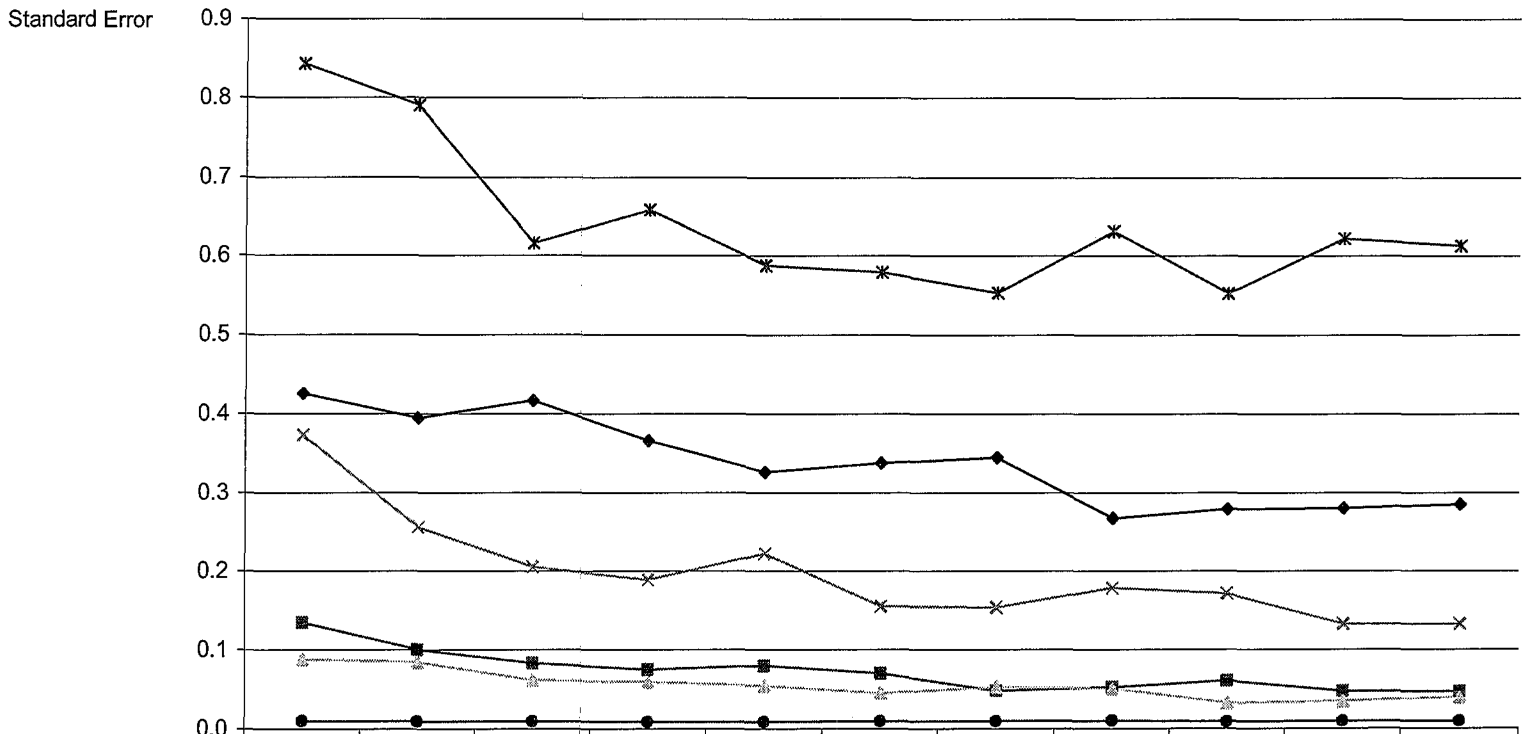
	TAZ	1 km2	2 km2	3 km2	4 km2	5 km2	6 km2	7 km2	8 km2	9 km2	10 km2
Grocery stores	0.0624	0.0587	0.0536	0.0456	0.0396	0.0328	0.0255	0.0268	0.0219	0.0189	0.0183
Meijer	0.3436	0.3446	0.3665	0.3057	0.3346	0.3126	0.3019	0.2437	0.2925	0.2550	0.2464
Department stores	0.0872	0.0818	0.0635	0.0594	0.0566	0.0468	0.0521	0.0493	0.0335	0.0374	0.0368
Variety stores	0.3620	0.2124	0.1720	0.1748	0.1410	0.1356	0.1401	0.1242	0.1291	0.1311	0.1336
Store diversity	0.6974	0.6678	0.5192	0.5022	0.4974	0.5159	0.4524	0.5211	0.4968	0.5422	0.5277
Activity duration	0.0112	0.0118	0.0113	0.0109	0.0108	0.0102	0.0103	0.0106	0.0107	0.0108	0.0098

Figure 3.7: Standard error of explanatory variables among 11 zoning systems (Model 2).



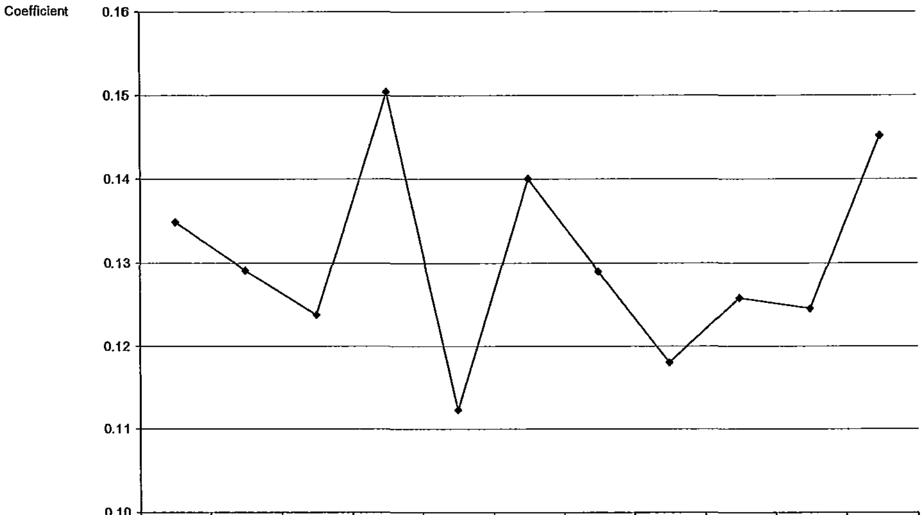
	TAZ	1 km2	2 km2	3 km2	4 km2	5 km2	6 km2	7 km2	8 km2	9 km2	10 km2
Meijer	3.1811	2.8544	2.5762	1.4363	1.4141	1.5276	0.7759	2.0051	1.4633	1.6608	1.9157
Drug stores	0.3479	0.1878	0.0500	0.1787	0.0470	0.1220	-0.0497	-0.0566	-0.0473	0.0447	-0.0090
Department stores	0.3581	0.3669	0.1564	0.3042	0.2770	0.1550	0.1437	0.3239	0.1924	0.1479	0.2091
Sewing stores	-0.7536	0.5727	-0.2831	0.0675	-0.5664	0.1890	-0.0478	0.0922	-0.3130	0.1988	0.0841
Store diversity	4.8886	2.6547	4.8088	2.7945	3.8846	3.9776	4.6211	4.3519	3.4188	3.3457	3.6737
Activity duration	0.0912	0.0942	0.0922	0.0948	0.0837	0.0959	0.0958	0.1094	0.0917	0.1000	0.1106

Figure 3.8: Effects of explanatory variables among the 11 zoning systems (Model 3).



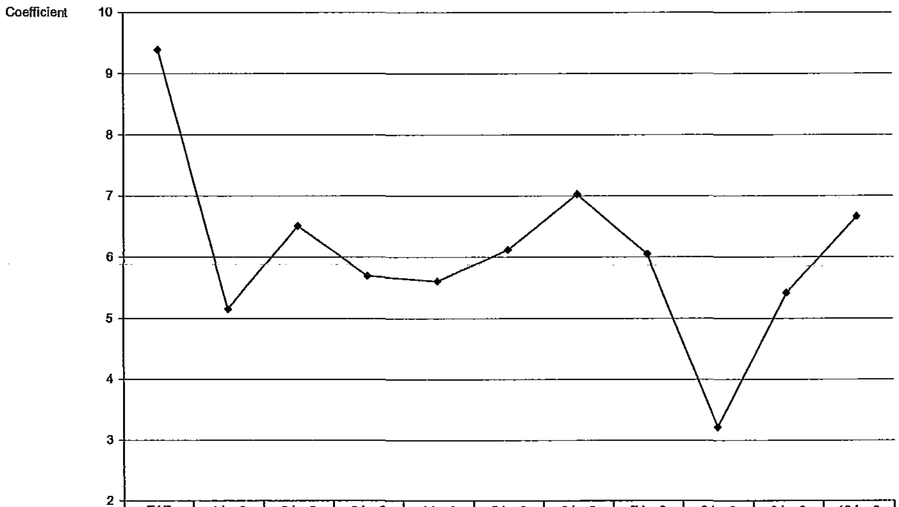
	TAZ	1 km2	2 km2	3 km2	4 km2	5 km2	6 km2	7 km2	8 km2	9 km2	10 km2
Meijer	0.4253	0.3939	0.4162	0.3656	0.3259	0.3372	0.3439	0.2675	0.2789	0.2801	0.2849
Drug stores	0.1349	0.1005	0.0830	0.0745	0.0799	0.0697	0.0478	0.0522	0.0607	0.0470	0.0467
Department stores	0.0882	0.0848	0.0617	0.0591	0.0544	0.0451	0.0532	0.0507	0.0324	0.0347	0.0396
Sewing stores	0.3726	0.2563	0.2051	0.1887	0.2225	0.1552	0.1535	0.1785	0.1711	0.1323	0.1324
Store diversity	0.8431	0.7901	0.6156	0.6584	0.5872	0.5791	0.5530	0.6312	0.5520	0.6224	0.6128
Activity duration	0.0097	0.0094	0.0092	0.0088	0.0088	0.0091	0.0094	0.0096	0.0086	0.0092	0.0094

Figure 3.9: Standard error of explanatory variables among 11 zoning systems (Model 3).



	TAZ	1 km2	2 km2	3 km2	4 km2	5 km2	6 km2	7 km2	8 km2	9 km2	10 km2
Activity duration	0.1349	0.1291	0.1238	0.1504	0.1123	0.1400	0.1290	0.1180	0.1258	0.1245	0.1452

Figure 3.10: Effect of activity duration among 11 zoning systems (Model 1).



	TAZ	1 km2	2 km2	3 km2	4 km2	5 km2	6 km2	7 km2	8 km2	9 km2	10 km2
Store diversity	9.3849	5.1504	6.5070	5.6909	5.6030	6.1212	7.0214	6.0429	3.2087	5.4115	6.6634

Figure 3.11: Effect of store diversity among 11 zoning systems (Model 1).

Since there are more explanatory variables in Model 1 than in Models 2 and 3, we take Model 1 as an example to test how differently the explanatory variables responded to the aggregation effect. In Table 3.8, the shaded cells indicate t-statistics significant at the 0.05 level (i.e., the hypothesis that the coefficient estimate in the grid model is equal to the TAZ model's coefficient is rejected). The greater the number of shaded cells in a row, the greater the dissimilarity between the TAZ and grid systems. Activity duration shows greater similarity between the TAZ and grids relative to the attractiveness variables. This finding is likely due to the parameters in the formula for activity duration, including actual travel time, actual activity duration and generated travel time. The shoppers' actual travel time and actual activity duration are not affected by aggregation. Generated travel time is a frameless network-based variable, not affected by the aggregation effect. Parameters such as Meijer, variety stores, motor vehicle dealers, furniture stores and stationery stores as well as the store diversity index differ most.

Table 3.8: t-statistics of difference between parameters estimated in grid zoning systems and TAZ (Model 1)

Variables	t-statistic									
	1 km^2	2 km^2	3 km^2	4 km^2	5 km^2	6 km^2	7 km^2	8 km^2	9 km^2	10 km^2
Grocery stores	1.2720	-1.2611	-1.2446	-2.2277	-1.8039	-2.9200	-2.8398	-1.3102	-2.4881	-2.6429
Candy stores	1.0568	0.9013	0.8216	-0.1758	1.3330	0.8126	1.0467	1.5929	1.7291	1.5862
Bakeries	1.1315	1.1253	1.4825	0.0983	1.3137	2.5267	3.1069	1.3833	2.4809	3.3727
Meijer	0.9653	-1.3818	-3.3970	-3.2922	-3.6531	-4.3654	-2.8695	-3.9165	-3.1303	-1.2015
Department stores	0.5913	-0.7485	1.0314	1.1531	0.0857	-0.0445	0.6486	-0.6183	-0.9947	-0.4114
Variety stores	3.0984	3.9197	4.3491	3.6478	4.0494	4.1296	4.2142	3.3471	4.2161	4.9326
Motor vehicle dealers	2.4507	2.7792	3.3076	2.9487	2.2591	2.8366	1.9633	3.1759	2.3523	1.7516
Furniture stores	-2.4021	-2.6944	-3.9718	-3.6489	-2.0519	-4.0310	-2.9859	-3.1648	-2.2780	-2.7265
Record tape stores	1.6398	1.3113	0.7858	3.0910	2.0015	0.8845	1.7238	2.6765	2.8041	2.2647
Stationery stores	2.5992	2.8746	3.7946	4.8782	3.8564	4.8053	4.4682	4.8014	4.7880	4.7838
Tobacco stores	2.3702	1.1505	1.0269	1.8091	1.4938	1.5117	1.1339	2.6467	0.4157	-0.5341
Shopping malls	-0.6707	-0.5510	0.5563	-1.2064	-1.6861	-2.4534	-2.8214	-0.4413	-1.8291	-3.7294
Store diversity	-3.6576	-2.6133	-3.3145	-3.4246	-2.8801	-2.2257	-3.0159	-5.8431	-3.5481	-2.4578
Activity duration	-0.5018	-0.9634	1.3052	-2.0342	0.4430	-0.5157	-1.4974	-0.8110	-0.9239	0.8985

However, candy store, department store and tobacco store parameters are surprisingly stable in different zoning schemes. In addition, few of the coefficients from the grid systems are approximate to the TAZ system with the finest grids producing coefficients that are closest to the TAZ system. For example, 1 km^2 has eight variables (un-shaded cells) while the 10 km^2 has six variables that have similar estimates to the TAZ values. However, the largest t-statistic for each explanatory variable is not always found for the largest grids, rather it is frequently found in the medium grids. For instance, the largest t-statistic for the furniture stores is found in grid size 6 km^2.

3.4.3 Goodness-of-fit

The same set of explanatory variables is used across the zoning system for each type of model. As expected, not all explanatory variables that are significant in the TAZ remained significant in the grid models. The inclusion of a few insignificant variables in the grid models will not dramatically increase their goodness-of-fit. To examine how the parameters in the TAZ model are changed in the grid models, a consistent set of parameters is needed. In this sense, we compare the goodness-of-fit of these models regardless of the significance of parameters. It is found that most of the ρ^2 of the grid models are lower than the TAZ model in the general shopping models, but higher in the grocery shopping models. The goodness-of-fit increases as the grid size increases in the case of nongrocery shopping (Table 3.9).

In order to measure the model prediction ability, we computed the percentage right for the three shopping models. Table 3.10 shows that the percentage right and expected percentage right of the model increases as the grid size increases, but not in a monotonic fashion. The higher percentage right of the larger grid is credited to fewer

Table 3.9: ρ^2 for general shopping, grocery shopping and nongrocery shopping models

	TAZ	**1 km^2**	**2 km^2**	**3 km^2**	**4 km^2**	**5 km^2**	**6 km^2**	**7 km^2**	**8 km^2**	**9 km^2**	**10 km^2**
General shopping											
Model 1	0.3801	0.3383	0.3728	0.3923	0.3587	0.4028	0.3947	0.3747	0.3596	0.3756	0.3939
Grocery shopping											
Model 2	0.3235	0.3140	0.3583	0.3411	0.3301	0.3462	0.3097	0.3172	0.3283	0.3800	0.3407
Nongrocery shopping											
Model 3	0.3809	0.3238	0.3664	0.3615	0.3657	0.4224	0.4176	0.4487	0.3871	0.4166	0.4240

Table 3.10: Percentage right using the full constrained choice set

		TAZ	1 km^2	2 km^2	3 km^2	4 km^2	5 km^2	6 km^2	7 km^2	8 km^2	9 km^2	10 km^2
General shopping												
Size of choice set	Mean	338	563	302	210	163	134	114	99	89	80	73
	Std. err.	232	649	345	237	180	147	124	107	95	86	77
n		481	481	481	481	481	481	481	481	481	481	481
Hits		38	42	64	68	69	101	94	76	94	118	112
% right		7.9002	8.7318	13.3056	14.1372	14.3451	20.9979	19.5426	15.8004	19.5426	24.5322	23.2848
Expected % right		5.3701	3.9162	6.0946	8.3804	7.1936	10.9415	11.4683	10.2954	11.163	14.1779	14.9358
Grocery shopping												
Size of choice set	Mean	315	483	260	181	140	115	99	86	77	70	64
	Std. err.	219	518	275	189	144	117	100	86	77	69	62
n		295	295	295	295	295	295	295	295	295	295	295
Hits		34	25	40	44	56	55	59	50	58	79	65
% right		11.5254	8.4746	13.5593	14.9151	18.9831	18.6441	20	16.9492	19.661	26.7797	22.0339
Expected % right		4.252	3.7251	7.0109	7.0909	7.5956	9.5591	9.6601	9.4915	10.6288	15.0234	12.7711
Nongrocery shopping												
Size of choice set	Mean	370	666	357	248	191	157	134	117	104	94	86
	Std. err.	237	730	388	266	203	165	139	120	106	96	86
n		264	264	264	264	264	264	264	264	264	264	264
Hits		39	28	31	37	43	60	55	56	71	77	67
% right		14.7727	10.6061	11.7424	14.0152	16.2879	22.7273	20.8333	21.2121	26.8939	29.1667	25.3788
Expected % right		7.5086	4.0557	6.0932	6.2727	7.4288	11.2138	12.266	13.5998	13.214	17.2243	16.5323

alternatives in an individual's constrained choice set. Percentage right of the TAZ model is lower than most of the grid models, which can also be attributed to the size of choice set (i.e. an average of 338 alternatives in the TAZ constrained choice sets compared to an average of 73 alternatives in the 10 km^2 constrained choice sets).

Based on the number of correctly predicted zones, we further investigate how the size of grid impacts the model prediction by introducing a GIS-based measurement of goodness-of-fit. For each zoning system, zones that have been correctly predicted[6] by the models are retrieved. The travel time from the actual shopping location to the centroid of the predicted zone is then computed in TransCAD. Figure 3.8 shows that the centroid of a larger predicted zone (P10) can be closer to (Figure 3.12A) or farther (Figure 3.12B) from the actual shopping opportunity (S) than the centroid of a smaller predicted grid (P1).

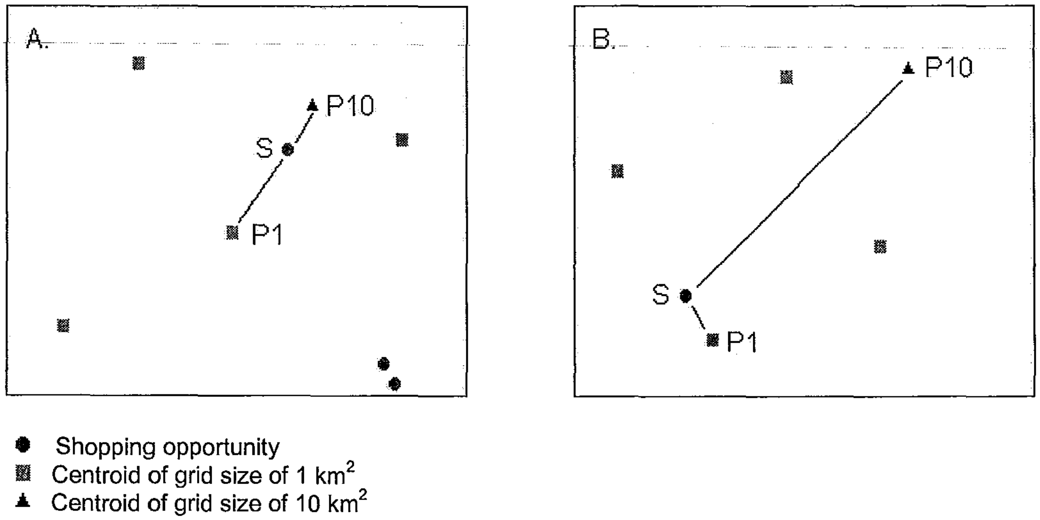


Figure 3.12: Point-by-point matching measurement of predictions.

[6] The correctly predicted zones refer to the predicted zones where actual shopping activities were conducted.

Table 3.11: Travel time from the actual shopping stores to the centroid of the predicted zone (in minutes)

	TAZ	1 km^2	2 km^2	3 km^2	4 km^2	5 km^2	6 km^2	7 km^2	8 km^2	9 km^2	10 km^2
General shopping											
MEAN	1.9918	0.7562	1.2255	1.0584	1.3870	1.7752	1.7676	1.0680	0.9906	1.2108	2.1768
STD	1.1799	0.7517	0.5774	0.6680	0.8269	0.9136	0.9840	0.4225	0.6179	0.5410	1.2616
MAX	4.12	2.93	3.12	2.32	4.69	4.34	4	2.41	2.96	2.51	6.49
MIN	0.12	0.07	0.16	0.06	0.11	0.23	0.36	0.22	0.2	0.22	0.2
Grocery shopping											
MEAN	1.6050	0.9040	0.9047	1.3080	1.2704	1.5362	1.7241	0.9240	0.8822	1.1358	2.0391
STD	0.7434	0.8921	0.5593	0.9081	0.6652	0.8622	0.9684	0.3556	0.5003	0.4790	1.1795
MAX	3.24	2.93	1.94	3.25	4.69	4.90	3.78	1.67	2.18	2.30	6.49
MIN	0.13	0.13	0.18	0.06	0.26	0.41	0.30	0.13	0.17	0.22	0.20
Nongrocery shopping											
MEAN	2.2659	0.8471	1.2074	0.9819	1.4274	1.9400	1.7931	1.0500	0.9582	1.2849	2.5007
STD	1.3020	0.6736	0.5606	0.6719	0.7598	0.8646	0.9252	0.3922	0.5638	0.5756	1.2606
MAX	4.12	2.93	2.08	2.09	2.70	4.34	3.71	2.31	2.24	2.51	5.24
MIN	0.13	0.30	0.18	0.06	0.11	0.91	0.60	0.11	0.20	0.02	1.13

Table 3.11 shows the statistics of the travel time between the actual location and predicted centroids. The higher the travel time, the farther is the predicted point from the true shopping location. TAZ exhibits an unsatisfactorily large travel time, which indicates that the predicted zone is not close to the actual location. The grid models, especially the 1 km^2 model, display an overall small travel time. Although it seems self-explanatory that the travel time from the predicted zone to the actual chosen location should be rising as the grid size increases, our results do not show a continuous trend, which is interrupted by the grids of 7 km^2, 8 km^2 and 9 km^2. The result implies that a monotonic trend can only exist in an area where opportunities are evenly spread and/or sample size is very large.

3.5 Summary

The 10 grid zoning systems, when imposed on the study area, produce different numbers of universal choice sets, from 625 to 5713. The finer the grids, the larger the number of destinations in the universal choice set, and the larger the number of destinations in the constrained choice sets. By using the same set of explanatory variables across different zoning systems, the scale effect has been examined in three shopping models. The results suggest that the magnitudes of the coefficients and the standard errors of the estimates become smaller as the grid size increases. Summary of t-statistics reveal the finer grid zoning systems, on average, resulted in coefficient estimates closer to the TAZ estimates.

Estimates of certain types of stores appear more sensitive to variations in zoning systems. The marginal effect of variables sensitive to variations in zoning levels decreases with aggregation. In order to determine if the differences between the grid model estimates are statistically different from the TAZ model, we report the differences by using the t-statistic that takes into account the standard errors of the estimates. Certain store estimates such as department stores are more consistent in varied zoning schemes. The temporal factor, activity duration, exhibits resistance to aggregation schemes whilst the store diversity index is sensitive to variations in aggregation level. A number of estimates from the grid models show opposite signs of the TAZ models.

The differences between the TAZ model estimates and grid model estimates are not necessarily unacceptable. The shape and the size of the TAZs are different from the grids. The TAZ boundaries are irregular and TAZs are different in size from one another, leading to large standard deviations in the distribution of the data. On the other hand, in each of the grid zoning systems, the shape and size of the grid are identical. Therefore, the gap between the estimates from grid systems and the TAZ estimate is not unexpected. However, the results do deepen our suspicion of the suitability of the predefined TAZ as an analysis unit for transportation-related studies.

According to the parameter estimations and model performance, different systems can be recommended in different contexts:

1. TAZ vs. grid zoning system: the preference would depend on whether individual data are available. A large proportion of transportation data are collected based on the TAZ system and it can only make sense if the data are analyzed and results are reported on the

same TAZ zoning system or at higher aggregated levels. However, if individual data are available, the TAZ system is not recommended for situations such as shopping. Our results show that the standard errors of the attractiveness parameters in the TAZ model are nearly always higher than the grid systems, which makes the reliability of the TAZ system questionable. Based on several measures of goodness-of-fit, grid systems also show their strength in data description and prediction. The ρ^2 in more than half of the grid systems is higher than that for the TAZ. Moreover, percentage right are higher and the point-by-point measure lower in almost all the grid systems than the TAZ. Compared to the grid system, TAZs have irregular shapes and different areas, making them less tractable. Another reason why TAZ is not a good analysis unit is that the individual cognitive realm is not coterminous with the TAZ boundary. It is felt more likely that when deciding destination zones, due to limited knowledge and personal preference an individual will split his/her constrained choice set into zones of manageable size which are presumably smaller than the TAZ (i.e., the size of a specific neighborhood is approximately three-block groups or one-fourth of a census tract, Price 2002).

2. Small grids vs. large grids: 1 km^2 grid is advantageous if we consider the GIS-based point-by-point measurement of the model performance. As the grid size decreases, the predicted locations are generally closer to the actual shopping opportunities, although not monotonically. From a different perspective, the10 km^2 grid would be a better choice in that the standard error of the estimates is lower than those from the smaller grids. Using the percentage right measurement of goodness-of-fit, larger grids also lead to higher correct prediction, although it is not in a monotonic fashion either. However, 10 km^2 grid

may be too large to serve as a shopping unit from an individual perceptive point of view. The cognitive map is not static, but is a dynamic entity (Kitchin 1994). It consists of different levels of details for specific tasks and the difference varies from person to person, from task to task. From the simplest to the more complex level, different details and integration of knowledge of an environment are retrieved to cope with decision making (Golledge and Timmermans 1990). When detailed information is needed, say the way finding of a specific teahouse, it is plausible that a smaller information analysis unit like the smaller grids would apply. Conversely, when looking at things at a large scale, say the location of a state's capital, a larger information analysis unit would be adopted. Shopping is an activity that requires detailed information such as the location, price and quality of the products, service, parking facilities etc. With the ever enriched information and knowledge, most individuals are believed to adopt smaller analysis units. Comparatively speaking, smaller grids have more flexibility in information aggregation. Hence among the grid zoning systems tested, the authors of this paper suggest a smaller grid (i.e., 1 km^2) would be more suitable than a larger grid (i.e., 10 km^2) in the context of shopping.

3.6 References

Amrhein C.G. 1995, Searching for the elusive aggregation effect: evidence from statistical simulations, *Environment and Planning A*, vol. 27, pp. 105-119.

Arentze T.A. and Timmermans H. 2005, An analysis of context and constraints-dependent shopping behaviour using qualitative decision principles, *Urban Studies*, vol. 42, pp. 435-448.

Barnard P.O. 1987, Modeling shopping destination choice behaviour using the basic multinomial logit model and some of its extensions, *Transportation Reviews*, vol. 7, pp. 17-51.

Beardwood J.E. and Kirby H.R. 1975, Zone definition and the gravity model: the separability, excludability and compressibility properties, *Transportation Research*, vol. 9, pp. 363-369.

Bell D.R., Ho T.-H., and Tang C.S 1998, Determining where to shop: fixed and variable costs of shopping, *Journal of Marketing Research*, vol. 35, pp. 352-369.

Ben-Akiva M. and Lerman S.R. 1985, *Discrete choice analysis: theory and application to travel demand,* The MIT Press, Cambridge.

Buliung R.N. 2004, *Data organization and the design of measures for exploration and analysis of household activity/travel behaviour,* Ph.D. Dissertation, McMaster University, Hamilton, ON.

Cliff A.D. and Haggett P. 1970, On the efficiency of alternative aggregations in region-building problems, *Environment and Planning*, vol. 2, pp. 285-294.

Coulton C.J., Korbin J., Chan T., and Su M. 2001, Mapping residents' perceptions of neighborhood boundaries: a methodological note, *American Journal of Community Psychology*, vol. 29, pp. 371-383.

Ding C. 1998, The GIS-based human-interactive TAZ design algorithm: examining the impacts of data aggregation on transportation-planning analysis, *Environment and Planning B: Planning and Design*, vol. 25, pp. 601-616.

Ettema D. and Timmermans H. 2004, Social exclusion and space-time accessibility under conditions of uncertain travel times: theory and numerical simulations, in *Annual Meeting of Association of American Geographers*, Philadelphia, March 14-19

Ewing R., Haliyur P., and Page G.W. 1994, Getting around a traditional city, a suburban planned unit development, and everything in between, *Transportation Research Record*, vol. 1466, pp. 53-62.

Fotheringham A.S., Brunsdon C., and Charlton M. 2000, Challenges in spatial data analysis, in *Quantitative Geography*, Sage Publications, London, pp. 236-248.

Fotheringham A.S. and Rogerson P.A. 1993, GIS and spatial analytical problems, *International Journal of Geographical Information Systems*, vol. 7, pp. 3-19.

Fotheringham A.S. and Trew S. 1993, Chain image and store choice modeling: the effects of income and race, *Environment and Planning A*, vol. 25, pp. 179-196.

Fotheringham A.S. and Wong D.W.S. 1991, The modifiable areal unit problem in multivariate statistical analysis, *Environment and Planning A*, vol. 23, pp. 1025-1044.

Fox E.J., Montgomery A.L., and Lodish L.M. 2004, Consumer shopping and spending across retail formats, *Journal of Business*, vol. 77, pp. 25-60.

Gehlke C.E. and Biehl K. 1934, Certain effects of grouping upon the size of the correlation coefficient in census tract material, *Journal of the American Statistical Association*, vol. 29, pp. 169-170.

Golledge R.G. and Timmermans H. 1990, Applications of behavioural research on spatial problems I: cognition, *Progress in Human Geography*, vol. 14, pp. 57-99.

Guo J.Y. and Bhat C.R. 2004, Modifiable Areal Units: A problem or a matter of perception in the context of residential location choice modeling?, *Transportation Research Record*, vol. 1898, pp. 138-147.

Guy C. 1985, The food and grocery shopping behaviour of disadvantaged consumers: some results from the Cardiff consumer panel, *Transactions of the Institute of British Geographers*, vol. 10, pp. 181-190.

Holt D., Steel D.G., and Tranmer M. 1996, Area homogeneity and the modifiable areal unit problem, *Geographical Systems*, vol. 3, pp. 181-200.

Horner M.W. and Murray A.T. 2002, Excess commuting and the modifiable areal unit problem, *Urban Studies*, vol. 39, pp. 131-139.

Kim B.-D. and Park K. 1997, Studying patterns of consumer's grocery shopping trip, *Journal of Retailing*, vol. 73, pp. 501-517.

Kitchin R.M. 1994, Cognitive maps: what are they and why study them?, *Journal of Environmental psychology*, vol. 14, pp. 1-19.

Landau U., Prashker J.N., and Alpern B. 1982, Evaluation of activity constrained choice sets to shopping destination choice modeling, *Transportation Research Part A*, vol. 16, pp. 199-207.

Lembo A.J. Jr, Lew M.Y., Laba M., and Baveye P. 2006, Use of spatial SQL to assess the practical significance of the Modifiable Areal Unit Problem, *Computers and Geosciences*, vol. 32, pp. 270-274.

Limanond T., Niemeier D.A., and Mokhtarian P.L. 2005, Specification of a tour-based neighborhood shopping model, *Transportation*, vol. 32, pp. 105-134.

Lloyd R. and Jennings D. 1978, Shopping behaviour and income: comparisons in an urban environment, *Economic Geography*, vol. 54, pp. 157-167.

McFadden D. 1974, Conditional logit analysis of qualitative choice behaviour, in *Frontiers in Econometrics*, Zarembka P., ed., Academic Press, New York, pp. 105-142.

Miller E.J. and O'Kelly M.E. 1983, Estimating shopping destination choice models from travel diary data, *The Professional Geographer*, vol. 35, pp. 440-449.

Miller H.J. 1999, Measuring space-time accessibility benefits within transportation networks: basic theory and computational methods, *Geographical Analysis*, vol. 31, pp. 187-212.

O'Neill W.A. 1991, Developing optimal traffic analysis zones using GIS, *Institute of Transportation Engineers Journal*, vol. 61, pp. 33-36.

Openshaw S. 1977a, Optimal zoning systems for spatial interaction models, *Environment and Planning A*, vol. 9, pp. 169-184.

Openshaw S. 1977b, A geographic solution to scale and aggregation problems in region-building, partitioning and spatial modeling, *Transactions of the Institute of British Geographers*, vol. 2, pp. 459-472.

Openshaw S. 1978, An empirical study of some zone-design criteria, *Environment and Planning A*, vol. 10, pp. 781-794.

Openshaw S. 1984, Ecological fallacies and the analysis of areal census data, *Environment and Planning A*, vol. 16, pp. 17-31.

Openshaw S. and Taylor P. 1979, A million or so correlation coefficients: three experiments on the modifiable areal unit problem, in *Statistical Applications in the Spatial Sciences*, Wrigley N., ed., Pion, London, pp. 127-144.

Openshaw S. and Taylor P. 1981, The modifiable areal unit problem, in *Quantitative Geography: A British View*, Wrigley N. and Bennett R., eds., Routledge, London, pp. 60-69

Páez A. and Scott D.M. 2004, Spatial statistics for urban analysis: a review of techniques with examples, *GeoJournal*, vol. 61, pp. 53-67.

Park C.W., Iyer E.S., and Smith D.C. 1989, The effects of situational factors on in-store grocery shopping behaviour: the role of store environment and time available for shopping, *Journal of Consumer Research*, vol. 15, pp. 422-433.

Pellegrini P.A., Fotheringham A.S., and Lin G. 1997, An empirical evaluation of parameter sensitivity to choice set definition in shopping destination choice models, *Regional Science*, vol. 76, pp. 257-284.

Plante M., Lowell K., Potvin F., Boots B., and Fortin M.-J. 2004, Studying deer habitat on Anticosti Island, Quebec: relating animal occurrences and forest map information, *Ecological Modeling*, vol. 174, pp. 387-399.

Price S. 2002, Surface interpolation of apartment rental data: can surfaces replace neighborhood mapping?, *Appraisal Journal*, vol. 70, pp. 260-273.

Putman S.H. and Chung S-H. 1989, Effects of spatial system design on spatial interaction models. 1: the spatial system definition problem, *Environment and Planning A*, vol. 21, pp. 27-46.

Ratcliffe J.H. 2005, Detecting spatial movement of intra-region crime patterns over time, *Journal of Quantitative Criminology*, vol. 21, pp. 103-123.

Robinson W.S. 1935, Ecological correlation and the behaviour of individuals, *American Sociological Review*, vol. 15, pp. 351-357.

Scott D.M. 2006, Constrained destination choice set generation: A comparison of GIS-based approaches, in *85th Annual Meeting of the Transportation Research Board: Compendium of Papers CD-ROM*, Washington D.C., January 9-13.

Timmermans H. 1981, Multiattribute shopping models and ridge regression analysis, *Environment and Planning A*, vol. 13, pp. 43-56.

Timmermans H. 1996, A stated choice model of sequential mode and destination choice behaviour for shopping trips, *Environment and Planning A*, vol. 28, pp. 173-184.

Timmermans H., van der Haijden R., and Westerveld H. 1984, Decision making between multiattribute choice alternatives: a model of spatial shopping behaviour using conjoint measurements, *Environment and Planning A*, vol. 16, pp. 377-387.

Tobler W. 1989, Frame independent spatial analysis, in *The Accuracy of Spatial Databases*, Goodchild M.F. and Gopal S., eds., Taylor & Francis, London, pp. 115-122.

Wong D.W.S. 2004, Comparing traditional spatial segregation measures: a spatial scale perspective, *Urban Geography*, vol. 25, pp. 66-82.

Zhang M. and Kukadia N. 2005, Metrics of urban form and the modifiable areal unit problem, *Transportation Research Record*, vol. 1902, pp. 71-79.

Chapter 4 Conclusions

Incorrect specification of destination choice sets (Williams and Ortuzar 1982; Landau et al. 1982; Thill 1992; Thill and Horowitz 1997) and changes in scale or unit definition (Openshaw 1984; Fotheringham and Wong 1991; Amrhein 1995; Plante et al. 2004; Ratcliffe 2005) can lead to biased parameter estimates and predictions as well as influence findings in statistical tests. The motivation of this thesis is to assess inaccuracy resulting from unconstrained destination choice sets and scale effects.

Starting with the features of the activity-based approach, Chapter 1 provided a brief review of data collection as well as various research areas and related model specifications in activity/travel behaviour studies. Given the deficiency of most current destination choice models, which is the lack of integration of spatio-temporal constraints in generating destination choice sets, the activity-based approach is proposed as a solution by taking into consideration both spatial and temporal constraints in the generation process. Applying a shortest path PPA algorithm (Scott 2006), constrained shopping destination choice sets are generated for both TAZ and grid systems.

The virtues of constrained over unconstrained destination choice sets are exhibited fully in Chapter 2 in the sense that the constrained models have greater efficiency in parameter estimations and higher predictive ability in terms of percentage right. Based on specific circumstances, comparisons between TAZ and grid systems as well as evaluation among different grid systems are presented in Chapter 3. While Chapter 2 shows the importance of taking constraints into account in the choice set generation, Chapter 3 shows that efficient parameter estimates and improvement in model goodness-of-fit rely not only on a well-specified choice set, but also on the underlying zoning system since MAUP effects occur as zoning schemes change.

4.1 Contributions to shopping behaviour research

Previous shopping destination choice studies have focused on grocery or general shopping trips. Nevertheless, the large number of nongrocery shopping opportunities in the study area and the considerable proportion of nongrocery shopping trips imply that the demand for nongrocery shopping is pronounced. This gives rise to concern that the conventional approach in shopping studies may not be appropriate. With adjoined SIC code to the shopping opportunities, classification of shopping stores becomes possible, which provides a detailed view into what types of shopping trips shoppers have conducted. Discrimination of 31 types of shopping trips in this thesis facilitates the separation between grocery and nongrocery shopping so that nongrocery shopping behaviour can be investigated through different shopping models from grocery shopping models. The breakdown of shopping opportunities to specific store types by SIC code provides a new disaggregated way to study how differently shoppers behave in response to various shopping stores.

Another contribution of this paper is the employment of a GIS-based space-time prism to retrieve choice sets constrained in both time and space. Suggested in previous studies that spatial shopping models used for analysis or prediction should be based on activity-constrained choice sets at an individual level (Arentze and Timmermans 2005), the activity approach in choice set generation differs from other approaches by taking account of the spatial distribution of shopping opportunities, realistic travel conditions and link-specific travel cost as well as an individual's time budget. By applying a network-based PPA algorithm and overcoming the main challenges in the substantial computational effort, the spatio-temporal effect has been carefully considered in the specification of destination choice set. Although mentioned in the literature that the individual deterministic choice set extracted from the space-time prism can be incorporated into a conventional choice model and its implementation is completely compatible with the standard discrete choice model and maximum likelihood estimation (Thill 1992), to the authors' knowledge, researchers have not gone that far. This thesis is the first attempt to incorporate the individual constrained choice set into a conventional choice model and compare the model estimations between a constrained and an unconstrained model. Despite the time and effort in data preparation, the virtues of constrained models revealed by this study will no doubt deepen our faith in constrained choice sets.

The contribution of the thesis also lies in the documentation of MAUP effects, in particular the scale effect, on shopping destination choice. Empirical studies have shown that analytical results obtained by techniques sensitive to scale effects or zoning effects are likely to change as the aggregation level or area boundary varies. It is pointed out that documenting the results on model estimations at different scales and zoning levels is important to assess the MAUP effects and critically evaluate the reliability of the estimates (Fotheringham and Rogerson 1993). Nonetheless, there are very limited empirical studies and documentation of the MAUP on logit models, let alone in the context of destination choice. In the thesis, we compare the parameter estimations and model goodness-of-fit between a TAZ system and 10 randomly generated grid systems. Sensitivity of explanatory variables exhibit remarkable, but tractable variations. Temporal factors are more resistant to variation in zoning schemes than attractiveness variables. Under a series of criteria, the best zoning system is recommended with certain conditions applied. Our results support the suspicion on the suitability of predefined analysis units like TAZ and suggest grid systems would be a potential substitution.

4.2 Future Research

Regarding the constrained shopping destination choice set, more constraints can be imposed on the universal choice set to define a more restrictive and realistic choice set. Cross-disciplinary collaboration between geography, sociology, economics and psychology needs to be consolidated. Particularly, the individual's perspective on shopping opportunities is relatively hard to measure, yet desirable. In this vein, expertise from cognitive science and psychology would be extremely helpful. Concerning the optimal underlying analysis unit, more empirical experiments with variations in zoning schemes and the documentation of outcomes will help assess the sensitivity of the analytical results, which will ultimately foster the development of other techniques in dealing with MAUP such as optimal zoning design and model specifications that can minimize the zoning and scale effects. We hope the inaccuracy due to destination choice set definition and zoning systems will gain sufficient attention and be carefully studied in the foreseeable future.

4.3 References

Amrhein C.G. 1995, Searching for the elusive aggregation effect: evidence from statistical simulations, *Environment and Planning A*, vol. 27, pp. 105-119.

Arentze T.A. and Timmermans H. 2005, An analysis of context and constraints-dependent shopping behaviour using qualitative decision principles, *Urban Studies*, vol. 42, pp. 435-448.

Fotheringham A.S. and Rogerson P.A. 1993, GIS and spatial analytical problems, *International Journal of Geographical Information Systems*, vol. 7, pp. 3-19.

Fotheringham A.S. and Wong D.W.S. 1991, The modifiable areal unit problem in multivariate statistical analysis, *Environment and Planning A*, vol. 23, pp. 1025-1044.

Landau U., Prashker J.N., and Alpern B. 1982, Evaluation of activity constrained choice sets to shopping destination choice modeling, *Transportation Research Part A*, vol. 16, pp. 199-207.

Openshaw S. 1984, Ecological fallacies and the analysis of areal census data, *Environment and Planning A*, vol. 16, pp. 17-31.

Plante M., Lowell K., Potvin F., Boots B., and Fortin M.-J. 2004, Studying deer habitat on Anticosti Island, Quebec: relating animal occurrences and forest map information, *Ecological Modeling*, vol. 174, pp. 387-399.

Ratcliffe J.H. 2005, Detecting spatial movement of intra-region crime patterns over time, *Journal of Quantitative Criminology*, vol. 21, pp. 103-123.

Scott D.M. 2006, Constrained destination choice set generation: A comparison of GIS-based approaches, in *85th Annual Meeting of the Transportation Research Board: Compendium of Papers CD-ROM*, Washington D.C., January 9-13.

Thill J.C. 1992, Choice set formation for destination choice modeling, *Progress in Human Geography*, vol. 16, pp. 361-382.

Thill J.C. and Horowitz J.L. 1997, Travel-time constraints on destination choice sets, *Geographical Analysis*, vol. 29, pp. 108-123.

Williams H.C.W.L. and Ortúzar J. de D. 1982, Behavioural theories of dispersion and the misspecification of travel demand models, *Transportation Research Part B*, vol. 16, pp. 167-219.