Catalysts and Magnets: Built Environment Effects on Bicycle Commuting

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Jessica Elizabeth Schoner

IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

David M. Levinson, Adviser
Xinyu Cao, Co-Adviser

September 2013

Acknowledgements

I would like to thank the many people who have helped me through the completion of this thesis. The first of these are my advisors and committee members. It is an honor to have worked with Dr. David Levinson, Dr. Jason Cao, and Dr. Gary Davis. Each has had a strong and significant impact on my education trajectory.

Dr. Levinson's *Transportation Geography and Network Science* course was the first engineering class I had taken in almost a decade, and after completing the course, he strongly encouraged me to add the Civil Engineering dual degree to my program of study. I am grateful for his countless hours of education and advising that have enabled me to pursue this dream.

This thesis would not have been possible without the support of my Urban Planning adviser, Dr. Jason Cao. I worked as a research assistant with Dr. Cao on his study *The Transportation Impacts of Transitways: A Case Study of the Hiawatha Line*. I owe my deepest gratitude to Dr. Cao for the use of his *Transitways* study data for my thesis, and for his mentorship over the past three years. I would also like to thank the Center for Transportation Studies Transitway Impacts Research Program and its sponsors and partners.

It is a pleasure for me to thank the many other people who have supported my thesis work in countless ways: my labmate in the Nexus Research Group, Andrew Owen, who provided technical assistance in constructing the bicycle accessibility measure used in this thesis; Humphrey School faculty and staff, especially Dr. Greg Lindsey and Katherine Murphy; professional mentor Tony Hull; and the brilliant and insightful classmates I have worked with over the past three years.

Finally, I would like to thank my family and friends for their encouragement through decades of school; and Paul, for his immeasurable companionship, patience, and support.

Abstract

What effects do bicycle infrastructure and the built environment have on people's decisions to commute by bicycle? While many studies have considered this question, commonly employed methodologies fail to address the unique statistical challenge of modeling such a low mode share. Additionally, self selection effects that are not adequately accounted for may lead to overestimation of built environment impacts.

This study addresses these two key issues by using a zero-inflated negative binomial model to jointly estimate participation in and frequency of commuting by bicycle, controlling for demographics, residential preferences, and travel attitudes. The findings suggest a strong self selection effect and modest contributions of bicycle accessibility: that bicycle lanes act as "magnets" to attract bicyclists to a neighborhood, rather than being the "catalyst" that encourages non-bikers to shift modes. The results have implications for planners and policymakers attempting to increase bicycling mode share via the strategic infrastructure development.

Contents

List of Figures

List of Tables

Acronyms

ACS American Community Survey. 2, 7, 8, 15, 22, 24

AM1 Alternate Model 1. 38–41, 43

AM2 Alternate Model 2. 39–41

AM3 Alternate Model 3. 39–41

AM4 Alternate Model 4. 40, 41

FHWA Federal Highway Administration. 1

FT Full-time. 14, 16

GEE Generalized Estimating Equation. 14, 16

GIS Geographic Information System. 13, 27

GMM Generalized Mixed Model. 13, 16

HEV Heteroscedastic Extreme Value Model. 14, 16

IIA Independence of Irrelevant Alternatives. 14

LEHD Longitudinal Employer-Household Dynamics. 28

LR Test Likelihood Ratio Test. 33

LRT Light Rail Transit. 19, 21

MLE Maximum Likelihood Estimation. 30, 49

MNL Multinomial Logistic Regression. 2, 9, 12, 14, 16, 38

NBREG Negative Binomial Regression. 33

NTPP Nonmotorized Transportation Pilot Program. 1

OLogit Ordered Logistic Regression. 13, 16

OLS Ordinary Least Squares Regression. 11, 16, 33

OSM OpenStreetMap. 28

PT Part-time. 14, 16

RCA Random Coefficient Analysis. 14, 16

RELogit Rare Events Logistic Regression. 15, 16

ROW right-of-way. 21

RP Revealed Preference. 9

SP Stated Preference. 9, 13

SURE Seemingly Unrelated Regression Equations. 15, 16

TAZ Transportation Analysis Zone. 15

ZINB Zero-Inflated Negative Binomial Regression. 2, 32, 40, 41, 43, 45

Chapter 1

Introduction

The question of a relationship between bicycling and the built environment, particularly dedicated bicycle lanes and trails, has captivated the attention of researchers and planners for decades. In a state of the practice and research needs paper, Porter *et al.* (1999) identified critical questions about the role of bicycle infrastructure: how to forecast use of new facilities, how to estimate mode shift due to building new facilities, and how these new facilities may affect mobility, congestion, and air quality.

Despite many advances in the field, their questions about the impacts of infrastructure are still salient today. Differentiating between self-selection effects and true causal relationships continues to challenge researchers. Cao *et al.* (2009a) found that despite ongoing attempts to control for this residential self-selection effect in studies about built environment impacts on travel, not many studies can indicate just how big the built environment contribution is after controlling for self-selection.

Even with this uncertainty, cities are scrambling to build new bike infrastructure in attempt to encourage mode shift to bicycling. The Federal Highway Administration (FHWA) recently concluded Nonmotorized Transportation Pilot Program (NTPP), a $100M experiment to evaluate the extent to which investments in infrastructure, education, and enforcement can increase rates of walking and biking (Federal Highway Administration, 2012).

However, common strategies for researching and evaluating transportation projects fail to address the nuances of bicycling. The utility of bicycling, more so than any other mode, is strongly affected by weather phenomena and day-to-day variation in travel needs, such as hauling cargo or goods. As a consequence, many bicyclists are in fact multi-modal travelers (Heinen *et al.* , 2010). Because bicycling has such a small mode share, standard survey and data collection strategies, especially those that assume people tend to stick to a single mode throughout the week such as the American Community Survey (ACS), underestimate its prominence.

Furthermore, much of the research about bicycling uses discrete Binary Logit or Multinomial Logistic Regression (MNL) mode choice models to predict participation in bicycling. Many also employ research design strategies that skew the sample in favor of people who are already prone to bicycling, producing coefficients that are not accurate for modeling behavior among the general population. Because many bicyclists are multimodal, distinguishing between participation and frequency is critical for being able to model impacts of bicycling on the transportation system, environmental issues, and health (Heinen *et al.* , 2010).

What effect do bicycle infrastructure and the built environment have on people's decisions to commute by bicycle, and are some people more inclined to be "bikers" than others? This study explores frameworks developed for studying residential self-selection to account for this dichotomy between participation ("modal" self-selection) and frequency of bicycle commuting, with aims of expanding the understanding of bicycling and the built environment. Existing survey data from Minneapolis, MN and a Zero-Inflated Negative Binomial Regression (ZINB) model are employed to jointly estimate participation in and frequency of bicycle commuting as a function of the built environment, controlling for demographics, residential preferences, and travel attitudes.

This research is significant in two primary ways: First, at the time of writing, the application of a zero-inflated model (functionally, a sample selection model with count data) to control for self-selection in nonmotorized travel participation is unique (Mokhtarian & Cao, 2008). For comparison, this study estimates several additional models of bicycling participation and frequency using the same data and variables and following common strategies employed in the literature. The results of comparing these and the zero-inflated model show that in particular, modeling behavior based on a subset of the population known to be bicyclists substantially overestimates bicycling when applied to the population at large.

Second, while the magnitude and direction of the coefficients are consistent with other studies and the comparative models of modeling bicycling frequency in the general population, the unique structure of the zero-inflated model provide deeper insight to the relationships between individual preferences, self-selection, and the built environment. When interpreted in this framework, it is easy to identify ways of harnessing the residential self-selection effect to increase rates of bicycling.

The extent to which bicycling infrastructure acts as a "catalyst" to induce mode shift among non-bicyclists to biking is unknown, given the difficulty of establishing causality in cross-sectional studies (Cao *et al.* , 2009a). However, the evidence of a self-selection effect suggests that certain infrastructure types function as "magnets" for people who are already prone to bicycling for work, due to their demographic, residential preference, and travel attitude profiles. Combined with evidence from variables used to predict frequency after controlling for residential self-selection, the findings from this study can be used to locate new bicycling infrastructure strategically for providing housing choices for current and would-be bicyclists, and maximizing the number of bicycle trips they choose to make.

This thesis is organized as follows: Chapter 2 reviews literature about bicycling and the built environment, with a specific focus on how studies manage low numbers of bicyclists among the general population, and briefly discusses literature on statistical techniques that can be used to address this issue. Chapter 3 describes the survey administration, data, and modeling procedure. Chapter 4 presents findings from using a zero-inflated negative binomial model to predict participation in and frequency of bicycle commuting among urban residents. Chapter 5 presents results from this model side-by-side with those from other common methods of estimating bicycling behavior and tests how accurately each model predicts the number of bicyclists and non-bicyclists in the data from which they were estimated. Finally, Chapter 6 concludes with a discussion of the implications of this research for urban planners, transportation engineers, policymakers, and researchers.

Chapter 2

Literature Review

This chapter briefly reviews literature on self-selection and the unique characteristics of bicyclists that make them difficult to model. It then presents strategies that have been employed to address the low mode share of bicyclists in research about bicycling and the built environment and review findings from these efforts. Finally, literature on methodological techniques that have been employed on parallel transportation research questions provides a framework for further analysis.

2.1 Self-Selection and Traveler Preference

Residential self-selection issues are pervasive in research about "alternate" modes of transportation. Any clear and causal relationship between travel behavior and the built environment is masked by this phenomenon of people self-selecting into neighborhoods that meet their needs and lifestyle. The built environment characteristics around a person's home, therefore, reflect that resident's preferred mode at least as much as they cause the resident to actually change behavior. If residential self-selection has a strong effect, then an observed relationship between travel behavior and the built environment may in fact be a spurious correlation.

In their two prominent papers on the residential self-selection effect, Cao *et al.* (2009a) found that residential preferences have a strong impact on both location choice and travel behavior, but the built environment has a separate contribution above and beyond the self-selection effect. Measuring and controlling for self-selection, therefore, is essential for understanding the true impacts of built environment and infrastructure characteristics on behavior. For nonmotorized travel in particular, this self-selection effect is characterized by correlation between a preference for biking or walking and choosing to live in a neighborhood with supportive built environment and infrastructure features like bike lanes, multipurpose trails, and good street network connectivity.

Mokhtarian & Cao (2008) identified seven methods that have applicability to studying residential self-selection in travel: (1) direct questioning, (2) statistical control, (3) instrumental variables models, (4) sample selection models, (5) joint discrete choice models, (6) structural equations models, and (7) longitudinal designs. The authors identified examples of six of these methods being used in practice to control for self-selection; at the time, however, they could not identify any uses of the sample selection model to control for self-selection. Since then, Cao (2009) demonstrated the use of sample selection modeling for predicting participation in and frequency of stops in trip-chaining behavior on the evening commute. However, applications of this method to nonmotorized travel is limited.

The self-selection literature focuses largely on *residential* self-selection because this is the confounding factor between built environment characteristics and travel behavior. However, travel attitudes and preferences clearly also have a direct impact on behavior beyond operating through residential self-selection. Given the limitations and challenges that bicycling presents for many common transportation needs, many people simply would never consider bicycling (Wardman *et al.* , 2007).

In their controversial essay "Bicycling in the United States: A Fringe Mode?", Gordon & Richardson (1998) assert that cities demonstrating higher rates of bicycling, notably Seattle at the time the paper was written, can likely be attributed to personal preference more so than any policy or built environment characteristic. Additionally, many of the same built environment features that support bicycling also appear in studies highlighting transit-friendly or pedestrian-friendly characteristics (Ewing & Cervero, 2010), so a single neighborhood may contain people self-selecting to satisfy on a wide range of travel needs. Therefore, with smaller modes like bicycling, the strong influence of individual preference on choosing the non-dominant mode may be easier to identify and control for than residential self-selection.

2.1.1 Multimodal Lifestyle

Complicating the study of bicycling is the fact that bicyclists, more so than other mode users, are distinctly "multi-modal" (Heinen *et al.* , 2010). More so than driving and even transit, bicyclists are vulnerable to day-to-day changes in weather or varying travel needs. Having to make additional stops, carry groceries or other bulky items, or travel when it is dark all decrease the utility of bicycling. Many bicyclists, therefore, can be thought of as "part-time" bicyclists.

Unfortunately, this phenomenon results in conventional survey questions underestimating bicycling. Surveys that ask about a single primary commute mode, such as the ACS, miss people who bike only 1-2 days per week, or only for non-work purposes. These questions tell us how many people are bicycling frequently, but do not tell us how many people are biking infrequently and how many trips this translates to.

Surveys that ask what mode you used "yesterday", such as the Minneapolis-St. Paul metro area Travel Behavior Inventory survey (Metropolitan Council, n.d.), in theory should average out over the whole population to a representative value of the amount of bicycling being done, but this assumes bicyclists choose their biking days randomly and that sample

Table 2.1: Bicycle Commute Mode Share from the 2011 5-year ACS Estimates

	Bicyclists		All Modes		
Geography	**Commuters**	**(SE)**	**Commuters**	**(SE)**	**Mode Share**
United States	744,560	(4030)	139,488,206	(80701)	0.53 %
Metro Areas	661,740	(3824)	118,354,454	(61120)	0.56 %
Principal Cities	437,865	(3065)	45,914,919	(24078)	0.95 %
Minnesota	19,706	(565)	2,684,619	(3149)	0.73 %
Minneapolis Metro	13,885	(450)	1,690,532	(2432)	0.82 %
Minneapolis City	7,808	(355)	202,386	(1327)	3.86 %

sizes are large enough to reflect the ground truth of bicycling. With small sample sizes and such a low mode share, these types of questions have low chances of catching a part-time bicyclist on their biking days.

Much of the literature on bicycling employs binary logit models that predict who is a bicyclist in any capacity, and do not tell us how much bicycling is actually being done. Heinen *et al.* (2010) describes this problem in this way: "It is of interest to distinguish between (1) mode choice in general, that is to say, the bicycle is at least one of the modes used; and (2) daily choice, in terms of frequency. The latter is useful because many bicycle commuters choose not to cycle every day."

2.2 Bicycling as a Small Mode Share

Bicycling represents a relatively small mode share, particularly for commuting. In the United States, the ACS estimates that only 0.51 % of commuters use a bicycle as their primary commuting mode. While the average is higher when focusing on central cities, as shown in Table 2.1, the overall rates are still extremely low relative to driving, and even other so-called "alternate" modes such as transit.

A review of the literature on bicycling behavior and the built environment found three distinct strategies for modeling bicycling, given the low mode share. A few studies employed multiple strategies in the same paper. Many studies did not directly address the low mode share concern. The resulting five categories, explained more fully in subsequent sections, are:

1. Inclusion Criteria (2.2.1)
2. Strategic Over-sampling (2.2.2)
3. Combined Inclusion Criteria and Strategic Over-sampling (2.2.3)
4. Statistical Distributions (2.2.4)
5. No specific technique for small mode share (2.2.5)

Table 2.2 summarizes the studies reviewed in each of these five categories. Some studies appear multiple times in the table because the paper includes several components employing one or more of the three techniques.

2.2.1 Inclusion Criteria

Inclusion criteria studies apply a filter to general population data to extract a subset sample that applies to their research question. Data sources typically come from original general population surveys or government-administered regional travel surveys and census. Many, but not all, of the studies select people based on past bicycling behavior (e.g., having bicycled within the past year) or expressed willingness to bicycle. The effect is that these studies model bicycling behavior among a subset of people already expected to have some propensity to bicycle. The results may not be easily extrapolated to the general population.

The clearest example of this technique comes from Wardman *et al.* (2007). The authors combine national census travel diary records with originally collected Revealed Preference (RP) and Stated Preference (SP) survey data to jointly estimate a MNL model.

They state that about 60 % of people both in general and among their SP survey participants indicate that they would "never contemplate switching to cycling", so these participants were removed from the dataset. The authors explain, "A model is hardly needed to predict the behavior of such individuals and their actual choices or SP responses would provide little information for modelling purposes." The authors also filter out trip records from individuals with commutes greater than 7.5 miles (about 12 km) because they were only interested in commute trips where bicycling is a viable option. Consequently, they omit commute distance as an explanatory variable in their model because this threshold eliminates most of the variation. They also anticipate circularity between mode choice and commute distance due to residential self-selection effects.

Wardman et al. found that dedicated, separate infrastructure was positively and strongly associated with bicycle commuting. Time spent on dedicated bicycle facilities was valued as 14 % to 17 % as much as time spent on streets with no facility. They felt that their estimates were too high: for example, upgrading a 20-minute trip from no facility to a segregated cycleway according to their data would be equivalent to reducing the travel time by 17 minutes, for a 3-minute long commute. But even after standardizing the coefficients, time on facilities was still valued at 28 % to 31 % of time on streets with no facility. In this case, the results may be exaggerated by the removal of respondents who would not bicycle under any circumstance because their value of time by mode would be heavily skewed to favor time in the car.

Winters *et al.* (2010), Handy & Xing (2011), and Xing *et al.* (2010) all used a similar technique for screening out dedicated non-cyclists. Winters *et al.* (2010) considered anyone with access to a bicycle and who has cycled within the past year a "current cyclist", and anyone who indicates willingness to cycle in the future as a "potential cyclist". Handy & Xing (2011) and Xing *et al.* (2010) used a stricter definition, only including people who had bicycled within the past year. While there is no direct comparison, it is highly probable that these techniques would screen out many of the same individuals who were removed in Wardman *et al.* (2007) for indicating that they would never commute by bicycle.

After screening participants, Wardman et al. collected trip data on frequently made non-recreational trips by bicycle and by another mode for comparison. For each trip, they measured built environment characteristics around the origin and destination, and along the route in between. They used three multilevel logistic modeling to predict the probability that any given trip occurred by bicycle as a function of (1) origin characteristics, (2) destination characteristics, and (3) characteristics along the route. They found bicycle facilities to be positively associated with the odds that a given trip was completed by bike in the origin and destination model, but not the route model, among other built environment effects and after controlling for demographic factors.

Xing *et al.* (2010)'s and Handy & Xing (2011)'s studies use the same data about six small cities in California and Oregon and "past year" inclusion criterion to model use of bicycling for utilitarian versus recreational trips, the quantity of each type of bicycling, and a specific focus on commuting by bicycle.

Xing *et al.* (2010)'s binary logit model of utilitarian bicycling found that physical environment measures, including shorter distances and more safe destinations, were associated with both a greater share of biking for transportation. Their Ordinary Least Squares Regression (OLS) found the same factors to be associated with longer distances biking for transport. Both of these findings are significant after controlling for individual travel attitude responses, though the magnitude of these effects is small relative to attitudinal factors. While one would expect longer trips to correlate with more mileage biking, long trips are a barrier to making the trip via bicycle in the first place, so the increased distance is offset by reduced probability of making the trip and (presumably) frequency. This finding underscores the importance of understanding both whether people bicycle and how much (duration and/or frequency) they bicycle.

Commute distance and parking costs are associated with odds of commuting by bicycle, but like the utilitarian versus recreational finding, the magnitude of these coefficients is an order of magnitude smaller than the travel attitude variable about personal comfort level while bicycling (Handy & Xing, 2011). The effects of restricting the sample to bicyclists only means that these coefficients may overstate the impacts of built environment characteristics relative to the general population in which the majority would not consider bicycling regardless of infrastructure. Additionally, their study only reported a singular commute mode, the one the respondent "usually" uses, so the model may exclude part-time cyclists.

2.2.2 Strategic Over-sampling

A common strategy for bicycle research and evaluation is to structure the sampling frame in a way that deliberately captures a greater than average proportion of cyclists. Over-sampling strategies range from subtle, such as pre-selecting geographies expected to have higher than average rates of bicycling (Akar & Clifton, 2009; Moudon *et al.* , 2005), to deliberate, such as snowball sampling bicycle clubs and local bike shops Sener *et al.* (2009) or bicyclist intercept surveys (Hunt & Abraham, 2007).

Like the inclusion criteria studies, over-sampling runs the risk of measuring effects on a concentrated population of people already prone to bicycling. Additionally, the outreach method for contacting bicyclists heavily biases the types of bicyclists who respond. Specifically, bike club members are more likely to fit the "fearless" category in Geller (2006)'s framework. These cyclists may be less affected by built environment characteristics than occasional cyclists because they are comfortable bicycling in mixed traffic. For modeling the effects of dedicated bike infrastructure on mode shift or bicycling behavior among infrequent cyclists, this strategy may not be appropriate.

Akar & Clifton (2009) and Heinen *et al.* (2011b) both sampled specific geographies to model bicycle commute mode choice. In Akar and Clifton's case, the sample contained students, faculty, and staff affiliated with a university campus. Since colleges and universities are significant predictors in bicycle commute mode share (Rodríguez, 2004), this sampling frame can be expected to have higher than average rates of bicycling. Heinen *et al.* (2011b) selected several large employers in cities in the Netherlands that are known for their especially high bicycling mode share in order to capture more employees who commute by bicycle. While Akar & Clifton (2009) constructed a MNL model for all modes and Heinen *et al.* (2011b) used a simple binary logit model of bicycling or not bicycling, they are both effectively modeling participation in bicycling. It should be noted that Heinen *et al.* (2011b) also appears in section 2.2.3 because additional analysis in the paper uses inclusion criteria.

Two other studies, Moudon *et al.* (2005) and Sener *et al.* (2009), model frequency of bike commuting, though their oversampling strategies and model structures differ substantially. Moudon *et al.* (2005) pre-selects geographies using Geographic Information System (GIS) that are expected to be conducive to bicycling. They then administered a general population telephone survey via random digit dialing within these geographies. The geography selection component facilitates capturing a higher number of bicyclists than average, while the administration ensures that within those geographies, the sample is relatively representative. The model is a binary logit regression of commuting by bicycle at least weekly. While the logit structure is typically associated with modeling participation, the weekly threshold is high relative to other studies reviewed for this thesis and seems to indicate frequency as much as participation. Where Moudon et al.'s sampling strategy was subtle, Sener *et al.* (2009) took the opposite approach. The researchers administered the survey using a snowballing technique, sending an online link to bicycling clubs, posting the link in local bike shops, and purchasing ads in local papers. The effort captured bicyclists in over 100 cities in Texas (USA), but clearly is not a representative sample of the general population. This study used an Ordered Logistic Regression (OLogit) model to predict frequency of commuting by bicycle.

One final strategy for oversampling is the intercept survey. Survey administrators reach out specifically to bicyclists while they are biking, either by stopping them along a trail or corridor, or by attaching a paper survey to parked bicycles. Hunt & Abraham (2007), Hunt (2009), and Thakuriah *et al.* (2012) all make use of this strategy. Hunt & Abraham (2007) and Hunt (2009) are similarly structured, but the latter includes a longitudinal comparison between two administrations of the same survey. The SP experiment asks bicyclists to choose between pairs of scenarios with varying trip length and infrastructure quality. Thakuriah *et al.* (2012)'s study is distinct in that it specifically focuses on suburban bicycle commuters using new bicycle facilities. Their model attempts to identify mode shift by using Binary Generalized Mixed Model (GMM) to model whether the survey respondent is a former driver who never bicycled.

2.2.3 Combined Inclusion Criteria and Strategic Sampling

Some studies made use of both of the aforementioned strategies. They first administered a survey using strategic sampling, and then screened their participants using inclusion criteria to focus on a subset of bicyclists. Heinen *et al.* (2011b) and Heinen *et al.* (2011a) sample employees in cities with high rates of bicycling, as described above. However, in some of their analyses, they also filter out non-cyclists. Heinen *et al.* (2011b) constructs one binary logit model of whether the participant bike commutes Full-time (FT) or Part-time (PT). For this model, the authors had to remove all non-cyclists. Despite the binary structure, this is effectively a model of frequency, where FT represents bicycling daily, and PT represents bicycling with any frequency greater than once per year and less than daily. Heinen *et al.* (2011a) followed up with PT cyclists from the original study periodically over the course of a year to survey them about how they commuted on that particular day. This set of models (Binary Logit, Generalized Estimating Equation (GEE), and Random Coefficient Analysis (RCA)) attempt to explain day-to-day factors that affect a

PT cyclist's choice of mode. From this study, the authors conclude that bicyclists are distinctly multi-modal.

Like Akar & Clifton (2009), Rodríguez (2004) surveys affiliates of a university campus to oversample bicyclists. They further filter their results by selecting only respondents within certain municipal boundaries that are considered "close enough" to be bicycling distance. They start with a classic MNL mode choice model, and additionally employ a Nested Logit and Heteroscedastic Extreme Value Model (HEV) model to loosen the Independence of Irrelevant Alternatives (IIA) assumption that tends to break down among "alternate" modes.

2.2.4 Statistical Techniques

While relatively uncommon among the literature on bicycling and the built environment, statistical techniques can be used to address the relatively low mode share for bicycling among general population surveys. Buehler (2012) models bicycling for any given commute trip in a large regional travel diary survey using Rare Events Logistic Regression (RELogit). For these individual commute trips, the authors found strong associations with bicycle facilities provided at work, including bike parking. Free car parking was negatively associated with making the commute trip by bicycle.

2.2.5 No Specific Low Mode Share Technique

Several studies model bicycling behavior, controlling for residential self-selection, without using any inclusion, sampling, or statistical technique designed to manage the low mode share issue. These range from general population studies in cities that coincidentally have a high bicycling mode share (versus Heinen *et al.* (2011b) and Heinen *et al.* (2011a) that specifically sought high mode share cities in their research design) to mode choice models that jointly estimate all modes at once so the excess zeroes for bicycling aren't as significant of an issue.

Cao *et al.* (2009b) used Seemingly Unrelated Regression Equations (SURE) to model nonwork travel mode choice and found that after controlling for self-selection, built environment still had some effect on nonmotorized travel. Both attitudes and the built environment had a stronger effect on nonmotorized travel than auto and transit, presumably because driving in particular is the default mode for most people. Auto ownership disappeared from the model after controlling for self-selection.

Krizek & Johnson (2006) and Parkin *et al.* (2007), and Dill & Voros (2007) all explore bicycling using general population data. Parkin *et al.* (2007) in particular uses national Census data in England and Wales to model bicycling mode share within each district. While their paper made no explicit mention of it, a parallel concern using US Census or ACS data is the standard errors relative to small mode share values. Krizek & Johnson (2006) uses census data at the Transportation Analysis Zone (TAZ) level to compare mode share over time close to and far from new bicycle facilities.

2.3 Methodology Review

While use of specific statistical techniques for modeling low bicycle mode share is infrequent, examples from related fields suggest that these techniques are an opportunity for bicycling research.

Handy *et al.* (2006) and Schoner & Cao (2013) use Negative Binomial regression to estimate frequency of utilitarian and recreational walking. The Negative Binomial model relaxes the Poisson distribution's strong assumption of variance equal to the mean by adding a separate dispersion parameter. Large numbers of people who do not bike appear in a dataset as excess "zeroes", which in effect appears as overdispersion.

Table 2.2: Research Design Techniques for Managing Low Mode Share in Selected Studies of Bicycling and the Built Environment

Citation	Data Source	Technique	Model
Inclusion Criteria to Select Bicycling Subset			
Handy & Xing (2011)	Original Survey	Biked within past year	Logit - Primary bike commute
Wardman *et al.* (2007)	Census & Survey	Current/Potential Bicyclists	MNL - Mode Choice
Winters *et al.* (2010)	Original Survey	Current/Potential Bicyclists	Multilevel Logistic - Bike (vs. car) trip
Xing *et al.* (2010)	Original Survey	Biked within past year	Logit - Utilitarian v. Recreation Biking
Xing *et al.* (2010)	Original Survey	Biked within past year	OLS - Log-miles of Utilitarian Bike
Strategic Survey to Oversample Bicyclists			
Akar & Clifton (2009)	Original Survey	University Affiliates	MNL - Mode Choice
Heinen *et al.* (2011b)	Original Survey	High biking cities	Logit - Bike commute
Hunt & Abraham (2007)	Original Survey	Bicyclists	Logit - SP experiment
Moudon *et al.* (2005)	Original Survey	Suitable geography	Logit - Biking at least weekly
Sener *et al.* (2009)	Original Survey	Bicyclists	OLogit - Bike commute frequency
Thakuriah *et al.* (2012)	Original Survey	Bicyclists	Binary GMM - Former captive car user
Inclusion Criteria & Strategic Survey			
Heinen *et al.* (2011b)	Original Survey	High biking cities & Cyclists	Logit - FT vs. PT Bike Commute
Heinen *et al.* (2011a)	Original Survey	High biking cities & PT Cyclists	GEE/RCA Logit - Mode Choice
Rodríguez (2004)	Original Survey	City & University Campus	MNL, Nested, & HEV - Mode Choice
Statistical Techniques			
Buehler (2012)	Regional Survey	RELogit	RELogit - Bike commute
No Specific Technique Used - General Population			
Cao *et al.* (2009b)	Original Survey		SURE - Bike/Walk Frequency
Krizek & Johnson (2006)	Regional Survey		Logit - Bike trip(s) in travel diary
Parkin *et al.* (2007)	Census		Logit - Bike commute share

Several of the methodologies identified by Mokhtarian & Cao (2008)'s for controlling residential self-selection may also be appropriate for this low mode share situation. Specifically, sample selection and joint discrete choice models provide opportunities to predict participation in bicycling separately from frequency, which accounts for the high numbers of people who simply do not bike at all. Pinjari *et al.* (2008) employed this technique by jointly modeled residential location choice and bicycle ownership to account for self-selection. While this is not directly commute related, it demonstrates the possibility of using a joint process to predict a binary discrete value (neighborhood type) and a discrete frequency value (number of bicycles owned) together. Neighborhood type in this example could represent self-selection into a bicycle friendly neighborhood. However, given bicycling's low representation among commute modes and strong correlation between bicycle suitability and other transportation preferences such as transit service level or walkability, self-selecting into a bicycle-friendly neighborhood may not be sufficient to identify bicyclists among all other residents.

Sample selection models use a logit or probit function to predict participation separately from frequency. Cao *et al.* (2008) employ the Heckman Selection method to predict participation in trip-chaining behavior on the evening commute and, given participation, the number of stops comprising that trip chain.

Zero-inflated models function similarly to the Heckman Selection model, but they use a logit model to predict non-participation (excess zeroes) and either a Poisson or Negative Binomial model to predict frequency. While a traditional Negative Binomial model would treat excess zeroes simply as overdispersion, the zero-inflated model assumes a separate process generates the extra zeroes so it can be modeled using different parameters. This type of model is used commonly for modeling infrequent events such as traffic crashes (Chin & Quddus, 2003).

Chapter 3

Methodology

3.1 Survey Administration

The data for this thesis came from a self-administered ten-page survey mailed in May 2011 to households in five corridors in the Twin Cities as part of a study on the effects of Light Rail Transit (LRT) and associated built environment on travel behavior (Cao & Schoner, 2013). These corridors, shown in Figure 3.1, were selected with the help of local planners because they had similar demographic trends. Three of the corridors are located in South Minneapolis: Nicollet Avenue, Bloomington Avenue, and Hiawatha Avenue from Lake Street to 50^{th} Street. The two remaining corridors were in suburban communities outside the City of Minneapolis: Coon Rapids, 12 miles north of downtown Minneapolis, and Bloomington, 17 miles south of downtown.

For each corridor, we purchased two databases of residents from AccuData Integrated Marketing, a commercial data provider: a database of "movers" and a database of "nonmovers." The "movers" included all current residents who had moved to the corridor after 2004. From this database, we drew a random sample of about 1,000 residents from the Hiawatha corridor and about 500 residents from each of Nicollet, Bloomington, Coon

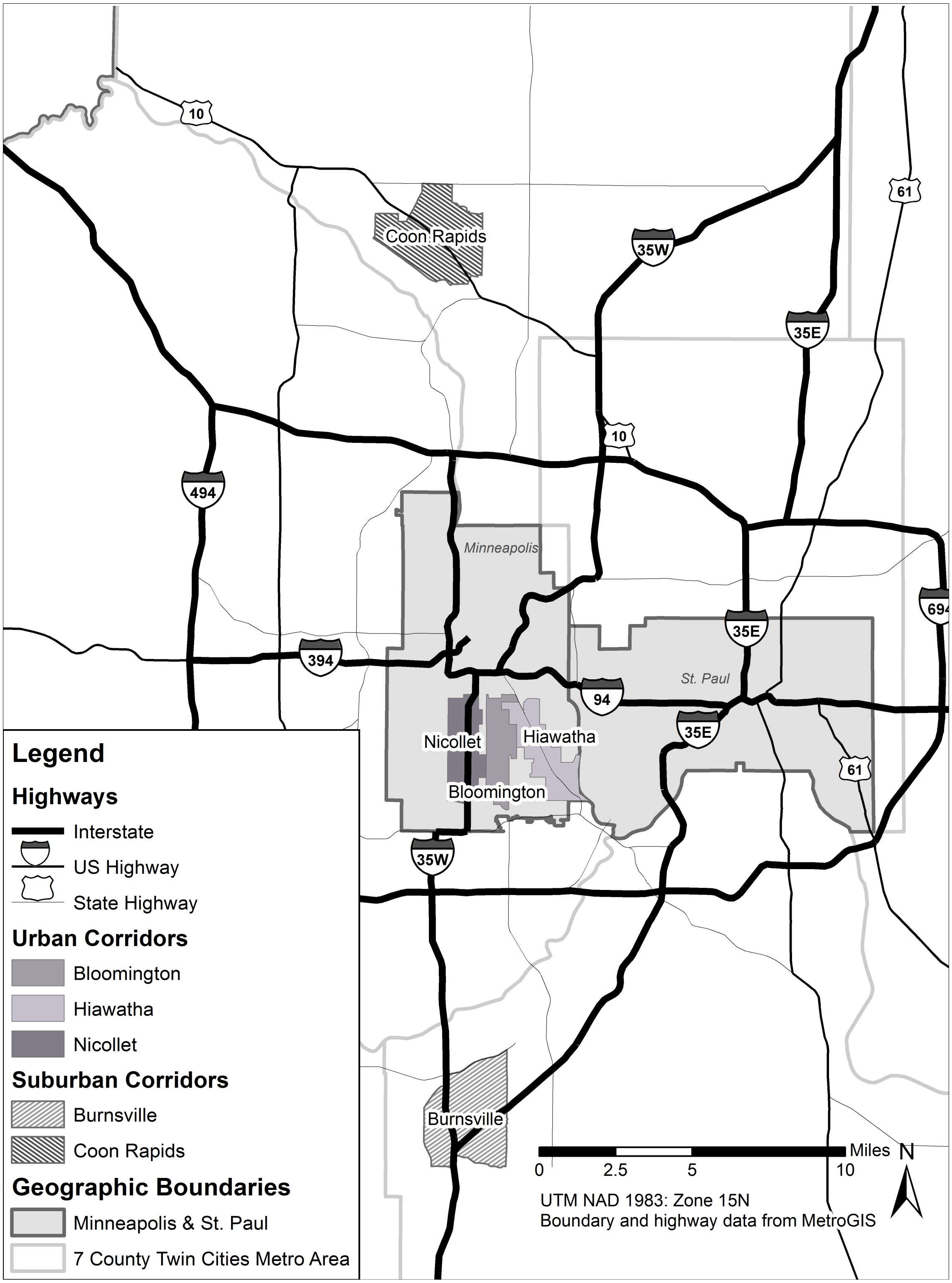

FIGURE 3.1: Five Original Corridors Surveyed
(Only urban corridors are analyzed in this thesis)

Rapids, and Burnsville corridors. The database of "nonmovers" consisted of a random sample of about 1,000 residents from the Hiawatha corridor and about 500 residents from each of the four corridors, who were not included in the "movers" list for each corridor.

The survey was pretested by students and staff members at the University of Minnesota and neighbors and friends of the investigators. Survey content was revised based on the feedback from pre-testers. The survey and two reminder postcards (1 and 2 weeks later) were mailed in May 2011. Ten $50 gift cards were provided as the incentive for the survey. The original database consisted of 6,017 addresses but only 5,884 were valid. The number of responses totaled 1,303, equivalent to a 22.2% response rate based on the valid addresses only. This is considered quite good for a survey of this length, since the response rate for a survey administered to the general population is typically 10-40% (Sommer & Sommer, 1997).

3.1.1 Sample Characteristics

This study focuses specifically on residents in the three urban corridors. All three urban corridors exhibit traditional urban development patterns: a well-connected street grid, high levels of transit service (LRT in the Hiawatha corridor and bus in Nicollet and Bloomington), a variety of land uses and housing types, and similar built environment context, as shown in Figure 3.2.

Although the sample was initially constructed to for the purpose of studying LRT, the three urban corridors include a variety of bicycling conditions. The Midtown Greenway trail runs east to west across the northern end of all three corridors, shown in Figure 3.2. Paired one-way bike lanes on Portland and Park Avenue run parallel to Nicollet and Bloomington avenues, about halfway in between them. The Hiawatha corridor contains both a separated trail along the LRT right-of-way (ROW) and bike lanes along Minnehaha Avenue, an arterial that parallels Hiawatha for most of the length of the corridor. The bike lanes on Portland, Park, and Minnehaha, and the trail along the Hiawatha LRT, all

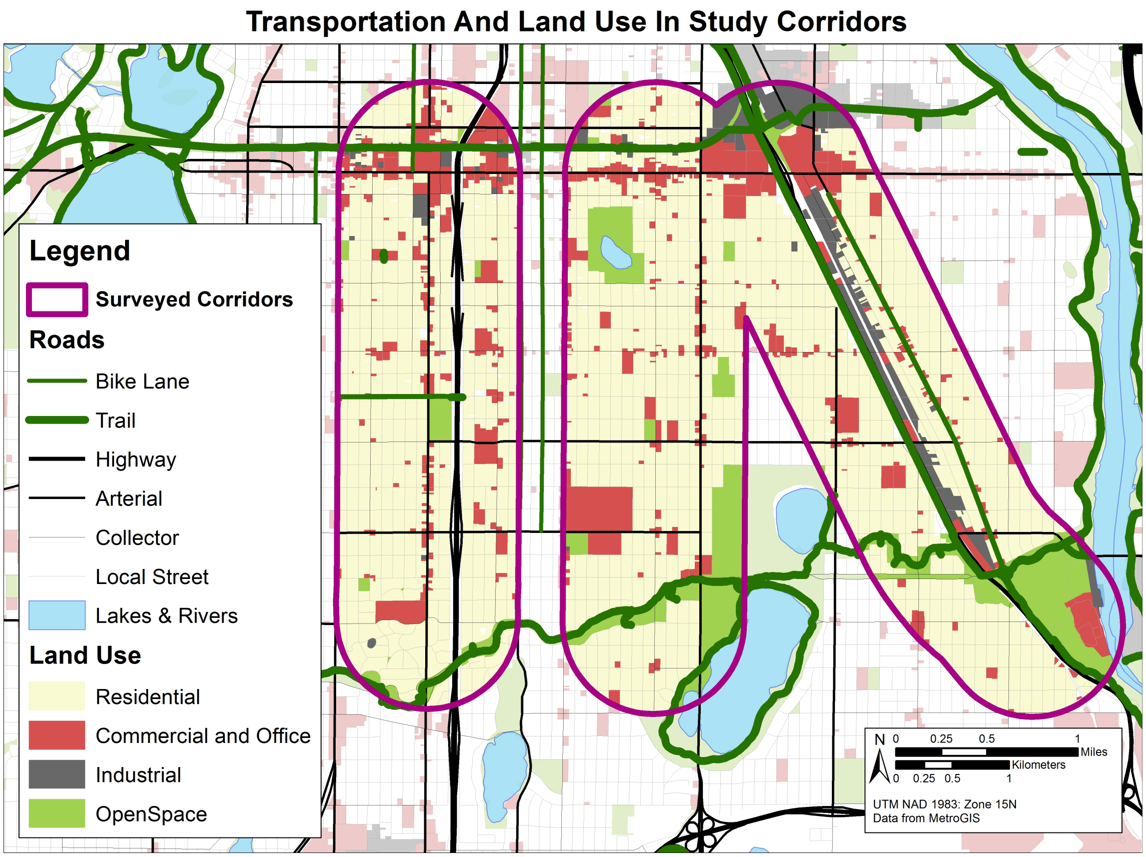

FIGURE 3.2: Street Network, Bicycle Infrastructure, and Land Use in Three Urban Study Corridors

serve Downtown Minneapolis. The bike trails toward the south end of the corridors are part of the Grand Rounds Scenic Byway and serve primarily recreational purposes.

Since this study focuses specifically on bicycling for the journey-to-work commute trip, survey participants who indicated that they are non-employed students or not working (e.g., retired, homemaker, or unemployed) were removed from the sample, for a final N of 614 respondents (161 bikers and 453 non-bikers). Table 3.1 demonstrates all adjustments made to the sample for purposes of this analysis.

Table 3.2 compares characteristics of survey respondents and the working sample with the 2011 ACS. Overall, homeowners and households with children are overrepresented

Table 3.1: Number removed, retained, and percent bicycle commuters at each step in sample reduction

	Bikers		Non-bikers		Total	
Removal Step:	**Remove**	**Remain**	**Remove**	**Remain**	**Remove**	**Remain**
(All Respondents)		230 (22%)		809		1,303
– Missing DV*	0	230 (22%)	0	809	264**	1,039
– Suburbs	18	212 (27%)	230	579	248	791
– Missing IV(s)***	30	182 (28%)	111	468	141	650
– Nonworkers	20	162 (26%)	15	453	35	615
– Bike Commute >35km	1	161 (26%)	0	453	1	614
Final Sample		**161 (26%)**		**453**		**614**

* Missing Dependent Variable: Respondent did not answer question about commute modes.

** Total number removed does not equal sum of Bikers and Non-bikers because respondents were not identifiable as either Bikers or Non-bikers.

*** Missing Independent Variable(s): Respondent did not answer one or more items having a hypothesized relationship with bicycle commuting in this study.

among survey respondents due to the oversampling of residents who have lived in their homes since before 2004. Respondents have a higher than average income and less likely to live in a zero-vehicle household. These are typical results for voluntary self-administered surveys.

3.2 Commuting

The dependent variable in this study is a measure of bicycle commuting frequency. Survey respondents were asked, "In a typical week with good weather, how many days do you use each of the following as your primary means of transportation between home and work/school?". Available modes included teleworking, driving alone, carpooling, transit, walking, biking, and other. For each mode, they were presented with six ordinal categories: (1) "Never", (2) "Less than once per month", (3) "1-3 days per month", (4) "Once per

Table 3.2: Demographics & Commute Mode Split
for Respondents and General Population

	2011 ACS		**2011 Survey**
	City	Tracts	Sample
Population / N	381,833	113,614	614
Pct. Female	49.8 %	50.0 %	50.8 %
Avg. HH Size	2.3	2.3	2.3
Pct. HH with Kids	24.2 %	28.8 %	26.1 %
Pct. Owner Occupied	50.4 %	57.5 %	86.1 %
Median Income	$ 47,478	$ 50,231	$ 75,561
Pct. Fulltime	58.1 %	62.6 %	85.0 %
Pct. Part time	23.9 %	21.1 %	15.0 %
Pct. Not Working	17.9 %	16.3 %	0 %
Avg. Vehicles/HH	1.7	1.7	1.7
Pct. Zero-Vehicle HH	8.8 %	8.1 %	4.0 %
Percent primarily commuting by:			
SOV	61.4 %	60.9 %	63.7 %
Carpool	8.5 %	11 %	4.2 %
Transit	14.0 %	16.1 %	14.4 %
Bike	3.9 %	4.2 %	8.5 %
Walk	6.4 %	2.6 %	3.8 %
Other	0.9 %	0.9 %	1.3 %
Telecommute	4.9 %	4.4 %	5.6 %

* Results displayed here do not sum to 100 % due to estimation procedure used to make survey results more comparable to ACS.

week", (5) "2-3 days per week", and (6) "4-5 days per week". This question can be found on Page 5 of the survey instrument, included in full in Appendix A.

These categories were recoded to approximate a count variable using the midpoints of each category (0, 1, 2, 4, 10, and 18 times per 4-week "month" respectively) to represent commute trips per month made by each mode. Participants were asked to report additional details about their commute, including the distance in both miles and minutes and whether their employer/school provides free parking.

Table 3.3: Bike Commuting Frequency

Response	Frequency	Percent
Never	453	73.8%
Total Non-bikers:	**453**	**73.8%**
Less than once per month	37	6.0%
1-3 days per month	30	4.9%
Once per week	22	3.6%
2-3 days per week	39	6.4%
4-5 days per week	33	5.4%
Total Bikers:	**161**	**26.2%**

Participation in commuting by bicycle was defined as any respondent who indicated that they commute by bike at least infrequently (categories 2 through 6).

The lower half of Table 3.2 focuses specifically on the journey to work mode for survey respondents and the general population. The ACS asks respondents to indicate their single primary mode of transportation for the commute trip (McKenzie & Rapino, 2011), assuming that people use only a single mode each and every day. This assumption is problematic for all commuters, and particularly for bicyclists because they experience barriers such as weather events or the need to transport goods that change from day to day (Heinen *et al.* , 2011b, p. 103)

Since the survey allowed people to indicate frequency of using a variety of modes, they have been adjusted slightly for presentation in Table 3.2. This is the percentage of respondents who indicated using each mode at least 4-5 times per week, plus 50 % of respondents who indicated using each mode 2-3 times per week, thus estimating the number of people who might indicate this as their "primary" mode. Table 3.3 shows the actual distribution of survey respondents among bike commute frequency choices. The sample contains fewer people who carpool and more bicyclists and telecommuters than the City of Minneapolis or the census tracts in which the respondents live. However, these differences are small and should not bias the results significantly.

3.3 Independent Variables

Table 3.4 presents the hypothesized relationship and descriptive statistics for all independent variables considered in this study, and the following sections explain how each variable is measured.

3.3.1 Built Environment

Built environment characteristics were measured in two ways: (1) through a set of survey questions asking respondents to indicate how true each of 29 neighborhood characteristics was of their neighborhood, and (2) using a GIS to objectively measure the infrastructure and land use around their homes.

3.3.1.1 Perceived Built Environment

Respondents rated how true each of 29 characteristics, such as "Large back yards" and "Easy access to transit stop/station", was of their current neighborhood. The ordinal scale ranged from (1) "Not at all true" to (4) "Entirely true". The two primary characteristics included in this study are "Good bicycle routes beyond the neighborhood" and "Close to where I work". Two additional characteristics, "Low crime rate within neighborhood" and "Low level of car traffic on neighborhood streets," were tested but found to be insignificant. The full list of characteristics is on the third page of the survey instrument, reproduced in Appendix A.

3.3.1.2 Objectively Measured Built Environment

To evaluate the built environment and its impacts on travel behavior, we constructed network distance buffers around each participant's homes at about 400−, 800−, and

Table 3.4: Variables with Descriptive Statistics

Variable	Description	Hypothesis	Mean	(S.D.)
Y	Number of days in a typical month with good weather that respondent commutes by bicycle *overall*	Dependent Variable	1.90	(4.58)
Y_b	Days per month commuting by bicycle *for bicyclists*		7.26	(6.43)
A	Jobs accessible by bike within 10 minutes (1000's)	+	0.61	(0.52)
C	Number of children under 12 in household	-	0.43	(0.83)
D	Respondent has college degree or higher	+	0.74	(0.44)
E_d	Commute distance in km	-	13.37	(13.17)
E_p	Employer provides free parking	-	0.69	(0.46)
E_t	Respondent works part time	-	0.15	(0.36)
F_b	Pro-biking Factor	+	0.29	(1.06)
F_d	Pro-driving Factor	-	−0.12	(1.15)
F_u	Pro-travel Factor	+	−0.03	(1.26)
G	Respondent's age in years	-	45.14	(12.64)
H	Land use entropy within 1600 m	+	0.35	(0.11)
I_b	Residential preference for "Good bicycle routes"	+	3.03	(1.07)
I_c	Residential preference for "Living unit on cul-de-sac"	-	1.36	(0.74)
I_w	Residential preference for "Close to where I work"	+	3.16	(0.88)
K	Income ($1000)	-	7.56	(3.34)
L	Respondent has a limitation that makes biking difficult	-	0.05	(0.21)
N_l	km of bike lane within 1600 m	+	3.37	(2.45)
N_t	km of bike trail within 1600 m	+	3.84	(2.79)
N_{4w}	Intersection density within 400 m	+	0.19	(0.03)
P_b	Perception: Good bicycle routes beyond the neighborhood	+	3.66	(0.62)
P_c	Perception: Living unit on cul-de-sac rather than through street	-	1.13	(0.52)
P_w	Perception: Close to where I work	+	2.87	(1.06)
V	Fewer than 1 car per driver in household	+	0.18	(0.38)
W	Respondent is female	-	0.51	(0.50)

1,600−meter ($\frac{1}{4}$−, $\frac{1}{2}$−, and 1−mile) distances. The network distance buffer includes only areas that the respondent could actually walk to.

Bicycle facilities around each respondent's home were measured by summing the total distance of bike lanes in meters within each respondent's network distance buffers.

An additional measure of bicycle cumulative opportunity accessibility summed the number of jobs accessible by bicycle within 10 minutes using OpenTripPlanner Development Team (2013), assuming an average bicycling speed of 16.1 kilometers per hour (10 miles per hour). Data inputs were 2010 Longitudinal Employer-Household Dynamics (LEHD) job estimates by census block and an OpenStreetMap (OSM) street network shapefile.

3.3.2 Residential Preferences

Respondents rated how important each of 29 characteristics, such as "Large back yards" and "Easy access to transit stop/station", was when they were last looking for a place to live. The ordinal scale ranged from (1) "Not at all important" to (4) "Extremely important". Respondents could also choose "I never considered it". For this same list of characteristics, respondents rated how accurately each of the statements represents their current neighborhood on a scale from (1) "Not at all true" to (4) "Entirely true". In this study, the two characteristics considered are "Good bicycle routes beyond the neighborhood" and "Close to where I work". The full list of characteristics is on the third page of the survey instrument, reproduced in Appendix A.

3.3.3 Travel Attitudes

To measure attitudes regarding travel, the survey asked respondents whether they agreed or disagreed with a series of 21 statements on a 5-point scale from "strongly disagree" (1) to "strongly agree" (5). Factor analysis was then used to extract the fundamental dimensions spanned by these 21 items, since some of the items are highly correlated. As shown in Table 3.5, seven underlying dimensions were identified: pro-drive, pro-walk, pro-bike, pro-transit, safety of car, status of car and pro-travel.

Pro-bike (F_b) and pro-travel (F_u) were selected for use in both models to control for travel attitudes.

3.4 Modeling Approach

The zero-inflated Poisson model was developed to account for count data that are overdispersed due to an excess number of zeroes in the dataset (Lambert, 1992). A binary logistic function is used to predict the probability of the dependent variable assuming a value of zero. Given this probability, the Poisson model is then fit to the non-zero data. These two equations are jointly modeled using Maximum Likelihood Estimation (MLE)

In human behavior, data are often overdispersed even after accounting for the presence of excess zeroes. A zero-inflated negative binomial distribution is used to model count data with excess zeroes that otherwise would violate the assumption of equal mean and variance in Poisson (Yau *et al.* , 2003). The distribution of zeroes follows a binary logistic distribution, similar to zero-inflated Poisson, but the values of non-zero observations are generated by a negative binomial process, as shown in equation 3.1:

$$Y_i \sim \begin{cases} 0 & \text{with probability } p_i \\ \mathrm{NB}(\lambda_i) & \text{with probability } 1 - p_i \end{cases} \tag{3.1}$$

In this thesis, Y_i is the number of days in a typical month with good weather on which person i commutes by bicycle. Zero outcomes ($Y_i = 0$), indicating non-participation in bicycle commuting, occur with probability p_i. The remaining count outcomes ($Y_i > 0$)

Table 3.5: Pattern Matrix for Travel Attitude Factors

	Safety of car	**Status of car**	**Pro-travel** (F_u)	**Pro-drive** (F_d)	**Pro-transit**	**Pro-bike** (F_b)	**Pro-walk**
Traveling by car is safer overall than walking.	0.748						
Traveling by car is safer overall than taking transit.	0.582						
Traveling by car is safer overall than riding a bicycle.	0.335					-0.308	
It does not matter to me which type of car I drive.		-0.642					
To me, the car is nothing more than a convenient way to get around.		-0.601					
To me, the car is a status symbol.		0.324		0.341			
Travel time is generally wasted time.			-0.615				
The only good thing about traveling is arriving at your destination.			-0.544				
Getting there is half the fun.			0.446				
I like to drive just for fun.				0.692			
I like driving.				0.665			
I feel free and independent if I drive.				0.491			
I like taking transit.					0.754		
Public transit can sometimes be easier for me than driving.					0.737		
I prefer to take transit rather than drive whenever possible.					0.711		
I prefer to bike rather than drive whenever possible.						0.846	
Biking can sometimes be easier for me than driving.						0.829	
I like riding a bike.						0.783	
I prefer to walk rather than drive whenever possible.							0.734
I like walking.							0.650
Walking can sometimes be easier for me than driving.							0.546

occur with probability $1 - p_i$, and are assumed to follow a Negative Binomial distribution with expected frequency λ and overdispersion parameter τ.

Thus for any value of y, the probability $P(Y_i = y)$ is represented by Equation 3.2 as per Mwalili *et al.* (2008).

$$P(Y = y) = \begin{cases} p + (1-p)(1+\frac{\lambda}{\tau})^{-\tau}, & y = 0 \\ (1-p)\frac{\Gamma(\lambda+\tau)}{y!\Gamma(\tau)}(1+\frac{\lambda}{\tau})^{-\tau}(1+\frac{\tau}{\lambda})^{-y}, & y = 1, 2, \ldots \end{cases} \tag{3.2}$$

It should be noted that the survey item used for this question collected grouped responses (e.g., "4-5 days per week"), so the values will not perfectly follow a Negative Binomial distribution (Schader & Schmid, 1985). This is discussed further in Section 6.4.

3.4.1 Model Development

Variables listed in Table 3.4 were selected based on their expected relationship with bicycle commuting from the literature. All statistical modeling was done using STATA 10.0. While many variable combinations were tested based on hypotheses about whether each variable contributed more strongly to the participation portion of the model or the frequency portion, the final model with best fit was derived by adding all variables to both halves of the equation and using a stepwise removal process based on $p-$values until all insignificant variables had been removed. The final model presented in this thesis shows each category of variables added separately in order to understand the relative contribution of built environment, demographic, residential preference, and travel attitude factors to the overall model fit.

Chapter 4

Results

4.1 Model Results

Table 4.1 shows the results from modeling participation and frequency of commuting by bicycle using Zero-Inflated Negative Binomial Regression (ZINB). First, we explain how to read the table. Step 1 includes only built environment predictors, and the subsequent models add controls for demographics and individual commute characteristics (Step 2), residential preferences (Step 3), and travel attitudes (Final Model). The two components of the ZINB model are shown separately, with participation on top and frequency below.

In unformatted ZINB results from statistical software, the coefficients represent “zero inflation”, or the relationship between that variable and the probability of the dependent variable assuming a value of 0. Coefficients have been reversed here for clarity. A positive coefficient indicates a positive association with the probability of participating in bicycle commuting.

Since these results come from the binary logit process in the ZINB model, the a one-unit increase in the independent variable can be interpreted as an increase in the log-odds of being a bicycle commuter (at any frequency) by the value of the coefficient. Using the final model (Model 4) as an example, a 1-unit increase in N_l (kilometers of bicycle lanes within 1600 meters of the respondent's home) is associated with a 0.15 increase in the log-odds of being a bicycle commuter. Exponentiating this gives us an odds ratio of 1.15 (not shown).

4.1.1 Model Fit

The third section of the table shows the $ln(\alpha)$ and α parameters from the negative binomial regression. Significant values indicate that α differs statistically from 1. Insignificant values, such as the model in Step 2, indicate a lack of overdispersion and suggest that a Poisson process would be sufficient. In the final model, the α parameter has the value of 0.62. In all iterations, it is greater than 0. α is significant at the 0.05 level for steps 1 and 3, as well as the final model, indicating the presence of overdispersion in the data. Finally, tests of model fit are shown. A significant value for the Likelihood Ratio Test (LR Test) is another indicator that the negative binomial model is a better fit than a Poisson model. Vuong's test statistic compares two competing non-nested models (Greene, 1994; Vuong, 1989). A positive and significant value in the Vuong test (e.g., 5.28 in the final model) indicates that the zero-inflated model is a better fit than Negative Binomial Regression (NBREG). These tests are significant in all steps of the model.

The Pseudo-R^2 value is a McFadden's Adjusted pseudo-R^2. There is no directly comparable measure in negative binomial or binary logistic regression to the classic R^2 used in OLS, which represents the percent of variation in the dependent variable that can be explained by the independent variables. The McFadden's Adjusted pseudo-R^2 takes on values from approximately 0 to 1, but it represents the relative improvement of this model's log likelihood over that of a null or constant-only model, with a penalty for additional parameters in the model.

Table 4.1: Results of Zero-Inflated Negative Binomial Regression

	Step 1 Built Environment		Step 2 + Demographics		Step 3 + Residential Pref.		Final Model + Travel Attitudes	
Variable	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Participation								
N_l km of bike lane	0.13***	(0.047)	0.089*	(0.047)	0.12**	(0.058)	0.15**	(0.065)
E_d km of commute			−0.086***	(0.016)	−0.095***	(0.019)	−0.11***	(0.022)
G Age			−0.041***	(0.0096)	−0.047***	(0.012)	−0.031**	(0.013)
F_b Pro-bike Factor							1.86***	(0.24)
F_u Pro-travel Factor							0.20*	(0.12)
Constant	−0.93***	(0.26)	1.87***	(0.57)	2.58***	(0.74)	0.35	(0.73)
Frequency								
P_c Cul-de-sac	−0.50**	(0.23)	−0.47**	(0.22)	−0.65***	(0.23)	−0.36*	(0.18)
A Accessibility	0.52***	(0.19)	0.43**	(0.17)	0.55***	(0.19)	0.25**	(0.12)
E_p Free parking			−0.48**	(0.20)	−0.58***	(0.21)	−0.50***	(0.15)
K Income			−0.097***	(0.030)	−0.12***	(0.030)	−0.075***	(0.023)
I_b Pref Bike Route					0.61***	(0.10)	0.23***	(0.085)
I_w Pref Near Work					0.22**	(0.11)	0.17*	(0.085)
F_b Pro-bike Factor							0.83***	(0.12)
F_u Pro-travel Factor							0.15**	(0.062)
Constant	1.77***	(0.31)	2.86***	(0.40)	0.14	(0.64)	0.21	(0.51)
$ln(\alpha)$	0.66**	(0.32)	0.37	(0.25)	0.64**	(0.26)	−0.48**	(0.22)
α	1.93	(0.61)	1.45	(0.37)	1.89	(0.49)	0.62	(0.13)
LR Test	492.75***		431.66***		422.29***		200.38***	
Vuong Test	1.91**		4.65***		4.18***		5.28***	
McFadden Pseudo−R^2	0.005		0.044		0.060		0.220	

*** *Significant at the* $p < 0.01$ level ** *Significant at the* $p < 0.05$ level * *Significant at the* $p < 0.1$ level

Examining the Pseudo-R^2 values for these four models provides a simple comparison of the relative importance of each set of variables. Model 1, with built environment variables only, has a very low pseudo-R^2 (0.005). Adding demographics increases this by an order of magnitude, and residential preferences cause another modest increase. Adding the travel attitude factors, however, causes the pseudo-R^2 to more than triple, suggesting that these are the most influential components in the model. The final model has a pseudo-R^2 of 0.220.

4.1.2 Built Environment

Of all the built environment variables hypothesized to have a relationship with reported bicycle commuting, only bike lanes (N_l), job accessibility (A), and living on a cul-de-sac (P_c) are significant. Interestingly, bike lanes appear in the participation portion of the equation. Bicycle lanes are more strongly associated with *whether* a person bike commutes at all than *how much* they do so. This result is likely connected to residential self-selection of people who like bicycling into suitable neighborhoods. People who like bicycling choose to live into neighborhoods with bike lanes, which enables them to commute by bicycle.

The coefficient for job accessibility (A) suggests that close proximity to jobs is an important predictor in how frequently one can make that commute trip by bicycle. It also probably serves as a proxy for other built environment variables that were pushed out of the model due to multicollinearity, such as density.

The cul-de-sac variable (P_c) is intuitive in that cul-de-sacs represent interruptions in the street grid, but the number of cul-de-sacs in the portion of South Minneapolis where this study occurred are limited. Further exploration of what street characteristics are associated with people's perceptions of living on a cul-de-sac is warranted. Additionally, the perceived built environment measure of living *on a cul-de-sac* was significant, where other street grid measures, such as number of cul-de-sacs within 400 meters of the respondent's home, were not. Given the low number of cul-de-sacs in South Minneapolis, it is reasonable to believe that only close proximity to this kind of street network interruption has a depressive effect on bicycling, rather than the presence of one or even a few cul-de-sacs farther away.

4.1.3 Demographics and Commute Characteristics

The coefficients on demographic and commuting variables are all intuitive, given existing literature on bicyclists, but the interpretation is not necessarily as straightforward on some of them. Longer commutes (E_d) decrease the probability that someone will commute by bicycle. The interpretation of this variable is unambiguous: each additional kilometer of commute distance represents a decrease in the log-odds of commuting by bicycle by 0.11.

Free parking (E_p) is associated with a decrease in the frequency with which a respondent commutes by bicycle. Given the spatial distribution of free parking in Minneapolis, however, it may be that this variable is simply serving as a proxy for working in Downtown Minneapolis or at the University of Minnesota since the respondents' actual work locations are not available in a geocoded format. Bicycle infrastructure connectivity to Downtown and the University of Minnesota is very strong, with several major north-south bike lanes and the Hiawatha LRT Bike Trail connecting the study area to Downtown.

The rest of the demographic variables are consistent with literature. Age (G) is negatively associated with participation in bicycle commuting, and income (K) is negatively associated with frequency after selecting for participation. Regarding income specifically, higher income does not decrease a person's odds of being a bicycle commuter, but given their chances of commuting by bike, they do so with lower frequency. This may indicate the presence of people who like bicycling but have more commute options available to them, so they can choose to bike when it is convenient and comfortable, and drive otherwise.

4.1.4 Residential Preferences

The positive coefficients on the importance of strong bike routes beyond the neighborhood (I_b) and living close to work (I_w) further suggest the self-selection effect. Curiously, however, these variables fit best in the frequency portion of the model. This reinforces the interpretation of free parking as a proxy for Downtown or University employment: a person who works at one of these pay-for-parking destinations is already prone to bicycling, but if they value (and then presumably self-select into) a neighborhood that is both close to work and has good bike routes beyond the neighborhood, these South Minneapolis corridors provide adequate infrastructure for bicycling often. If work locations could be secured, a network analysis may confirm this finding or provide a stronger interpretation.

4.1.5 Travel Attitudes

Both the Pro-bicycling factor (F_b) and the Pro-travel factor (F_u) are positively and significantly associated with participation and frequency. This is unsurprising, but still noteworthy because they represent such a large contribution in this model's explanatory power. The pro-travel factor contained sentiments about enjoying the journey as much as reaching the destination and valuing time spent in travel (versus believing it to be wasted). Utility theory for predicting mode share assumes that people will choose the mode that minimizes their cost and time investments, but these findings suggest that bicyclists derive value specifically from their commute. This is consistent with literature on positive utility of commuting in which people value mental separation from work and self-report non-zero ideal commute distances (Redmond & Mokhtarian, 2001). Further supporting this finding is a recent study by Paige Willis *et al.* (2013), in which bicyclists in particular derive satisfaction from many aspects of their commute, including the independence that bicycling provides, pleasure of the act of bicycling itself, and the ability to express one's self-defined identity as a cyclist.

Chapter 5

Comparison of Alternate Estimation Methods

5.1 Model Comparison Methodology

This chapter compares the results from the model presented in Chapters 3 and 4 to other modeling strategies found in the literature. Chapter 2 reviewed four strategies for managing the low bicycling mode share: inclusion criteria (2.2.1), strategic sampling (2.2.2), hybrid methods (2.2.3), and statistical techniques (2.2.4). Additionally, much of the literature about bicycle commuting and the built environment relies on binary logit or MNL discrete choice modeling to predict choice of the bicycle mode, but frequency does not receive much attention. Four alternate models were developed to test these different strategies using the same data and variables as the original model. They are summarized in 5.1.

Table 5.1: Models Tested to Compare Against Zero-Inflated Results

Model	Function	Cases Estimated On	DV	IV
Alternate Model 1	NBREG	Whole Sample	Frequency	All Variables
Alternate Model 2	NBREG	Subset of Actual Participants	Frequency	Frequency Variables
Alternate Model 3	LOGIT	Whole Sample	Participation	Participation Variables
Alternate Model 4	NBREG	Subset of Predicted Participants from Binary Logit Model	Frequency	Frequency Variables

5.1.1 Negative Binomial Model

Alternate Model 1 (AM1) uses negative binomial regression to predict days commuting by bicycle in a typical month with good weather in the entire sample. In this model, the α parameter addresses general overdispersion, but no joint estimation process is used to address excess zeroes. AM1 uses all 11 variables found in the original zero-inflated model.

5.1.2 Actual Subset Model

Alternate Model 2 (AM2) uses negative binomial regression, similar to AM1, but the cases only include people who indicate bicycling at least infrequently. This method approximates study designs described in section 2.2.1 that use an inclusion criteria to remove non-bicyclists from their sample, and focuses specifically on the relationship between the built environment and frequency of bike commuting for known bicyclists. Because participation is pre-determined in this model, only variables that appeared in the frequency section of the original zero-inflated model are used.

5.1.3 Logit Model

Alternate Model 3 (AM3) uses binary logistic regression to predict whether a person participates in bicycle commuting at all, similar to Krizek (2006) and Heinen *et al.* (2011b). Bicycling frequency is not addressed. This model uses only variables that appeared in the participation section of the original zero-inflated model.

5.1.4 Predicted Subset Model

Alternate Model 4 (AM4) predicts bicycling frequency among a subset of the sample that was predicted to participate in bicycling by the logit model in section 5.1.3. This method is an imperfect proxy for the strategic sampling method (section 2.2.2, in which the researcher deliberately surveys people who have a high probability of being bicyclists. Heinen *et al.* (2011a) and Heinen *et al.* (2011b) achieved this goal by surveying large employers in cities with above average bicycling rates: neither subset criteria guarantees bicycle commuters in their sample, but they increase the probability that any person they survey will be one. Moudon *et al.* (2005) and Rodríguez (2004) used a geographic approach, sampling from areas that had some suitability for bicycle commuting. Like AM2, this model uses only frequency variables.

5.2 Model Comparison Results

5.2.1 Model Coefficients

Table 5.2 and Table 5.3 present a comparison of model fit among the final model and four alternates. The first thing to note is that the signs on all coefficients in Table 5.2 are the same across the final ZINB model and four comparative models. Thus none of these other common strategies for predicting bicycle commuting give conflicting results, using the same data and set of variables. Some variables are not significant in the alternates: accessibility (A) is not a significant predictor of frequency in AM1, and free parking is not associated in AM4, for example. The different magnitudes and significances of coefficients is reasonable given the different model structures. Even though AM3 and AM4 contain the same components as the final ZINB model, the components of the zero-inflated model are jointly estimated by a maximum likelihood procedure.

McFadden's Adjusted Pseudo-R^2 values are reported for all models, but caution should be used in interpreting them. Since this type of pseudo-R^2 measures relative improvement in each model over its respective null model, the pseudo-R^2 does not provide a meaningful comparison point for use across different models (UCLA: Statistical Consulting Group, n.d.). The pseudo-R^2 in AM3 stands out because it is almost twice as high as the next best model. However, AM3 is a much simpler model than the rest. The dependent variable is a binary indicator of participation in commuting by bicycle. As shown in Table 5.3, AM3 predicts participation about as well as the final ZINB model and AM1, both of which have considerably lower pseudo-R^2 values.

The pseudo-R^2 for both subset models are very low, suggesting that they do not add much improvement over a null model.

5.2.2 Model Evaluation

Table 5.3 shows the results of tabulating predicted participation by actual participation and summarizing expected probability or frequency by actual frequency. For all results shown in this table, the coefficients from the respective models were used to predict results on the entire sample. The values in the top half of the table tell us how accurately each is predicting participation among the full sample. The final ZINB model and AM1 perform equally well at predicting participation, as indicated by the "Proportion Correct" row (sometimes referred to as the "Count R^2"). AM2 and AM4, the models estimated on

Table 5.2: Comparative modeling between final ZINB model and four alternate models.

Model:		**Final ZINB Model**		**Alternate Model 1**	**Alternate Model 2**	**Alternate Model 3**	**Alternate Model 4**
Variable		Pcpn. Eqn.	Freq. Eqn.	Freq. Eqn.	Pcpn. Eqn.	Freq. Eqn.	Freq. Eqn.
N_l	km of bike lane	0.15 **		0.069*	0.13 **		
P_c	Cul-de-sac		−0.36 *	−0.35 *		−0.25	−0.20
A	Accessibility		0.25 **	0.22		0.20 **	0.42 **
E_p	Parking		−0.50 ***	−0.70 ***		−0.37 ***	−0.24
K	Income		−0.075***	−0.047***		−0.061***	−0.069 **
E_d	km of commute	−0.11 ***		−0.054***	−0.0980***		
G	Age	−0.031**		−0.011	−0.026 **		
I_b	Pref Bike Route		0.23 ***	0.36 ***		0.13 *	0.08
I_w	Pref Near Work		0.17 *	0.13		0.12	0.12
F_b	Pro-bike Factor	1.86 ***	0.83 ***	1.71 ***	1.96 ***	0.69 ***	0.96 ***
F_u	Pro-travel Factor	0.20 *	0.15 **	0.23 ***	0.23 **	0.12 **	0.18 **
Constant		0.35	0.21	−1.11	−0.53	0.81 *	−0.0030
$ln(\alpha)$		−0.48 **		0.76 ***		−1.00 ***	0.055
α		0.62		2.15 ***		0.37 ***	1.06 ***
Pseudo−R^2		0.220	0.220	0.179	0.419	0.073	0.020

*** *Significant at the* $p < 0.01$ level ** *Significant at the* $p < 0.05$ level * *Significant at the* $p < 0.1$ level actual and predicted bicycling subset respectively, strongly overpredict participation in bicycling, as one might expect from that type of sample.

The second half of this table shows the average probability of participation (for the participation equation of the final ZINB model and AM3) or expected frequency (for the other four equations), both overall and stratified by the respondents' actual bicycling frequency. The frequencies predicted by the final ZINB model and AM1 are fairly close, though AM1 equation predicts higher values on average, with an increasing gap for more frequent bicyclists. Both subset equations overpredict frequency among infrequent cyclists, but produce better results for the most frequent cyclists.

The subset models result, while not unexpected, reinforces the limitation of study designs that strategically sample bicyclists or remove non-bicyclists from the study. The results can be used to describe what makes some bicyclists' behavior different from others', but they are not generalizable for estimating effects of the built environment, demographics, residential preference, or travel attitudes on the general population. While Wardman *et al.* (2007) claimed that a model is hardly needed to predict the bicycling behavior of people who have no interest or intention to bike, the model that includes these members of the population may perform better when predicting bicycling across a population that includes non-cyclists, such as number of bicycle trips in a region.

TABLE 5.3: Model Comparison: Predicted participation and frequency using final ZINB model and four alternate models.

Model:	**Final ZINB Model** (Pcpn. Eqn.)	**Final ZINB Model** (Freq. Eqn.)	**Alternate Model 1** (Freq. Eqn.)	**Alternate Model 2** (Freq. Eqn.)	**Alternate Model 3** (Pcpn. Eqn.)	**Alternate Model 4** (Freq. Eqn.)
Predicted Participation Count:						
Nonbiker	395	383	398	109	419	262
Biker	123	134	127	161	103	147
False Positive	56	70	55	344	34	191
False Negative	36	27	34	0	58	14
Proportion Correct	0.85	0.84	0.86	0.44	0.85	0.67
Predicted Probability & Frequency:						
Overall	0.32	2.00	2.54	3.69	0.26	2.57
Predicted average for people who selected						
"Never"	0.19	0.64	0.64	2.39	0.14	1.36
"Less than 1 day/month"	0.49	2.17	2.26	4.02	0.40	2.51
"1-3 days/month"	0.58	2.67	2.89	4.79	0.47	3.18
"1 day/week"	0.71	3.60	4.02	5.52	0.62	3.91
"2-3 days/week"	0.83	8.15	10.9	9.49	0.78	8.05
"4-5 days/week"	0.82	11.39	17.73	12.21	0.78	11.25

Chapter 6

Discussion and Conclusions

6.1 Catalysts & Magnets

The model results show that bicycle commuting participation and frequency are associated with different built environment measures. However, interpreting these results requires some measure of caution. As discussed extensively in other literature, establishing causality between travel behavior and the built environment is challenging (Mokhtarian & Cao, 2008; Cao *et al.* , 2009a). The Zero-Inflated Negative Binomial Regression (ZINB) model helps control for residential self-selection effects that confound an observed relationship between built environment characteristics and commute mode choice by predicting participation (self-selection) separately from frequency.

A variable in a model of bicycling participation that has a causal impact on bicycling behavior could be considered a “catalyst”; these variables encourage people who are not already doing so to start bicycling. Conversely, a variable that indicates residential self-selection of people already prone to bicycling into bike-friendly neighborhoods can be called a “magnet”. The presence of one of these magnets in a neighborhood does not cause a non-bicyclist to become a regular bike commuter, but it does provide a neighborhood that meets the travel needs and preferences of bicyclists, giving them better opportunities to bicycle. Given the structure of the model employed here, coefficients in the participation half of the model are likely to have a "magnet" effect that overshadows any possible "catalyst" effect.

6.2 Implications for Practice

Bicycle infrastructure, measured as kilometers of bike lane within 1 kilometer of the respondent's home, is significant in the participation portion of the model. However, subsequent bivariate correlation tests (not shown) between survey respondents' pro-bike travel attitude factor (F_b) and their length of tenure in a neighborhood (stratified by age) failed to provide evidence of temporal precedence. Travel attitudes are stronger among residents who moved into their current more recently, suggesting that travel attitudes precede location choice. This suggests that bicycle infrastructure functions more like a magnet than a catalyst. While evidence to infer causality is lacking, it is evident that people who are more likely to use a bicycle for commuting do in fact live near these facilities. This is an important finding because it clarifies how bicycle infrastructure can be used to support bicycling for transportation. It implies that placing new bicycle infrastructure around other built environment characteristics that do appear to influence bicycling will magnify their effects by attracting residents with a propensity to commute by bicycle.

In the frequency half of the equation, cul-de-sacs, job accessibility, and free parking all have significant coefficients. The relationships to job accessibility and cul-de-sacs are intuitive: cul-de-sacs interrupt the street network, and respondents with greater accessibility to jobs by bicycling have a higher probability of working within a reasonable bike distance from home.

While it is possible that free parking directly influences how often a respondent commutes by bicycle, this is probably not its only mechanism of action. Pay-for-parking is located primarily Downtown and around the University of Minnesota, whereas free parking is the norm in most other parts of the city. There are some exceptions, but the general trend suggests a probable association between a lack of free parking at the respondent's work and that workplace being located either Downtown or at the University. Connectivity to Downtown and the University via bicycle is excellent due to several closely-spaced major north-south bicycle routes, whereas connectivity throughout the rest of the city is limited. Someone who lives in South Minneapolis and commutes to one of these locations likely has a much more comfortable route via existing lane and trail infrastructure than someone who commutes to another part of the region. Thus the free parking variable in this model is probably functioning as a proxy for how well the bicycle network serves a person's commute trip.

The accessibility and parking variables have significant implications for practice, given the finding about bicycle lanes. After controlling for self-selection, which includes self-selection into close proximity of bike lanes, these two variables still have a significant relationship with frequency of commuting by bicycle. If a city aspires to increase bicycling, these results suggest that new bicycle infrastructure should be deployed in neighborhoods with high accessibility to employment, and the routes should be designed to provide connections to major job centers. As new residents self-select into the neighborhood because of the bike lane, these other factors will enable them to bike more frequently.

6.3 Implications for Research

This research showed that modeling bicyclists as a subset does not produce generalizable results. While this finding is not surprising, it provides justification for studies that compare bicyclists to the general population, despite their low numbers. Negative Binomial regression produced similar estimates to the zero-inflated model employed in this thesis, but the results do not provide the additional insight of piecing together separate influences on participation and frequency.

This study's methodology and results provide a framework for future research on nonmotorized travel behavior. While commuting is a common research topic because peak hour traffic congestion garners so much interest, many bicyclists use bikes primarily for non-work trips. Given the relative inflexibility of the commute trip to other types of travel, such as household errands, shopping, or socializing, identifying factors associated with participation in and frequency of bicycling for non-work trips may prove even more useful for practice. Additionally, walking and transit face many challenges similar to bicycling, such as exposure to daily weather phenomena. This methodology may have applications to walking and other underrepresented modes.

6.4 Limitations and Areas for Further Study

The issue of free parking at a respondent's work location raises several questions about how the route choices available to a person might influence their mode choice. This study measured built environment characteristics around participants' homes, but not their work locations nor along possible routes connecting the two. Previous studies have found that characteristics along the route are stronger predictors of nonmotorized travel (Winters *et al.* , 2010; Srinivasan, 2002), and the free parking variable in this study demonstrates the need of exploring this issue in greater detail.

Future research should identify respondents' work locations and construct separate measures for bicycle infrastructure along the routes in between home and work. A measure of job accessibility that uses a modified network with stronger weights on jobs accessible via a route comprised mostly of dedicated infrastructure might serve as a sufficient proxy, if work locations are not available. While this still would not resolve whether infrastructure serves as a catalyst or magnet, it would clarify the contexts that make infrastructure relevant to travel decisions.

We modeled the participation component of bicycle commuting as a choice between being a bicycle commuter, or not being a bicycle commuter who uses any and all other modes. Some studies consider bicycling versus a specific choice, such as driving (Handy & Xing, 2011). Both strategies have merit: focusing specifically on bicycling versus driving removes possible dampening effects on the results from consolidating walking, transit, and driving into a single "non-bicycle" mode category. Built environment characteristics that support walking and transit have more in common with bicycle-friendly spaces than car-supportive environments. However, walking and transit are also small mode shares, so any dampening effect may not have much of an impact. Additionally, despite research and policy that collapses all non-auto modes into a broader category of "alternative transportation", there are distinct differences in how each mode functions and what needs it serves. Walking is much slower than bicycling and is better suited for short trips. Bicycling allows for on-demand transportation, similar to private auto travel, versus the fixed schedule constraints of transit. So the boundaries are not clear-cut. Comparing bicycling to all other modes in aggregate, as was done here, has a lower potential to overestimate results.

As mentioned in Section 3.4, the commute mode data collected in the survey represented grouped counts (e.g., "4-5 days per week"). The analysis in this thesis assumed that these grouped counts could be sufficiently approximated by a Negative Binomial distribution, but several other techniques are worth exploring for future studies. Schader & Schmid (1985) describe three algorithms that can be used for MLE of parameters from grouped data resulting from a Negative Binomial distribution. Alternatively, Pinjari *et al.* (2008) jointly estimate a logistic model of neighborhood choice and an ordered logit model of bicycle ownership. The ordered logit model may be more suitable for the grouped response data provided by the survey.

Finally, this study was not able to establish causality the relationship between bicycling and the built environment. The four main criteria required for *causality* include (1) association, (2) non-spuriousness, (3) time-precedence, and (4) causal mechanism.

Mokhtarian & Cao (2008)'s review of methodological techniques from the framework of requirements for inferring causality concludes that they have a strong capacity to show association, they are lacking in ability to establish time-precedence. Cross-sectional data, as used in this thesis, is notably weak in this regard. As previously mentioned, a bivariate correlation test between the strength of travel attitudes length of tenure in a neighborhood confirms that temporal precedence is a likely issue in this study. Nonetheless, the rest of the evidence presented in this study considerably advances the conversation about self-selection and the built environment.

Bibliography

Bibliography

Akar, Gulsah, & Clifton, Kelly J. 2009. Influence of Individual Perceptions and Bicycle Infrastructure on Decision to Bike. *Transportation Research Record*, **2140**, 165–172.

Buehler, Ralph. 2012. Determinants of Bicycle Commuting in the Washington, DC Region: The Role of Bicycle Parking, Cyclist Showers, and Free Car Parking at Work. *Transportation Research Part D: Transport and the Environment*, **17**(7), 525–531.

Cao, Xinyu. 2009. Disentangling the influence of neighborhood type and self-selection on driving behavior: an application of sample selection model. *Transportation*, **36**(2), 207–222.

Cao, Xinyu, & Schoner, Jessica E. 2013. The Influences of Light Rail Transit On Transit Uses: An Exploration of Station Area Residents along the Hiawatha Line in Minneapolis. *TRB 92nd Annual Meeting Compendium of Papers*, **#13-0277**.

Cao, Xinyu, Mokhtarian, Patricia L., & Handy, Susan L. 2008. Differentiating the Influence of Accessibility, Attitudes, and Demographics on Stop Participation and Frequency During the Evening Commute. *Environment and Planning B: Planning and Design*, **35**(3), 431–442.

Cao, Xinyu, Mokhtarian, Patricia L., & Handy, Susan L. 2009a. Examining the Impacts of Residential Self Selection on Travel Behaviour: A Focus on Empirical Findings. *Transport Reviews*, **29**(3), 359–395.

Cao, Xinyu, Mokhtarian, Patricia L., & Handy, Susan L. 2009b. The Relationship Between the Built Environment and Nonwork Travel: A Case Study of Northern California. *Transportation Research Part A: Policy and Practice*, **43**(5), 548–559.

Chin, Hoong C., & Quddus, Mohammed Abdul. 2003. Modeling Count Data with Excess Zeroes: An Empirical Application to Traffic Accidents. *Sociological Methods & Research*, **32**(1), 90–116.

Dill, Jennifer, & Voros, Kim. 2007. Factors Affecting Bicycling Demand: Initial Survey Findings from the Portland, Oregon, Region. *Transportation Research Record*, **2031**, 9–17.

Ewing, Reid, & Cervero, Robert. 2010. Travel and the Built Environment. *Journal of the American Planning Association*, **76**(3), 265–294.

Federal Highway Administration. 2012. *Report to the U. S. Congress on the Outcomes of the Nonmotorized Transportation Pilot Program SAFETEA LU Section 1807*. Tech. rept. April. Submitted by the Federal Highway Administration With the Assistance of the U.S. Department of Transportation's Volpe National Transportation Systems Center, Washington, D.C.

Geller, Roger. 2006. *Four Types of Cyclists*. Tech. rept. Portland Office of Transportation, Portland, OR.

Gordon, Peter, & Richardson, Harry W. 1998. Bicycling in the United States: A Fringe Mode? *Transportation Quarterly*, **52**(1), 9–11.

Greene, John. 1994. Accounting for excess zeroes and sample selection in poisson and negative binomial regression models. *Biology & Philosophy*, **9**(3), 265–265.

Handy, Susan L., & Xing, Yan. 2011. Factors Correlated with Bicycle Commuting: A Study in Six Small US Cities. *International Journal of Sustainable Transportation*, **5**(2), 91–110.

Handy, Susan L., Cao, Xinyu, & Mokhtarian, Patricia L. 2006. Self-Selection in the Relationship Between the Built Environment and Walking: Empirical Evidence from Northern California. *Journal of the American Planning Association*, **72**(1), 55–74.

Heinen, Eva, van Wee, Bert, & Maat, Kees. 2010. Commuting by Bicycle: An Overview of the Literature. *Transport Reviews*, **30**(1), 59–96.

Heinen, Eva, Maat, Kees, & van Wee, Bert. 2011a. Day-to-Day Choice to Commute or Not by Bicycle. *Transportation Research Record*, **2230**, 9–18.

Heinen, Eva, Maat, Kees, & Wee, Bert Van. 2011b. The Role of Attitudes Toward Characteristics of Bicycle Commuting on the Choice to Cycle to Work over Various Distances. *Transportation Research Part D: Transport and the Environment*, **16**(2), 102–109.

Hunt, J. D. 2009. An Examination of Bicycle Use Sensitivities Over Time Using Stated Preference. *IAPR Technical Paper Series*, **TP-06015**.

Hunt, J. D., & Abraham, J. E. 2007. Influences on Bicycle Use. *Transportation*, **34**(4), 453–470.

Krizek, Kevin J. 2006. Two Approaches to Valuing Some of Bicycle Facilities Presumed Benefits. *Journal of the American Planning Association*, **72**(3), 309–320.

Krizek, Kevin J., & Johnson, Pamela Jo. 2006. Proximity to Trails and Retail: Effects on Urban Cycling and Walking. *Journal of the American Planning Association*, **72**(1), 33–42.

Lambert, Diane. 1992. Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. *Technometrics*, **34**(1), 1–14.

McKenzie, Brian, & Rapino, Melanie. 2011. *Commuting in the United States : 2009*. Tech. rept. September. US Census Bureau, Washington, D.C.

Metropolitan Council. *Travel Behavior Inventory (TBI) Surveys*.

Mokhtarian, Patricia L., & Cao, Xinyu. 2008. Examining the Impacts of Residential Self-Selection on Travel Behavior: A Focus on Methodologies. *Transportation Research Part B: Methodological*, **42**(3), 204–228.

Moudon, Anne Vernez, Lee, Chanam, Cheadle, Allen D., Collier, Cheza W., Johnson, Donna, Schmid, Thomas L., & Weather, Robert D. 2005. Cycling and the Built Environment, a US Perspective. *Transportation Research Part D: Transport and the Environment*, **10**(3), 245–261.

Mwalili, Samuel M, Lesaffre, Emmanuel, & Declerck, Dominique. 2008. The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. *Statistical methods in medical research*, **17**(2), 123–39.

OpenTripPlanner Development Team. 2013. *OpenTripPlanner*. Version 0.9.1: OpenPlans, GNU Lesser Public General License.

Paige Willis, Devon, Manaugh, Kevin, & El-Geneidy, Ahmed. 2013. Uniquely Satisfied: Exploring Cyclist Satisfaction. *Transportation Research Part F: Traffic Psychology and Behaviour*, **18**, 136–147.

Parkin, John, Wardman, Mark, & Page, Matthew. 2007. Estimation of the Determinants of Bicycle Mode Share for the Journey to Work Using Census Data. *Transportation*, **35**(1), 93–109.

Pinjari, Abdul Rawoof, Eluru, Naveen, Bhat, Chandra R., Pendyala, Ram M., & Spissu, Erika. 2008. Joint Model of Choice of Residential Neighborhood and Bicycle Ownership: Accounting for Self-Selection and Unobserved Heterogeneity. *Transportation Research Record*, **2082**, 17–26.

Porter, Christopher, Suhrbier, John, & Schwartz, William L. 1999. Forecasting Bicycle and Pedestrian Travel: State of the Practice and Research Needs. *Transportation Research Record*, **1674**, 94–101.

Redmond, LS, & Mokhtarian, PL. 2001. The positive utility of the commute: modeling ideal commute time and relative desired commute amount. *Transportation*, **28**, 179–205.

Rodríguez, Daniel A. 2004. The Relationship Between Non-Motorized Mode Choice and the Local Physical Environment. *Transportation Research Part D: Transport and the Environment*, **9**(2), 151–173.

Schader, Martin, & Schmid, Friedrich. 1985. Computation of M.L. Estimates for the Parameters of a Negative Binomial Distribution from Grouped Data. A Comparison of the Scoring, Newton-Raphson, and E-M Algorithms. *Applied Stochastic Models and Data Analysis*, **1**, 11–23.

Schoner, Jessica E., & Cao, Xinyu. 2013. Influences of Light Rail Transit, the Built Environment, and Residential Self Selection on Pedestrian Travel Behavior: Evidence from Minneapolis, MN. *Presented at the 2012 ACSP Annual Conference*, 1–14.

Sener, Ipek N., Eluru, Naveen, & Bhat, Chandra R. 2009. Who Are Bicyclists? Why and How Much Are They Bicycling? *Transportation Research Record*, **2134**, 63–72.

Sommer, Barbara B., & Sommer, Robert. 1997. *A Practical Guide to Behavioral Research: Tools and Techniques*. 4th edn. Oxford: Oxford University Press.

Srinivasan, Sumeeta. 2002. Quantifying Spatial Characteristics of Cities. *Urban Studies*, **39**(11), 2005–2028.

Thakuriah, Piyushimita (Vonu), Metaxatos, Paul, Lin, Jane, & Jensen, Elizabeth. 2012. An Examination of Factors Affecting Propensities to Use Bicycle and Pedestrian Facilities in Suburban Locations. *Transportation Research Part D: Transport and the Environment*, **17**(4), 341–348.

UCLA: Statistical Consulting Group. *FAQ: What are pseudo R-squareds?*

Vuong, Quang H. 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, **57**(2), 307–333.

Wardman, Mark, Tight, Miles, & Page, Matthew. 2007. Factors Influencing the Propensity to Cycle to Work. *Transportation Research Part A: Policy and Practice*, **41**(4), 339–350.

Winters, Meghan, Brauer, Michael, Setton, EM, & Teschke, Kay. 2010. Built Environment Influences on Healthy Transportation Choices: Bicycling Versus Driving. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, **87**(6), 969–993.

Xing, Yan, Handy, Susan L., & Mokhtarian, Patricia L. 2010. Factors Associated with Proportions and Miles of Bicycling for Transportation and Recreation in Six Small US Cities. *Transportation Research Part D: Transport and the Environment*, **15**(2), 73–81.

Yau, Kelvin K. W., Wang, Kui, & Lee, Andy H. 2003. Zero-Inflated Negative Binomial Mixed Regression Modeling of Over-Dispersed Count Data with Extra Zeros. *Biometrical Journal*, **45**(4), 437–452.

Appendices

Appendix A

Survey Instrument

UNIVERSITY OF MINNESOTA

Twin Cities Campus | *Hubert H. Humphrey School of Public Affairs* | *130 Hubert H. Humphrey Center*
301–19th Avenue South
Minneapolis, MN 55455

www.hhh.umn.edu

May 2, 2011

Dear Resident,

The University of Minnesota is conducting a study of choices about where to live and about daily travel of residents from different kinds of neighborhoods. It is sponsored by Minnesota Department of Transportation, Metropolitan Council, Hennepin, Dakota, and Anoka Counties, among others. By understanding how the characteristics of the places where we live affect the transportation decisions we make, policy-makers can better address our transportation challenges.

I am writing to ask for your household's participation in this study. Your neighborhood is one of five in the Twin Cities Metropolitan Area that we have selected for this study, and your household has been randomly selected within your neighborhood. Your participation in the survey is voluntary and your responses are completely confidential. In any report we might publish, we will not include any information that would make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records. Thus, the study does not entail any personal risks.

When deciding whether to answer the enclosed survey, please remember that your participation will allow us to study the transportation needs of households similar to yours. Your response to each question is therefore critical to the study. We provide space at the end of the survey for additional comments that you think may be relevant. The survey should take about 20 minutes to complete.

Any adult household member who shares in the decision making for your household and who participated in selecting your current residence can complete the survey. Please return the survey **no later than June 13, 2011** using the enclosed business reply mail envelope. To show our appreciation, ***every household that returns a completed survey will be entered into a drawing for the chance to win one of ten gift card prizes of $50 each.*** The ID number at the top of this letter is there simply to identify whether you are the winner and which neighborhood you live in. Your responses will not be linked to your name in any other way.

Thank you in advance for participating in this study. If you have any questions, I encourage you to contact me directly at (612) 625-5671 or cao@umn.edu, or my research assistant, Ms. Jessica Schoner, (612) 412-4273 or schon082@umn.edu. If you're interested in learning more about me and the kind of research I do, please visit my website at:
http://www.hhh.umn.edu/people/jcao/index.html

Sincerely,

Jason Cao

Professor Jason Cao, Principle Investigator

Driven to Discover℠

Your Residence, Neighborhood, and Satisfaction

The questions in this section ask about your residence, neighborhood, and satisfaction. By "residence" we mean the house, apartment, or any other type of housing unit in which you live. By "neighborhood" we mean the residential area in which your residence is located – the area you consider to be your neighborhood.

1. How would you describe the type of housing unit in which you currently live?
   - ☐1 Apartment/Condo
   - ☐2 Townhouse
   - ☐3 Duplex
   - ☐4 Single-family detached house
   - ☐5 Other (please specify) ________________________

2. When did you move to your current residence? Month: _______________ Year: _____________

3. Please indicate the approximate location of your **previous** residence – the residence you lived in before moving into your current residence. Don't include places you lived in temporarily while moving to your current residence.

Street or nearest cross-streets: __________________ City: __________________ State/Country: __________________

4. On a seven-point scale, how well do **the characteristics of your neighborhood** meet the ***current needs*** of your household?

Extremely poorly			*Neutral*			*Extremely well*
☐1	☐2	☐3	☐4	☐5	☐6	☐7

5. Please indicate the extent to which you agree or disagree with each of the following statements on a seven-point scale from "strongly disagree" to "strongly agree." There are no right or wrong answers.

	Strongly Disagree			*Neutral*			*Strongly Agree*
When I think of my daily travel, the positive aspects outweigh the negative	☐1	☐2	☐3	☐4	☐5	☐6	☐7
If I could live my life over again, I would change almost nothing	☐1	☐2	☐3	☐4	☐5	☐6	☐7
I do not want to change anything regarding my daily travel	☐1	☐2	☐3	☐4	☐5	☐6	☐7
The conditions of my life are excellent	☐1	☐2	☐3	☐4	☐5	☐6	☐7
So far I have achieved the important things I want in life	☐1	☐2	☐3	☐4	☐5	☐6	☐7
My travel facilitates my daily life	☐1	☐2	☐3	☐4	☐5	☐6	☐7
I am satisfied with my life	☐1	☐2	☐3	☐4	☐5	☐6	☐7
My daily travel makes me feel good	☐1	☐2	☐3	☐4	☐5	☐6	☐7
I am completely satisfied with my daily travel	☐1	☐2	☐3	☐4	☐5	☐6	☐7
In most ways my life is close to my ideal	☐1	☐2	☐3	☐4	☐5	☐6	☐7

6. Please indicate ***how important*** each of the following characteristics was ***when you were looking for a place to live*** on a scale from “not at all important” to “extremely important.”

		Not at all important			*Extremely important*	*I never considered it*
□7	Affordable living unit	□1	□2	□3	□4	□5
□7	High quality living unit	□1	□2	□3	□4	□5
□7	Living unit on cul-de-sac rather than through street	□1	□2	□3	□4	□5
□7	Good investment potential	□1	□2	□3	□4	□5
□7	High quality K-12 schools	□1	□2	□3	□4	□5
□7	Attractive appearance of neighborhood	□1	□2	□3	□4	□5
□7	Variety in housing styles	□1	□2	□3	□4	□5
□7	High level of upkeep in neighborhood	□1	□2	□3	□4	□5
□7	Large back yards	□1	□2	□3	□4	□5
□7	Lots of off-street parking (garages or driveways)	□1	□2	□3	□4	□5
□7	Sidewalks throughout the neighborhood	□1	□2	□3	□4	□5
□7	Good bicycle routes beyond the neighborhood	□1	□2	□3	□4	□5
□7	Good public transit service (bus or rail)	□1	□2	□3	□4	□5
□7	Easy access to transit stop/station	□1	□2	□3	□4	□5
□7	Parks and open spaces nearby	□1	□2	□3	□4	□5
□7	Shopping areas within walking distance	□1	□2	□3	□4	□5
□7	Easy access to a regional shopping mall	□1	□2	□3	□4	□5
□7	Easy access to downtown	□1	□2	□3	□4	□5
□7	Religious or civic buildings (ex., library) nearby	□1	□2	□3	□4	□5
□7	Close to where I work	□1	□2	□3	□4	□5
□7	Low crime rate within neighborhood	□1	□2	□3	□4	□5
□7	Low level of car traffic on neighborhood streets	□1	□2	□3	□4	□5
□7	Quiet neighborhood	□1	□2	□3	□4	□5
□7	Good street lighting	□1	□2	□3	□4	□5
□7	Safe neighborhood for walking	□1	□2	□3	□4	□5
□7	Safe neighborhood for kids to play outdoors	□1	□2	□3	□4	□5
□7	Lots of interaction among neighbors	□1	□2	□3	□4	□5
□7	Lots of people out and about within the neighborhood	□1	□2	□3	□4	□5
□7	Diverse neighbors in terms of ethnicity, race, and age	□1	□2	□3	□4	□5
□7	Economic level of neighbors similar to my level	□1	□2	□3	□4	□5

7. Now, looking back at the characteristics to which you gave the highest importance in Question 6, **which THREE do you consider MOST important?** Please check the boxes left to the items in Question 6.

8. In this question, we'd like to know ***what your current residence and neighborhood are like***. Please indicate ***how true*** each of the characteristics is for your current residence and neighborhood on a scale from "not at all true" to "entirely true."

	Not at all true			*Entirely true*
Affordable living unit	□1	□2	□3	□4
High quality living unit	□1	□2	□3	□4
Living unit on cul-de-sac rather than through street	□1	□2	□3	□4
Good investment potential	□1	□2	□3	□4
High quality K-12 schools	□1	□2	□3	□4
Attractive appearance of neighborhood	□1	□2	□3	□4
Variety in housing styles	□1	□2	□3	□4
High level of upkeep in neighborhood	□1	□2	□3	□4
Large back yards	□1	□2	□3	□4
Lots of off-street parking (garages or driveways)	□1	□2	□3	□4
Sidewalks throughout the neighborhood	□1	□2	□3	□4
Good bicycle routes beyond the neighborhood	□1	□2	□3	□4
Good public transit service (bus or rail)	□1	□2	□3	□4
Easy access to transit stop/station	□1	□2	□3	□4
Parks and open spaces nearby	□1	□2	□3	□4
Shopping areas within walking distance	□1	□2	□3	□4
Easy access to a regional shopping mall	□1	□2	□3	□4
Easy access to downtown	□1	□2	□3	□4
Religious or civic buildings (ex., library) nearby	□1	□2	□3	□4
Close to where I work	□1	□2	□3	□4
Low crime rate within neighborhood	□1	□2	□3	□4
Low level of car traffic on neighborhood streets	□1	□2	□3	□4
Quiet neighborhood	□1	□2	□3	□4
Good street lighting	□1	□2	□3	□4
Safe neighborhood for walking	□1	□2	□3	□4
Safe neighborhood for kids to play outdoors	□1	□2	□3	□4
Lots of interaction among neighbors	□1	□2	□3	□4
Lots of people out and about within the neighborhood	□1	□2	□3	□4
Diverse neighbors in terms of ethnicity, race, and age	□1	□2	□3	□4
Economic level of neighbors similar to my level	□1	□2	□3	□4

9. Now we'd like to know ***what your PREVIOUS residence and neighborhood were like***. Please indicate ***how true*** each of the characteristics was for your ***PREVIOUS*** residence and neighborhood on a scale from "not at all true" to "entirely true."

	Not at all true			*Entirely true*
Affordable living unit	□1	□2	□3	□4
High quality living unit	□1	□2	□3	□4
Living unit on cul-de-sac rather than through street	□1	□2	□3	□4
Good investment potential	□1	□2	□3	□4
High quality K-12 schools	□1	□2	□3	□4
Attractive appearance of neighborhood	□1	□2	□3	□4
Variety in housing styles	□1	□2	□3	□4
High level of upkeep in neighborhood	□1	□2	□3	□4
Large back yards	□1	□2	□3	□4
Lots of off-street parking (garages or driveways)	□1	□2	□3	□4
Sidewalks throughout the neighborhood	□1	□2	□3	□4
Good bicycle routes beyond the neighborhood	□1	□2	□3	□4
Good public transit service (bus or rail)	□1	□2	□3	□4
Easy access to transit stop/station	□1	□2	□3	□4
Parks and open spaces nearby	□1	□2	□3	□4
Shopping areas within walking distance	□1	□2	□3	□4
Easy access to a regional shopping mall	□1	□2	□3	□4
Easy access to downtown	□1	□2	□3	□4
Religious or civic buildings (ex., library) nearby	□1	□2	□3	□4
Close to where I work	□1	□2	□3	□4
Low crime rate within neighborhood	□1	□2	□3	□4
Low level of car traffic on neighborhood streets	□1	□2	□3	□4
Quiet neighborhood	□1	□2	□3	□4
Good street lighting	□1	□2	□3	□4
Safe neighborhood for walking	□1	□2	□3	□4
Safe neighborhood for kids to play outdoors	□1	□2	□3	□4
Lots of interaction among neighbors	□1	□2	□3	□4
Lots of people out and about within the neighborhood	□1	□2	□3	□4
Diverse neighbors in terms of ethnicity, race, and age	□1	□2	□3	□4
Economic level of neighbors similar to my level	□1	□2	□3	□4

Your Daily Travel

The questions in this section ask about **your** daily travel – for example, trips from home to work or to the store.

1. Please tell us about your work/school trip (if you are not employed and not a student, skip to Question 2).

a. How far is it in miles from your residence to your primary place of work/school? ________ miles

b. How long does it usually take to get to your primary place of work/school? ________ minutes

c. Where is your primary workplace/school located?

Street or nearest cross-streets: ____________________ City: ____________________

d. Does your employer/school provide free parking? □1 Yes □0 No

e. In a ***typical week with good weather***, how many days do you use each of the following as your ***primary*** means of transportation between home and work/school? By "primary" we mean the means of transportation you use for the longest portion of your trip.

	Never	*Less than once per month*	*1-3 days per month*	*Once per week*	*2-3 days per week*	*4-5 days per week*
Working at home instead of making the trip	□1	□2	□3	□4	□5	□6
Car, driving alone	□1	□2	□3	□4	□5	□6
Carpool/vanpool	□1	□2	□3	□4	□5	□6
Bus/rail	□1	□2	□3	□4	□5	□6
Walking	□1	□2	□3	□4	□5	□6
Biking	□1	□2	□3	□4	□5	□6
Other (please specify)________________	□1	□2	□3	□4	□5	□6

2. In a ***typical month with good weather***, how often do you ***drive or ride as a passenger in a private vehicle*** from your home to each of the following places ***for purposes other than work/school***?

	Never	*Less than once per month*	*Once or twice per month*	*About once every 2 weeks*	*About once per week*	*Two or more times per week*
A religious or civic building (ex., library)	□1	□2	□3	□4	□5	□6
A service provider (ex., bank, barber)	□1	□2	□3	□4	□5	□6
A store or place to shop	□1	□2	□3	□4	□5	□6
A restaurant or coffee place	□1	□2	□3	□4	□5	□6
A place for entertainment/recreation	□1	□2	□3	□4	□5	□6
A place to exercise (ex., a gym or a park)	□1	□2	□3	□4	□5	□6
To pick up or drop off a passenger	□1	□2	□3	□4	□5	□6

3. In a ***typical month with good weather***, how often do you ***take public transit*** from your home to each of the following places ***for purposes other than work/school***?

	Never	*Less than once per month*	*Once or twice per month*	*About once every 2 weeks*	*About once per week*	*Two or more times per week*
A religious or civic building (ex., library)	□1	□2	□3	□4	□5	□6
A service provider (ex., bank, barber)	□1	□2	□3	□4	□5	□6
A store or place to shop	□1	□2	□3	□4	□5	□6
A restaurant or coffee place	□1	□2	□3	□4	□5	□6
A place for entertainment/recreation	□1	□2	□3	□4	□5	□6
A place to exercise (ex., a gym or a park)	□1	□2	□3	□4	□5	□6
To pick up or drop off a passenger	□1	□2	□3	□4	□5	□6

4. In a ***typical month with good weather***, how often do you ***walk*** from your home to each of the following places ***for purposes other than work/school***?

	Never	*Less than once per month*	*Once or twice per month*	*About once every 2 weeks*	*About once per week*	*Two or more times per week*
A religious or civic building (ex., library)	□1	□2	□3	□4	□5	□6
A service provider (ex., bank, barber)	□1	□2	□3	□4	□5	□6
A store or place to shop	□1	□2	□3	□4	□5	□6
A restaurant or coffee place	□1	□2	□3	□4	□5	□6
A place for entertainment/recreation	□1	□2	□3	□4	□5	□6
A place to exercise (ex., a gym or a park)	□1	□2	□3	□4	□5	□6
To pick up or drop off a passenger	□1	□2	□3	□4	□5	□6

5. In a ***typical month with good weather***, how often do you ***bike*** from your home to each of the following places ***for purposes other than work/school***?

	Never	*Less than once per month*	*Once or twice per month*	*About once every 2 weeks*	*About once per week*	*Two or more times per week*
A religious or civic building (ex., library)	□1	□2	□3	□4	□5	□6
A service provider (ex., bank, barber)	□1	□2	□3	□4	□5	□6
A store or place to shop	□1	□2	□3	□4	□5	□6
A restaurant or coffee place	□1	□2	□3	□4	□5	□6
A place for entertainment/recreation	□1	□2	□3	□4	□5	□6
A place to exercise (ex., a gym or a park)	□1	□2	□3	□4	□5	□6
To pick up or drop off a passenger	□1	□2	□3	□4	□5	□6

6. Approximately how many miles do ***you*** drive in a typical week (including weekends)? _______ miles

7. During **the last 7 days**, on how many days did you **take a walk or a stroll** around your neighborhood - for example, to get exercise or to walk the dog?

0 days	*1 day*	*2 days*	*3 days*	*4 days*	*5 days*	*6 days*	*7 days*
☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$	☐$_6$	☐$_7$	☐$_8$

8. During the last 7 days, on how many days did you **walk** from your residence **to a local store or shopping area**?

0 days	*1 day*	*2 days*	*3 days*	*4 days*	*5 days*	*6 days*	*7 days*
☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$	☐$_6$	☐$_7$	☐$_8$

9a. During the last 7 days, on how many days did you **walk for at least 10 minutes** at a time to go **from place to place**?

0 days	*1 day*	*2 days*	*3 days*	*4 days*	*5 days*	*6 days*	*7 days*
☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$	☐$_6$	☐$_7$	☐$_8$

If 0 days, skip to question 10.

9b. How much time did you usually spend on one of those days **walking** from place to place?

____ hours _____ minutes

10a. During the last 7 days, on how many days did you **bike for at least 10 minutes** at a time to go **from place to place**?

0 days	*1 day*	*2 days*	*3 days*	*4 days*	*5 days*	*6 days*	*7 days*
☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$	☐$_6$	☐$_7$	☐$_7$

If 0 days, skip to question 11.

10b. How much time did you usually spend on one of those days **biking** from place to place?

____ hours _____ minutes

11. For this question, please think about your ***current*** daily travel and your daily travel when you lived at your ***previous residence not long before you moved.*** We would like to know about how your travel has changed, for whatever reason. Please answer for ***your own travel*** only.

a. How much do you ***drive*** now, compared to when you lived at your previous residence?

A lot less now	*A little less now*	*About the same*	*A little more now*	*A lot more now*
☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$

b. How much do you ***use public transit*** (bus or rail) now, compared to when you lived at your previous residence?

A lot less now	*A little less now*	*About the same*	*A little more now*	*A lot more now*
☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$

c. How much do you ***walk in your neighborhood*** now, compared to when you lived at your previous residence?

A lot less now	*A little less now*	*About the same*	*A little more now*	*A lot more now*
☐1	☐2	☐3	☐4	☐5

d. How much do you ***ride your bike*** now, compared to when you lived at your previous residence?

A lot less now	*A little less now*	*About the same*	*A little more now*	*A lot more now*
☐1	☐2	☐3	☐4	☐5

Travel Preferences

We'd like to ask about your preferences with respect to ***daily travel***. Please indicate the extent to which you agree or disagree with each of the following statements on a scale from "strongly disagree" to "strongly agree." There are no right or wrong answers; we want only your true opinions.

	Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*
Traveling by car is safer overall than taking transit	☐1	☐2	☐3	☐4	☐5
I prefer to walk rather than drive whenever possible	☐1	☐2	☐3	☐4	☐5
Walking can sometimes be easier for me than driving	☐1	☐2	☐3	☐4	☐5
Travel time is generally wasted time	☐1	☐2	☐3	☐4	☐5
Traveling by car is safer overall than riding a bicycle	☐1	☐2	☐3	☐4	☐5
I like driving	☐1	☐2	☐3	☐4	☐5
To me, the car is a status symbol	☐1	☐2	☐3	☐4	☐5
I prefer to take transit rather than drive whenever possible	☐1	☐2	☐3	☐4	☐5
The only good thing about traveling is arriving at your destination	☐1	☐2	☐3	☐4	☐5
I like walking	☐1	☐2	☐3	☐4	☐5
It does not matter to me which type of car I drive	☐1	☐2	☐3	☐4	☐5
Public transit can sometimes be easier for me than driving	☐1	☐2	☐3	☐4	☐5
I like riding a bike	☐1	☐2	☐3	☐4	☐5
Traveling by car is safer overall than walking	☐1	☐2	☐3	☐4	☐5
I prefer to bike rather than drive whenever possible	☐1	☐2	☐3	☐4	☐5
To me, the car is nothing more than a convenient way to get around	☐1	☐2	☐3	☐4	☐5
Getting there is half the fun	☐1	☐2	☐3	☐4	☐5
Biking can sometimes be easier for me than driving	☐1	☐2	☐3	☐4	☐5
I like to drive just for fun	☐1	☐2	☐3	☐4	☐5
I like taking transit	☐1	☐2	☐3	☐4	☐5
I feel free and independent if I drive	☐1	☐2	☐3	☐4	☐5

Your Household

The questions in this section ask a few things about you and the members of your household. These characteristics are important for analyzing your choices about where to live and your choices about daily travel. We guarantee the confidentiality of this information and assure you that we will use this information only for analysis purposes.

1. What is your gender? □$_1$ Female □$_0$ Male

2. When were you born? 19__________ Year

3. Please indicate your educational background.

□$_1$ Some grade school or high school	□$_4$ Four-year college degree or technical school degree/certificate
□$_2$ High school diploma	□$_5$ Some graduate school
□$_3$ Some college or technical school	□$_6$ Completed graduate degree(s)

4. What is your current employment status?

□$_1$ Full-time	□$_3$ Non-employed student
□$_2$ Part-time	□$_4$ Not employed (including homemaker, retired, and unemployed)

5. Do you have any physical or anxiety condition that seriously limits or prevents you from …?

a. Driving a vehicle	□$_1$ Yes	□$_0$ No
b. Walking outside the home	□$_1$ Yes	□$_0$ No
c. Riding a bicycle	□$_1$ Yes	□$_0$ No
d. Using public transit	□$_1$ Yes	□$_0$ No

6. Do you have a driver's license? □$_1$ Yes □$_0$ No

7. How many members in your household (including yourself) are licensed drivers? ______

8. How many personal vehicles (cars, SUVs, vans, small trucks, and motorcycles) does your household have?

________ vehicles

9. What are the make, model, and year of the vehicle you drive most often (□$_1$ Not applicable).

	Make	Model	Year
Example	*Ford*	*Focus*	*2010*
Your vehicle	________________________	_________________________	____________

10. How many personal vehicles did your household have **just before you moved**? ________ vehicles

11. How many working bicycles does your household have? ________ bicycles

12. How many working bicycles did your household have **just before you moved**? ________ bicycles

13. Please indicate the number of your current household members (including yourself) falling into the different age groups given below.

_________ persons under 6 years old	_________ persons 18 to 54 years old
_________ persons 6 to 12 years old	_________ persons 55 to 64 years old
_________ persons 13 to 17 years old	_________ persons 65 or more years old

14. Please indicate the number of your household members (including yourself) falling into the different age groups given below, **not long before you moved from your previous residence**.

_________ persons under 6 years old	_________ persons 18 to 54 years old
_________ persons 6 to 12 years old	_________ persons 55 to 64 years old
_________ persons 13 to 17 years old	_________ persons 65 or more years old

15. Do you rent or own your current residence? □$_1$ Rent □$_0$ Own

16. Did you rent or own your previous residence? □$_1$ Rent □$_0$ Own

17. To understand travel choices, and for statistical purposes, we need an idea of your total household income. Please place a check in the box below that it indicates the ***approximate total annual combined*** income of all the working adults in your household.

□$_1$ \$0 to \$14,999	□$_4$ \$35,000 to \$44,999	□$_7$ \$75,000 to \$99,999
□$_2$ \$15,000 to \$24,999	□$_5$ \$45,000 to \$59,999	□$_8$ \$100,000 to \$124,999
□$_3$ \$25,000 to \$34,999	□$_6$ \$60,000 to \$74,999	□$_9$ \$125,000 or more

18. We also need to know your total household income ***not long before you moved*** from your previous residence. Please indicate whether ***approximate total annual combined*** income of all the working adults in your household has increased, decreased, or stayed about the same.

Decreased by				Stayed about the same	Increased by			
\$17,500 or more	\$12,500 to \$17,499	\$7,500 to \$12,499	\$2,500 to \$7,499		\$2,500 to \$7,499	\$7,500 to \$12,499	\$12,500 to \$17,499	\$17,500 or more
□$_1$	□$_2$	□$_3$	□$_4$	□$_5$	□$_6$	□$_7$	□$_8$	□$_9$

Optional: If you would like to enter the drawing for one of the ten gift card prizes of \$50 each, please provide the following contact information. Providing this information is entirely optional.

Daytime phone number: ___________________________________

E-mail address: ____

Is it OK for us to contact you if we have questions about your survey? □$_1$ Yes □$_0$ No

Is there anything else you'd like to tell us regarding your choices about where to live and your choices about daily travel? Please provide comments on the next page, and feel free to attach additional pages if you'd like.