BASIC HUMAN DECISION MAKING: An Analysis of Route Choice Decisions by Long-Haul Truckers

John Holland Knorring
Advisor: Professor Alain L. Kornhauser
April 14, 2003

Submitted in Partial Fulfillment of the requirements for the degree of Bachelor of Science in Engineering Department of Operations Research and Financial Engineering Engineering and Management Systems Program Program in Finance Princeton University
I hereby declare that I am the sole author of this thesis.

I authorize Princeton University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

_______________________
John H. Knorring

I further authorize Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

_______________________
John H. Knorring
To my Mother and Father
Acknowledgements

I would like to thank Professor Alain Kornhauser for his help and guidance throughout this entire thesis process. His unending support and insight made this all possible.

I would also like to thank Rachel He. Without her help in working with the data in the TIDE lab, none of this would have been possible.

Lastly, I would like to thank Andrew C. Redman ’02 for his help and support of this project.
INTRODUCTION

1.1 Travel Demand Forecasting

1.1.1 Trip Generation

1.1.2 Trip Distribution

1.1.3 Modal Split

1.1.4 Trip Assignment

1.2 Treatment of Congestion in the Four-Step Approach

1.2.1 Costs of Congestion

1.3 Thesis Contents

REVIEW OF LITERATURE

2.1 Historical Perspective of Travel Demand Modeling

2.2 Basic Human Decision-Making: Utility Maximization

2.2.1 Completeness

2.2.2 Transitivity

2.2.3 Continuity

2.3 Quantifying Utility: UTILS

2.4 The Truck Driver Utility Function

2.5 Factors in the Utility Function

2.5.1 Income and Education

2.5.2 Number of Alternate Routes Considered

2.5.3 Risk Aversion

2.5.4 Traffic Information

2.5.5 Time of Day

2.5.6 Congestion

2.5.7 Additional Factors

2.6 Work Done by Todd Burner ‘99

2.6.1 Value of work done by Todd Burner

2.6.2 Drawbacks to work performed by Todd Burner ‘99

2.7 Prospect Theory

2.7.1 Profile of Kahneman and Tversky

2.7.2 Forecasting the Future

2.7.3 Asymmetry of Risk

2.7.4 Ambiguity Aversion

2.7.5 Prospect Theory as it Relates to this Study

2.8 Other Scholarly Truck Specific Studies

EXPLANATION OF THE DATA SET
3.1 COLLECTION OF DATA .............................................................. 52

  3.1.1 Time of Arrival Ranging ...................................................... 52
  3.1.2 Loran .................................................................................. 56
  3.1.3 Global Positioning System .................................................... 57
  3.1.4 Inertial Navigation Systems ................................................. 60
  3.1.5 General Comments on Accuracy of Navigation Equipment ....... 62

3.2 COMPLICATIONS AND DIFFICULTIES WITH DATA EXTRACTION ....... 63

  3.2.1 Origin-Destination Information .............................................. 63
  3.2.2 Non-Random Sample of Trucks ............................................ 64
  3.2.3 Random Time Spacing of Observations .................................. 64
  3.2.4 Observation Reporting Error and Map Matching Complications ...... 65
  3.2.5 Stop Determination ............................................................... 66

3.3 ADDITIONAL DATA CONSIDERATIONS ....................................... 67

  3.3.1 Observed Traffic on Routes .................................................. 67
  3.3.2 Socioeconomic Factors ........................................................ 67
  3.3.3 Characteristics of Specific Trucks ........................................ 68

4 METHODOLOGY ............................................................................ 69

  4.1 STOP DETERMINATION ............................................................. 69
  4.2 DETERMINATION OF REGIONS TO BE EXAMINED .................... 70
  4.3 TRIP DETERMINATION ............................................................. 72

    4.3.1 Route Determination Heuristic ............................................ 73
    4.3.2 The Box Algorithm ............................................................. 73
    4.3.3 Drawbacks Associated with the Box Algorithm ....................... 77
  4.4 FURTHER DATA REDUCTION USING MINIMUM AND MAXIMUM AVERAGE SPEEDS 79

5 BUILDING THE MODEL .................................................................. 81

  5.1 FOUNDATION FOR DISCRETE CHOICE MODELS ......................... 82
  5.2 UTILITY MAXIMIZATION .......................................................... 83
  5.3 THE LOGIT MODEL ................................................................. 84
  5.4 DERIVATION OF PERCEIVED SPEEDS ....................................... 90

6 COMMENTARY ON RESULTS .......................................................... 92

  6.1 INTERPRETATION OF PERCEIVED SPEED CURVES ................. 94
  6.2 POINTS OF INDIFFERENCE ...................................................... 95
  6.3 INTERPRETATION OF THE PARAMETERS .................................. 97
  6.4 PITFALLS IN THE ANALYSIS .................................................... 97

    6.4.1 Small Sample Size for Specific Cases .................................... 97
6.4.2 Infrequency of Data Collection ................................................................. 98
6.4.3 Inaccuracy of Stop Definition ...................................................................... 99
6.4.4 Problems Associated with the Box Algorithm ........................................ 99
6.4.5 Overall Impact on the Model ................................................................. 100

7 CONCLUSION ........................................................................................................ 101

7.1 Interpretation of the Results ........................................................................ 102
7.2 Areas for Further Study ............................................................................... 103

APPENDIX 1: SUMMARY STATISTICS OF TRUCKS ON ROUTES: FIRST
ATTEMPT ............................................................................................................... 107

APPENDIX 2: SUMMARY STATISTICS OF TRUCKS ON ROUTES: SECOND
ATTEMPT (POST DATA REDUCTION) .............................................................. 111

APPENDIX 3: C CODE FOR DATA ANALYSIS ................................................. 114

APPENDIX 4: GPS BOX COORDINATES ............................................................ 129

APPENDIX 5: ROUTE MAPS .................................................................................. 132

7.2.1 90-94 ......................................................................................................... 132
7.2.2 Cincinnati, OH ......................................................................................... 133
7.2.3 Columbus, OH ......................................................................................... 134
7.2.4 Houston, TX ............................................................................................ 135
7.2.5 Indianapolis, IN ......................................................................................... 136
7.2.6 Memphis, TN ........................................................................................... 137
7.2.7 Nashville, TN ........................................................................................... 138
7.2.8 Oklahoma City, OK .................................................................................. 139
7.2.9 Richmond, VA .......................................................................................... 140
7.2.10 San Antonio, TX ...................................................................................... 141
7.2.11 St. Louis, MO .......................................................................................... 142
7.2.12 Wilmington, DE ....................................................................................... 143
1 Introduction

Each and every day, humans make decisions. Though most decisions (such as breathing and blinking) are made subconsciously, humans often take an active role in the decision making process. Humans decide what shoes they are going to wear, what radio station they are going to listen to, and how they are going to get to work each day just to name a few. The processes that are involved in decision-making are quite complex. This thesis is will shed light on some of the more interesting decision-making processes that long-haul truckers utilize when making route choice decisions.

Every day of the year, freight and goods need to be transported around the United States. One of the most popular methods of hauling freight is via trucks. Trucks, in conjunction with a driver, are able to navigate the U.S. highway network to bring goods from where they are supplied to where they are demanded. While this process might seem quite simple, the trucker drivers are constantly faced with decisions. Should he speed up slow down? Should he change lanes or continue on his current path? If he is coming up to a city, should he stay on the current highway and potentially face congestion, or should he change lanes so that he can get on a bypass route around the city that will hopefully save him time? This last question is one that this thesis will explore.

The goal of this thesis is to perform a detailed empirical analysis of fundamental human decision-making processes by examining the behavior of truck drivers as they navigate the United States interstate highway network and to analyze the route choice decisions that they make when a number of plausible alternate routes are available. This thesis uses logistic models to describe route choice behavior when truck drivers are faced
with alternate routes. Previous empirical analyses of this nature have been plagued by data quality problems relating to objectivity issues as well as data collection methods. However, this thesis relies on data that comes from an objective, remotely sensed revealed-preference data set consisting of over 249,465 trucks over a thirteen-day period. Because of this, this study will be able to truly perform a detailed empirical analysis. Once the empirical analysis is completed, this thesis will then use the results of the logistic models and analyze the decisions made by drivers to determine perceived speeds on alternate routes as well as analyze the rationality of decision-making exhibited by the long haul truck drivers. An analysis will then be done to determine how accurately the logistic models predict driver behavior and how to more accurately capture driver preferences and risk aversion. The real value of this thesis comes from the vast data set that will allow thorough enough analysis to finally compare theoretical conjectures to real world results.

One set of locations that has been a popular target for other studies is major cities that have at least two alternatives for traveling through a city. Many cities usually have a route that passes through the downtown area and a route that circumvents the downtown area, oftentimes referred to as a bypass route. When a driver is nearing such a city he is faced with a decision to make. Should he take the downtown route or the bypass route? Each route has its strengths and weaknesses. The downtown route is oftentimes shorter than the bypass route; however, the downtown route usually has a slower speed than the bypass route due to increased traffic volume or reduced traffic capacity relative to demand on the downtown route. In addition, posted speed limits for traffic in the central business district are generally less on the downtown route than on the complimentary
bypass routes. The end result is that under light to normal traffic conditions on the downtown route, the average time spent is lower than on the bypass route, however, on the whole, the variance in travel times is greater on the downtown route than the bypass. By investigating the decisions made by the truck drivers when they are faced with a similarly, one can begin to piece together the driver’s preferences for distance and time in relation to their risk aversion.

In addition to cities that have alternate routes, due to the size and complexity of the interstate highway network, there are a number of alternate routes between cities. Interstates 90 and 94, two of the most heavily traveled truck routes, is an example of one such area where truckers have to make a decision on which route to take. It is not obvious why every trucker would not always take the shorter route. Both routes cross areas of the northern plains states that have very little problems with congestion. This thesis, given its large data set, will attempt to explain why many truckers choose a longer route.

1.1 Travel Demand Forecasting

Many would argue that present economic activity, as well as future economic growth, in any region is highly dependent on the ability to move goods and services around in an efficient manner. Transportation infrastructure plays an important role in the movement of goods. As a result, it is very important that the links in the transportation network are able to handle not only present load factors in a manner that minimizes congestion costs, but also future load demands on the network. However, it is
also important to note that building transportation networks is very costly and time consuming, so building too much capacity into a system should also be avoided. Coming up with forecasts of future link load demands is a very complex process that includes factors such as future population growth, increases in the number of links, as well as alternative transportation methods. Taking these factors into account, planners then need to be able to understand how individuals make routing decisions in order to optimally project future link load demands. The better planners understand route choice decisions, the better their optimization models (and thus solutions) will be, resulting in a more efficient flow of goods and services in the future.

In coming up with projections of future link load demand, planners need to have a common basic methodology to solve for choices that representative travelers will make when faced with travel alternatives while also being limited by budgetary constraints. This methodology needs to be able to determine accurate estimates for the numbers of trips on the transportation network in addition to being able to determine how the trips will be distributed throughout the network links. The most common method used in textbooks and other similar scholarly work is the four-step travel demand process.

These four steps consist of:¹

1. Trip Generation
2. Trip Distribution
3. Modal Split
4. Trip Assignment

¹ The following explanation of the four step travel demand process is adopted from Norbert Oppenheim’s Urban Travel Demand Modeling: From Individual Choices to General Equilibrium and from Moshe Ben-Akiva’s Structure of Urban Travel Demand Models
This four-step process is commonly used in general traffic planning assessments, and it works well for our database of truck traffic behavior.

In modeling travel demand it is useful to break the entire network into smaller areas that are easier to analyze than the network as a whole. However, it is important though to keep in mind that breaking the network down into areas that are too small will provide little benefit. These smaller, more workable zones will be the basis for this study. The four-step travel demand process breaks each zone in the network into either an origin/source zone or a destination/sink zone. This thesis however, will use a slight variation of this process in that it will also include transit zones that will be used to identify the particular routes trucks take either through the central business district, or via the bypass route. These zones will be instrumental in coming up with workable solutions for the four-step travel demand process.

1.1.1 Trip Generation:

The first step in the travel demand process is trip generation. This step is designed to generate the number of trips that originate from a given source zone. One small pitfall to this step is that it says nothing about where the trips will be going. Planners can calculate and project numbers of trips generated from a zone via a few different methods. The first method for calculating the number of trips generated is to use observed zonal trip rates and extrapolate those numbers onto other similar zones. While this method is somewhat crude, it does provide a fairly robust solution. In terms of projecting future link demands, a simple growth factor can be applied. The second method being used today, and growing in popularity, is the econometric modeling
method. This method is quite similar to multi-factor modeling and regression commonly
used in modern time series analysis. One example of a model might be:

$$T_{(\text{trips})} = \text{Constant} + 4.5X_1 + 1.05X_2 + 0.076X_3 - 2.036X_4$$

Where $X_1$ might be the average population density in the zone, $X_2$ might be the distance
to the central business district, $X_3$ might be the average household income, and, $X_4$ might
be a seasonality component. While this method has some obvious advantages in that it is
much more descriptive of the representative zone, the complexity of the analysis
increases enormously when considering a network as large as the United States interstate
highway system.

It is important to keep in mind that the trip generation phase is fairly limited. It
only produces characteristics of the respective zone and traveler attributes. All
characteristics relating to trip destinations, mode choice, and route choice are determined
in later stages. The easiest way to conceptualize the trip generation stage is to think of
the zones and then determine how many trips will be leaving the respective zone.

![Figure 1-1](image)

1.1.2 Trip Distribution:

The trip distribution phase of the travel demand forecasting process is a one-way
stage that is typically executed after the trip generation stage. The inputs to this stage are
the outputs from the trip generation phase. Trip distribution typically finds and allocates
a destination for each of the trips generated at source nodes. The most common method for determining trip distribution is the “Gravity Method.” This method assesses the accessibility and attractiveness of destination zones for each source zone and distributes proportionally the trips from the trip source zones to the destination zones.

1.1.3 Modal Split:

The next stage that transportation planners follow is the modal split stage. This stage determines the mode split that the representative travelers will use to get from their respective origins to their destinations. In other words, this stage splits the travelers into groups that will be driving cars, other groups that take busses, groups that take trains, and groups that fly, for example. Once the planner has ascertained the appropriate levels of modal splits, he can then determine the number of trips taken by road from one region to another and thus plan accordingly. Modal splits are determined similarly to trip distribution. Typically a planner will use a diversionary curve or S-shaped logistic curve to determine modal split.

1.1.4 Trip Assignment:

Trip assignment is the fourth step in the traditional four-step approach. This is the stage where planners attempt to determine the routes that representative travelers will take between their respective origin-destination pair. This phase will show how all of the travelers distribute themselves over the links in the network. There are many ways to derive trip assignments. Some models use a shortest path constraint, however this is not
always the optimal solution. A common constraint found in most models is that no single
traveler should be able to reduce his travel time by switching to another series of links.
This constraint is known as the Wardrop user equilibrium condition. The result of this
condition is that the model must be able to find alternate paths that a driver can take, and
it must be able to determine if any travelers will take the alternate route. In other words,
the model must examine routes that are longer than the shortest path solution and
determine if the time spent on the alternative route is less than the time spent on the
shortest path route.

1.2 Treatment of congestion in the four-step approach:

In the four-step approach to modeling travel demands, congestion rarely is a
factor in any calculations. The end result of this is that the final solution to the four-step
process might not be optimal for all situations. Congestion becomes increasingly
important in factors where link load costs and destinations are functions of link demands.
The most obvious example of this is when link loads approach capacity and traffic rates
are slowed down significantly. The result is the time spent traversing a link increases to a
level where the benefit to following that link would be decreased significantly, thus
resulting in a reduced demand for the specified link. The end result is a vicious feedback
loop that is very difficult to model and the four-step process becomes a highly complex
iterative algorithm. Stability in the models is oftentimes difficult to achieve with
congestion.
1.2.1 Costs of Congestion

When one thinks about the last time that he was in a traffic jam, most likely bad memories come to mind due to the unnecessary costs involved with sitting in traffic. Papacostas and Prevedouros assert that there are a number of costs associated with congestion including: loss of productive time, increased fuel usage, increased levels of atmospheric pollution, increased engine and mechanical wear, increased frequency of traffic accidents, and the negative psychological and emotional impact of sitting in traffic culminating in “road rage.”

In order to minimize these negative byproducts of congestion, the government has introduced legislation and has created guidelines for monitoring efficiency in the highway. The first step in this process began with the passage of the Federal-Aid Highway Act of 1962. According to the U.S. Department of Transportation’s Travel Model Improvement Program, this act was designed to improve the U.S highway network so that users would have access in all weather conditions. Additionally, the act stipulates that cities and urban areas with populations of 50,000 or more must provide a long-range plan for their respective area of the transportation network. More recently, the Transportation Efficiency Act of 1998 (TEA-21), among others, has put the responsibility for long range planning into the hands of the locally established Metropolitan Planning Organizations (MPO) established throughout the United States in areas with populations greater than 50,000 in an effort to increase flexibility in the design and implementation process.

---

3 http://tmip.fhwa.dot.gov/clearinghouse/docs/utp/
1.3 Thesis Contents

This thesis consists of an analysis of truck driver decision-making gathered from a large revealed-preference data set consisting of over 60,000,000 location records. The analysis will be done to determine the effects of availability of alternate routes in decision-making processes.

Chapter II consists of an analysis of current scholarly work in the area of route choice. The chapter will focus on other studies and the current theory on truck driver behavior. This section will contain a historical perspective on route choice modeling and basic decision modeling along with the challenges others faced with empirical studies. This section will also include a discussion of Daniel Kahneman’s Nobel Prize winning Prospect Theory. Because this thesis is primarily focused on building on the work that Todd Burner ’99 did for his senior thesis, there is an extensive review of his thesis included in this section.

Chapter III consists of the data that is used in this analysis. This section will include commentary on the size and relevance of the data set along with a discussion on the importance of the objectivity and remotely censored data. In addition, methods of collecting the data, basic formatting of the data, and derived characteristics of the data are discussed. Many of the definitions to be used throughout this thesis as well as drawbacks to the data set will also be covered in this chapter.

Chapter IV consists of the methodology for extracting the data. This section includes the summary statistics of the number of trips and number of observations. In
addition, this section contains analysis on average speeds maintained by the truckers. Lastly, this section covers the algorithm used for selecting the appropriate trucks.

Chapter V consists of the building of the Logit Model. This includes remarks on notable statistics as well as how assumptions in the definitions and calculations may affect the resulting summary statistics.

Chapter VI consists of commentary on model fitting. This section will cover how logistic models are used and what can be gained from their use. Additionally, a number of case studies are examined using the results of the logistic models.

Chapter VII consists of concluding comments about the results of the study along with a section on areas that deserve further study.

In addition to eight chapters, there are two appendices. The first appendix is a central grouping of the output data, models, and statistics. The second appendix is a more in-depth look, including road maps, at the case studies covered in this thesis.
2 Review of Literature

Any scholarly research must begin with a review of related work done by others. It is necessary to review the work of other individuals so that value can be added and research is not simply repeated. First, background information on the current theories, for the sake of the reader, must be examined in order to set a workable foundation for critical analysis. As stated previously, this thesis is primarily an extension of work done by Todd Burner ’99. Therefore, it is essential to review the work that he did and access what is of value in his work to this thesis. Additionally, it is important to seek out areas where further analysis needs to be done to better set a foundation for this thesis.

2.1 Historical perspective of travel demand modeling

There has been much scholarly work done concerning travel demands and individual driver route choice decision-making. Most agree that modern travel demand analysis began in the late 1930’s as concern with building and planning for new roads increased in lock step with the increasing number of automobiles. Most of the early work was fairly limited and more focused on theoretical considerations due to the limited computing resources available. The focus of the early work was also on more aggregate levels of travel demand than on disaggregate individual behavior modeling.

After sorting through the vast amount of research work done on traffic demand modeling, one name that keeps reappearing in most of the research on both aggregate and disaggregate travel demand and behavior modeling is professor of Civil and
Environmental Engineering at the Massachusetts Institute of Technology, Moshe Ben-Akiva. Professor Ben-Akiva has published a wide array of work in the field of transportation. One of his more pertinent studies relating to this thesis is his doctoral thesis *Structure of Passenger Travel Demand Models*. In his thesis, he covers a wide range of modeling options for transportation demand. The two primary models that he uses are logit models and probit models. The logit models are simple logistic models that attempt to predict aggregate probabilistic behavior of populations. The probit model that he focuses on is much more complex.

In probit models, joint probabilities are used to model choices. These models try to associate a probability to every decision combination that could be made from the time one decides to travel until the time an individual reaches his destination. Because these models attempt to quantify all decisions, the computation complexity is increased by many factors.\(^4\)

One of the major weaknesses of the Ben-Akiva’s method is that his empirical study is based on household surveys. His study modeled modal choice by individuals when they go shopping. Household surveys do not accurately reflect human decision making for a number of reasons. First, there are data quality issues that arise when surveys are given. The surveyed individual can lie on the survey, or at least deliberately misinform the surveyor. The surveyed individual also might not know exactly which mode they took at some time in the past and they might guess incorrectly which method they took. Next, by surveying households, one has the issue that by observing the system, one may have disturbed the system. In other words, a surveyed individual might change or have changed their behavior because of the data collection that was being

done. However, given the level of technology for collecting arms length data at that time, Ben-Akiva’s study provides a useful base of study for this thesis. Going forward with this thesis, access to remotely sensed, arms length data will allow for a more accurate analysis to be performed.

2.2 Basic human decision-making: Utility maximization

Every day individuals are faced with a series of decisions to make on how to allocate resources. One of the most interesting resources that people allocate is time. The concept of time is very interesting because of its non-fungible nature; each person gets a fixed number of hours every day and that amount of time is the same for every person. It is impossible to trade hours with another user, and it is also impossible to change the rate at which the resource is used. While some associate time allocation with a simple newsvendor problem with costs of overage and costs of underage, in the more general case of decision-making, time allocation really is dependent on utility maximization. In the context of transportation, time is especially interesting because of the fundamental nature of time in evaluating destination, mode, and route alternatives.5

Most scholarly quantitative work regarding decision making for travel demand forecasting, transportation planning, and congestion management is based on the concept of utility maximization. The process of making route choice decisions is quite complex and there are a number of factors that are hard to quantify. Utility maximization is used to study travel demand forecasting because it is the most plausible method of evaluation. It

is the most efficient way to account for as many of the factors that are important for making routing decisions as possible. A brief list of these decision factors includes such things as availability of alternate routes, length of alternate routes, perceived speeds on alternate routes, anticipated congestion, weather on alternate routes, scenery encountered, and hazards avoided. While this list is only a brief look at the factors involved with decision-making, it sheds light on the complexity of route choice decision making. When making decisions about planning for future use, designers look to optimize the usefulness or utility of the system. They look to get the most benefit from the least amount of input effort. This concept of cost/benefit analysis is the central tenet of utility maximization.

A good place to start when talking about utility maximization and preferences is to make a few definitions. The following properties are also the same definitions used to define rational decision-making. First, utility needs to be defined. Most economists find it easiest to define utility in terms of its characteristic properties.

The properties for Utility Maximization are as follows:\(^6\)

- Completeness
- Transitivity
- Continuity

2.2.1 Completeness

The first property that most refer to is the completeness property. This property is used to rank options available. For any two options available, \(A\) and \(B\), an individual can always select one of the three following statements to be true. Either:

---

1. “A is preferred to B”
2. “B is preferred to A”
3. “A and B are equally attractive.”

Given these three options to select from, the result is that the individual will always have exactly one option to select from. Given these statements, it is understood that an individual can always completely comprehend and make a decision about the attractiveness of two alternatives. The completeness assumption also eliminates the possibility that an individual can report that A is preferred to B and B is preferred to A. An easy way to conceptualize this concept is to think about two options. Option A is you receive $20. Option B is you die. Most decision makers would prefer decision A to decision B. The completeness assumption states that, all other things being equal, if you choose A over B, you will never choose dying over receiving $20. This assumption: “all other things being equal”, as we will see later, is very important as it relates to route choice decision-making.\(^7\)

### 2.2.2 Transitivity

The second property of preference relation is transitivity. The transitivity property is used when ranking three or more choices. This property states, “If an individual reports that ‘A is preferred to B’ and that ‘B is preferred to C,’ then he or she must also report that ‘A is preferred to C.’” This property strengthens the consistency argument in decision-making. In other words, individuals will act consistently when

\(^7\) Ibid.
faced with a decision between A and C. Since they already know that they prefer B to C, and A is preferred to B, the decision is essentially already made for the individual.  

2.2.3 Continuity

The last property of preference relation is continuity. This property states that “If an individual reports ‘A is preferred to B,’ then situations suitably ‘close to’ A must also be preferred to B.” This assumption is used when it is necessary to look at changes in responses due to a marginal change in the given information.

2.3 Quantifying Utility: Utils

Given the three properties of preference relation: completeness, transitivity, and continuity, it is possible to prove that decision makers can rank all possible decisions available to them from least attractive to most attractive. This ranking of prospects defines utility. More formally, if a persons favors A to B we would say that the utility afforded A, written as U(A), is greater than the utility afforded B, U(B). Utility however is more complex than just ranking prospects. It is possible to assign numbers to utility rankings. However, the numbers that are attached to an option or prospect do not signify a specific value judgment in terms of the unit of utility, utils. This assumption is referred to as the Non-uniqueness of Utility Measures. In other words, if U(A) = 50 and U(B) = 10, this does not mean that A is preferred to B at a rate of 5 times that of B.

---

8 Ibid.
9 Ibid.
10 Ibid.
11 Ibid.
Rather, these numbers are used to imply that A is simply preferred to B. The U(A) could have been equal to 1000 and the U(B) could have been equal to 1 and there would be no loss of significance to either statement. As a result, it does not make sense to ask, “How much more does the decision maker prefer A to B?” Rather, it is only possible to declare that A is preferred to B. One interesting result of the non-uniqueness principle is that now it is not possible to compare utilities of different individual decision makers because the value of utils for one person is not transferable.\(^\text{12}\)

2.4 The Truck Driver Utility Function

Having covered the concept of utility, it is now important to cover how utility relates to truck drivers as a decision making class. Traditional economic theory suggests that the summation of factors involved in the decision making process can be gathered in a single output utility function. Rankings and preferences can then be determined by examining the output values of several utility functions. These output values are compared using the three preference relation properties of utility maximization to come up with a final ranking of preferences. Each individual decision maker’s utility function is composed of a number of factors. Each of these factors carries a different weight in the utility function based on the decision maker’s preferences. More formally:\(^\text{13}\)

\(^\text{12}\) Ibid.

\[ U = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + \ldots + a_nX_n \]

Where:
- \( U \) = the utility of the decision
- \( a_0 \) = the calibrated constant that embodies the “unmodeled” aspects of the utility function
- \( a_i \) = the calibrated \( i^{th} \) coefficient
- \( X_i \) = the \( i^{th} \) utility attribute

Given this utility function, and also assuming that the given truck drivers are utility maximizers, it is now possible to model the truck drivers preferences. It is generally assumed that it is not possible to view each truck driver’s utility function directly, however, much work has been done in the past to suggest factors that contribute to the decision making process and thus the utility function.

2.5 Factors in the Utility Function

Now that the utility function has been defined, it is necessary to then determine what all of the \( a_i \) and \( X_i \) factors are. In general, the process for coming up with such values is done in one of two ways. The first method, while significantly easier to do, is not as precise and does not always capture all factors well enough. This method is referred to as the stated preference method. Essentially, this method uses surveys of decision makers to determine what their preferences are. A typical question involves figuring out what the individual’s aversion to risk by posing a real world tradeoff. The typical question asked in many psychological studies is: “Would you rather get $100
right now, or take a gamble where there is a 50% probability of getting $0 and a 50% probability of getting $200?" From a probabilistic and expected value perspective, these options are equal. From this question one can determine if someone is risk neutral, risk averse, or risk seeking via the completeness property of utility maximization. Also, by adjusting the probabilities and amounts, one can begin to determine a decision maker's risk aversion curve. Taken in the context of truck drivers and the real world options available to them, they will be asked if they would prefer to take a route that will take them 1 hour to cover every day, or to take a route that is 50 minutes with a probability of four out of five days, and 1 hour and 15 minutes on one out of five days. In the first case, the driver knows with certainty how long the trip is going to take him. In the second case, the driver has to choose to gamble that when he drives the second route that it will hopefully be faster. He knows that on the average choosing the second route will save him five minutes, but that the standard deviation of the trip time is over 10 minutes. This is a tradeoff of shorter times with greater variance versus longer times with certainty. This topic will be covered again in greater detail later in this thesis.

While relatively easy to do, there are significant drawbacks to the stated preference method. First, using a stated preference method requires that the surveyed individual be able to recount past actions with a very high level of detail. For most individuals this is hard to do. Next, the stated preference method requires that the decision maker be totally honest with his responses to questions. This is also hard to accomplish, for two reasons. First, the decision maker may not have known exactly what happened and is unable to entirely recount the past accurately, or it is possible that the decision maker knows what happened, but for some reason chooses to answer the survey.

14 Ibid.
in a non truthful manner. Lastly, the stated preference method is not nearly as scientific as the revealed preference method as it does not replicate the actual choice process that individuals face when they are driving.  

For this thesis, the revealed preference method will be used. Having a revealed preference data set means that we have data on what people actually have done, as opposed to what they think they did or what they think they would prefer to do. In general, the data for these types of studies is collected via some sort of third party observer. In collecting this data, it is important to try very hard to influence the system as little as possible in the data collection. This idea of not disturbing the system is very important because if the data collection influences the choices or behavior of the studied group, then the value of the data is significantly reduced. While harder to work with, a revealed preference data set is generally considered more valuable. As was alluded to earlier, revealed preference data sets reveal decisions that were made. Stated preference data sets allow for individuals to state their preferences. When formulated as a question, many individuals do not always state their preferences to be the same as the preferences that they demonstrate by their actions.

To come up with the $X_i$ factors for the utility function using a revealed preference data set, the analyzer suggests a possible factor that could contribute to the utility of the decision maker. Then, an analysis is done to see if this factor is relevant to the utility function. If it is relevant, it will be included in the next step of analysis. After the number of different possible $X_i$ factors has been exhausted, the analyzer moves to the next stage of analysis. He now needs to determine the level of contribution that each $X_i$ factor has to the overall utility. To accomplish this, the analyzer gathers a number of

\[15\text{ Ibid. pg. 22}\]
decisions that the decision maker has made and also looks at alternatives that were available. A regression is done to estimate the contribution of each factor to the overall utility function. The results are seen as the $a_i$ factors. As simple as this sounds, it is actually very difficult to determine all of the $X_i$ factors and their respective magnitudes of contributions to the utility function. Luckily though, there has already been much research on passenger vehicle route choice decision making to suggest some possible decision factors in the utility function. This thesis will assume that a portion of the error term is explained by non-considered decision factors.

In coming up with possible factors that would effect the decision making process for truckers, this study will start with factors that have been shown to be significant contributors in the utility function of passenger vehicles. While passenger vehicles are not expected to have utility functions that are identical to long haul truckers because of the different nature of their trips, it is safe to assume that their utility functions will be at least similar to long haul truckers. This thesis will use generally accepted passenger vehicle decision factors as a base of study for the truck driver analysis.

### 2.5.1 Income and Education

In looking at decision makers, individuals that have a higher household income have been shown by Abdel-Aty et al. to consider a larger number of alternate routes.\(^\text{16}\) Household income is an important factor in route choice determination because as income increases, drivers are willing to consider more possible routes in their analysis of

---

picking routes. A brief example of this might be that higher income individuals are more likely to consider taking expensive toll roads. While it is important to note that higher household income is correlated to a larger number of alternate routes considered, whether or not higher income is causing more routes to be considered is a much more important question. Additionally, Khattak et al. showed that higher income drivers not only considered more alternate routes, but they also were more likely to use the alternate routes. Some have suggested that the correlation between income and route choice is a result of higher income drivers placing a higher value on their time, and thus they seek out more alternate routes that could potentially save them time. While this relationship has not been proven entirely, there is a strong correlation between the two.

In addition to income, the level of education that the driver has obtained is strongly correlated with the number of alternate routes considered and selected. Abdel-Aty et al. were able to show through their research that people’s usage of alternate routes increased as the level of education increased. This result is not too surprising for two reasons. First and most obvious, greater levels of income have traditionally been associated with higher levels of education. Second, higher levels of education are strongly correlated with more intelligent individuals. Intelligent individuals have a higher propensity to think through the route that they will take and as a result will consider multiple routes.

---


2.5.2 Number of Alternate Routes Considered

Stephanedes et al. were able to show that in general, drivers rarely consider more than two alternate routes, and when they do consider a third route it is usually only under extreme circumstance such as weather. In their stated preference study of commuters in Minneapolis and St. Paul Minnesota, fewer than 3% of the commuters considered a third alternative route. These results would strongly suggest that passenger vehicle drivers essentially pick between one of two routes considered. Assuming that long haul truck drivers’ cognition is the same as commuter traffic, we will use this principle of limited alternate route considerations as a basis for this thesis by only considering the downtown route and one bypass route.

When considering alternate routes there are basically only two ways that this is possible for drivers. The first method is by looking at a road map. This is the primary method used for picking routes for individuals that are not familiar with the roads around them. If an individual has a road map in front of them, they have every single alternative route possible; however, most of the routes are discarded in order to reduce cognitive overload. In general they are left with a few practical routes to choose from. From there, the individual uses his own utility function and optimizes his route by maximizing his utility. The second way in which people decide on alternate routes is by use of detailed cognitive maps that an individual may have built up over time. The problem with this method is that individuals will only have detailed cognitive maps of their surroundings if and only if they are quite familiar with the roads and have driven on them for an extended period of time. When drivers go into unfamiliar areas, they are unlikely to have the

necessary cognitive maps to decide on routes and are left to either find a map or guess where the road that they are on leads.

2.5.3 Risk Aversion

One of the most important factors in choosing routes as it relates to this thesis is how the individual decision makers respond to risk and their willingness to take risks. When a driver becomes delayed on his selected route, he will, in general, attempt to determine how long he will be delayed based on his own observations of the traffic flow around his vehicle. The driver basically has two options. First, he can stay on the route and just build the delay into his expected time of arrival. His second option is that he can begin to calculate an alternate route that will avoid the road ahead of him. One disadvantage of switching to another route is that the driver most likely has no information about the traffic flow on the alternate route. In fact, the only way that he would have information about the alternate route is if he was told by an outside agent i.e. radio or another driver as to the conditions on other roadways. As a result, the variance in the amount of time that it will take to travel the whole route will be greater.

When talking about individuals risk preference and risk behavior, the concept of variance comes into account. In this case with the driver stuck in traffic, if he switches routes, there is a chance that the new route is congested as well and he will spend an equal or greater amount of time stuck on the alternate route. However, the driver does not know the answer to this question before he makes his decision to switch or to stay. One can only determine the likelihood or propensity that a driver will switch to alternate

---

20 Ibid.
routes given his willingness to take risks. By gambling on an alternate route when the
driver already has a good idea of the time that it will take him on his current route given
the level of congestion, the driver is showing that he has at least a fairly healthy appetite
for risk. Khattak et al modeled this concept of risk behavior in drivers. They were able
to show that a driver’s “inclination to adventure and discovery” does influence route
choice decisions.\textsuperscript{21} As a result, for any study of this nature, it will be important to have
an understanding of the risk preferences of the individuals involved in the study.
Obtaining individuals preference for risk is very difficult to do. It is almost impossible to
determine from a revealed preference data set as revealed preference data sets only show
the decisions people actually make and not why they make them or what factors they
took into account to formulate a decision. Risk preferences are determined through a
series of very specific questions similar to the question posed earlier about taking a
gamble on one outcome that varies in quantity or choosing to take another option that has
a reduced outcome, but has greater certainty in the outcome. The data set for this thesis
does not include such questions, so it will be very difficult to model the behavior of
drivers on an individual basis. However, it is possible in some cases to model aggregate
levels of risk preference for groups of drivers based on their choice to use toll type routes.

2.5.4 Traffic Information

Traffic information can come from a number of different sources: personal
observations, radio, television, roadside displays, CB radio, cellular phone and even the

\textsuperscript{21} Khattak, Asad et al. Factors Influencing Commuters’ En Route Diversion Behavior in Response to
Delay. In \textit{Transportation Research Record 1318}. TRB, National Research Council, Washington, D.C.,
1991 pg 143.
Khattak et al. looked at the effects of traffic information in the use of alternative routes for drivers. They suggest that drivers formulate route choice decisions in real time. In other words, drivers start off with a route in mind, but then when they see what the conditions are like on the chosen road, or find out from another previously listed source, they are then likely to make an in-vehicle decision to switch to an alternate route. The result of this is that as new information comes in, from local observations or from another agent, the driver will make new decisions on the route to take. The problem with this method though is that it insists that traffic information be correct. For the most part, traffic reports available from news services are reasonably accurate and can be trusted, but this topic is beyond the scope of this project.

2.5.5 Time of Day

Time of day considerations would seem to clearly be a factor in determining route choice decisions. For commuter drivers who drive to work and must be there at some specific time, choosing a route to follow would be very important depending on the time of day. If a driver is familiar with the traffic patterns at different times of the day, he may be able to avoid certain traffic tangles that at other times of the day do not exist. In a study done by Mahmassani and Stephan on commuters, they simulated real life decision making by asking actual commuter drivers route choice and departure time questions. The results of the study were that commuters would first pick a route, and then they would determine what time they had to leave their origin to arrive at their destination by

\[\text{Ibid.}\]
the appropriate time. It was shown that the drivers would modify their departure time by a significant amount before they would modify their route.\textsuperscript{23}

In considering time of day as a decision factor for long haul truckers, the analysis is most likely slightly different. There has been very little research done in this area, however, one could speculate. Long haul truckers are similar to commuter drivers in many aspects of route choice decision-making, and they are also quite different in other areas. In the area of time of day considerations, it is fairly easy to argue that long haul truckers are much more flexible than other drivers. In general, a truck driver will pick up a load and then have a scheduled time of delivery for the load. For long haul truckers this may be a few days in the future. Because the law limits truck drivers in the number of hours per day that they can drive, the truckers will want to pick the optimal times of the day for them to drive so that they can cover the most distance in the shortest period of time. Getting caught in a morning or evening rush hour is not a productive use of a trucker’s time, not to mention being in a traffic jam uses extra fuel and causes additional wear on the machinery.

2.5.6 Congestion

In the area of route choice decision-making, there have been numerous studies that have looked at congestion as a factor of route choice decision-making. Todd Burner ‘99, as an example, looked at congestion on alternate routes via perceived speed curves as influences on the route choice decision process. Burner raised an important question in

his analysis of perceived speed curves. Studies had shown that drivers, when either they observed congestion or were given actual congestion information, they modified their routes.\textsuperscript{24} Burner suggested that not only observed congestion or information about congestion ahead were the only factors in route choice decision making, but almost more importantly, the perception of congestion either ahead on the chosen route or on alternative routes was a very important decision factor in the utility functions. If one assumes this to be true, then the components of the Utility Function are not limited to observable factors and people’s perceptions and feelings also contribute to the utility function.\textsuperscript{25}

There has been some work done to try to explain this decision making process. Chee-Chung Tong attempted to explain this by using behavior decision theory. This theory stipulates, “An individual’s judgment and choice are two integral parts of the same process.”\textsuperscript{26} It is really quite simple. First, people perceive what is going on around them and formulate some sort of judgment. Essentially, people do not react or make decisions based on what is actually happening, rather they make decisions based on what they perceive is happening. The result as it pertains to this study is that there are other factors involved in the decision making process that contribute to the decisions that are being made that are going to be difficult to quantify.


\textsuperscript{25} Ibid.

\textsuperscript{26} Tong, Chee-Chung et al. Travel Time Prediction and Information Availability in Commuter Behavior Dynamics. In \textit{Transportation Research Record 1138}, TRB, National Research Council, Washington, D.C., 1987, pg 1.
2.5.7 Additional Factors

It is safe to assume that the seven previously stated factors are not the only factors that contribute to the utility function. However it is also likely safe to assume that these seven factors contribute a majority of the characteristic shape of the utility function as well as explain a large portion of its variance. In other words, these factors provide a very good first approximation. Other factors that could be considered in future studies that might influence decision-making and further quantify the characteristic shape of the utility function are: the scenery or type of neighborhood that the route goes through, the quality of the road, the number of turns along the route, or the number of rest areas along the route.

2.6 Work done by Todd Burner ‘99

The work done by Todd Burner has proven to be very helpful in the writing of this thesis. He did excellent work in setting a framework for congestion analysis. His thesis primarily focused on route choice decision making with anticipated congestion on alternate routes. He accomplished this by looking at a number of different case studies across the U.S. highway network. He chose the cases based on a few different criteria. He wanted to look at cities that had a major heavily traveled highway that passed through the central business district (CBD) of the city. In addition, each case city must have also had a bypass route around the downtown area that for the most part has a more limited access than the downtown route.
After selecting the case cities, Todd then developed perceived speed curves for the downtown route in relation to the bypass route. The method that he used to determine these perceived speed curves will be covered in greater detail later, as the same method will be used in this thesis. Using the results of the Logit model, he forecasted the expected percentage of traffic that would use the bypass route given the perceived speeds on the downtown route. If the perceived speeds on the downtown route were lower than the posted speed limits, the reduction in speed was considered a result of congestion.

2.6.1 Value of work done by Todd Burner

The most important conclusion that Todd came up with in his thesis relates to his work with the perceived speed curves that he generated from the Logit model. His conclusion on the risk aversion of the drivers in relation to avoiding congestion is a very valuable conclusion. According to Todd,

“Perhaps the most interesting characteristic of the perceived speed curves is how insensitive the perceived speed is to the percentage of trucks that use the bypass (or looking at the inverse, how sensitive the percentage of trucks on the bypass is to a very small changes in perceived speed on the downtown route)…. This implies that truck drivers are very sensitive to small changes in perceived speeds, which by extension indicates a very high value for time.”

These findings are quite interesting because they suggest that there is some sort cost function that can be associated with the time differences. Todd suggests that due to the level of risk aversion shown by the drivers, the cost should be fairly significant.
In addition to Burner’s work on the perceived speed curves of downtown routes, he also laid a solid foundation for continued study in his definitions and assumptions. When one considers stop definition, quite possibly the most important definition for this type of work given the data set, Burner’s definition, forward progress of less than two miles over a thirty-minute period, allows for an accurate benchmark to base stops.

2.6.2 Drawbacks to work performed by Todd Burner ‘99

The work that Todd Burner ’99 performed on his senior thesis is fantastic. His analysis of route choice via perceived congestion on alternate routes was very important to the field of route choice. The most important drawback to his work though is in his data set. His data set consisted of a group of around 20,000 trucks over a 7-day period. While this may seem to be a very sizeable data set, when considering the vast size of the United States interstate highway network, in many places it was quite difficult to get enough data points to be able to do a thorough analysis. With the considerably larger data set used in this study, it is now possible to fully carry out an analysis of this nature.

2.7 Prospect Theory

Perhaps some of the most important work being done on the cutting edge of economics is related to Princeton professor Daniel Kahneman. His ground breaking empirical research that resulted in what is now known as prospect theory is some of the most powerful current work on economics and decision-making. Professor Kahneman, together with Amos Tversky, developed a theory for explaining why individuals’
decision making under uncertainty deviates from what is predicted by standard economic theory. Standard economic theory suggests that individuals are “rational” decision makers. This means that they will make decisions that are totally rational. For example, when faced the option of receiving $100 with certainty, or with $5000 with certainty and all other things being equal, the rational decision maker will choose the $5000. This area of rational decision-making begins to get a little less clear though when the concept of variance or a gamble are introduced into the proposition. Let's say now we have a new situation. You can now receive $5000 with certainty, or after flipping one coin, if it is heads you receive $10,000 and if it is tails you receive nothing. What would you pick? Kahneman and Tversky attempted to explain why people act like they do with Prospect Theory.

2.7.1 Profile of Kahneman and Tversky

Both Daniel Kahneman and Amos Tversky are of Israeli dissent. They both were in the Israeli army where Tversky received a commendation for bravery while Kahneman developed a psychological profiling system for screening army recruits. Today however, both of these individuals work is more highly studied by Wall Street elites than by military scholars. This is a totally logical conclusion as the work done by these two individuals is primarily focused on decision making under uncertainty and Wall Street is one of the most uncertain places in the world. In addition, uncertainty influences almost every major decision. The outcome of every capital allocation is unknown, but Wall

---

28 Ibid.
Street professionals need to be able to make decisions without being paralyzed by the analysis of possible outcomes.

### 2.7.2 Forecasting the Future

In doing studies on pilots, Kahneman noticed that an individual’s performance on successive runs in general regressed towards the mean. In other words, if a pilot had a particularly good landing attempt, it was more likely that the next landing attempt would not be as good as the previous and visa versa. This idea raised some interesting questions for Kahneman. Would it be possible to forecast the future merely by studying a data set on the whole, then focusing on recent events and suggesting the subsequent events would occur closer to the mean? Essentially, what Kahneman was examining is a principle of auto-correlation in random events. An easy example to describe the idea of auto-correlation is in relation to basketball players. Often times while watching a game, fans will notice that a player has become “hot” or that they have a “hot hand” where their present probability of sinking a shot is greater than their historical probability of making a shot. Athletes often refer to this concept as “being in the zone.” The problem with this idea is that it is not true. Studies have shown that the probability that a player makes the next shot given that he has made successive shots is not statistically different from their regular historical probability of making a shot. Additionally, in the world of investing, if one looks at annualized daily volatilities of stock returns and compares them to annualized weekly volatilities of stock returns, one would see auto-correlation in practice if the annualized weekly volatility numbers were greater than the annualized daily

---

\(^{29}\text{Ibid.}\)
numbers. However, truth be told, the annualized numbers for both sets of calculations are identical. One would only see higher volatilities for weekly numbers if the data were auto-correlated, meaning that days when the stock market went up were followed by other up days and visa versa. These two examples relate to Kahneman and Tversky because they suspected that people err in forecasting because they falsely believe in auto-correlations and do not believe strongly enough in regression to the mean. In order to test their hypothesis, they designed a number of experiments to test how people make decisions when faced with uncertainty.

The results of their experiments were quite remarkable. The conclusions of Prospect Theory went against popular belief in the rationality assumption for decision makers. Kahneman and Tversky suggested that this was because of two basic shortcomings in humans. The first shortcoming is that emotion destroys self-control, which is essential to rational decision-making. For example, in the heat of the moment, it is totally realistic that individuals would change from being risk averse to risk seekers. The only problem with this proposition is that it violates the first principle of preference evaluation: completeness. As it relates to behavior of drivers, the incidence of “road rage” is significantly higher in cases involving congestion because of the increased emotional strain. Secondly, people suffer from what psychologists call “cognitive difficulties”, in that they are unable to fully understand what they are dealing with. In other words, people do not fully understand what they are dealing with because they distort their perception of reality around them. People are overly, and irrationally, afraid of low probability high drama events than they should be and they are under concerned

30 Ibid.
31 Ibid.
32 Ibid.
with routine events. For example, if asked about shark attacks, people conjure up ideas about Jaws and become overly worried when entering the water when the real probability that they get attacked is quite remote. However, if asked about the propensity to die from a fall down the stairs, people are less concerned because walking up and down the stairs is a routine occurrence. For this example, walking up and down the stairs is the base case where individuals make “normal” decisions. They are very familiar with the task of walking up and down the stairs. In reality though, the actual probability of dying from a fall down the stairs according to the National Safety Council is around 1:200,000 which is 25 times more likely then the 1:5,000,000 chance of dying from a shark attack.\textsuperscript{33} Because people don’t fully understand what they are dealing with as it relates to being attacked by a shark, they allow their emotions to destroy self-control; their behavior as it relates to decision-making is affected.

2.7.3 Asymmetry of Risk

Some of the more important work that Kahneman and Tversky did in the behavioral sciences was through examining individual’s profiles for risk and how they change given format changes in the questions. They found that people “treat costs and uncompensated losses differently, even though their impact on wealth is identical.”\textsuperscript{34} They showed this by performing a number of experiments.

One experiment that they performed was a basic coin flip gamble. Individuals had the option of either taking an amount with certainty, in this case $3000, or choose to gamble on the outcome of a “coin flip” with the probability of “heads” being 80% and

\textsuperscript{33} http://dsc.discovery.com/convergence/sharkweek2002/quiz/quiz.html
\textsuperscript{34} Ibid. Bernstein pg. 272.
resulting in a gain of $4000. Under the circumstance of a “tail” toss, they gambler would get $0. Kahneman and Tversky found that individuals overwhelmingly preferred the certain outcome even though the payoff of the risky bet was $3200.\textsuperscript{35} This result was not too unusual as the behavior could be sourced to risk aversion on the part of the participants. However, the results were much more interesting when Kahneman and Tversky changed the premise around from receiving $3000 with certainty to losing that same amount of money. The same holds for the gamble. In this case with gains now losses, they found that now the vast majority where now taking the gamble. People who just shortly before were risk averse were now risk seekers! This result is quite puzzling as it goes against the central tenets of rational behavior.

In another case study, Kahneman and Tversky asked a group of respondents about their risk aversion as it related to saving lives and death. The story they told the respondents was that there is a rare disease breaking out in a community that is expected to kill 600 people. They respondents have two options available to them to save lives. In the first questioning session, people were asked to choose between the first plan, which would save 200 lives with certainty, and the second plan that would save all 600 with a probability of one third and nobody would be saved with probability two thirds. Most of the respondents opted for the risk-averse choice of the first plan. In the second questioning session, Kahneman and Tversky switched the question around so that the questioning involved the words “die”. Option three now is that 400 of the 600 people will die while option four is that there will be a one third probability that nobody dies and a two-thirds probability that 600 people will die. It is important to note that now the options are posed in the frame of reference of death as opposed to survivors. This small

\textsuperscript{35} Ibid.
change in phrasing resulted in an incredible change in risk behavior. 78% of the subjects were now risk seekers compared to 72% previously being risk averse.\textsuperscript{36}

The question is now, what to do with this information? People switching from risk averse to risk seeking is inconsistent with many of the previously stated assumptions of rational behavior. Primarily, respondents originally preferred A (certainty) to B (the gamble), but then later preferred B (the gamble) to A (certainty) though the changes in wealth were equivalent. This violates the completeness property, assuming of course that certainty is not “suitably close” to gambling. Kahneman and Tversky came to the conclusion that people are not necessarily risk averse, as they would have chosen the gamble based on the alternative phrasing. Rather, they suggest that people, on the whole, understand variance and uncertainty, but what it really comes down to is that people strongly dislike losing and that decision makers appetite for risk is not in reference to how much wealth they currently have, but rather in reference to how the decision will affect their level of wealth. People remember their losses and carry them forward in their minds much more than their wins.\textsuperscript{37} This idea of loss aversion is concurrent also in the area of sport psychology. Dan Gable, a wrestler at Iowa State University and eventual Olympian was renowned for saying that avoiding the embarrassment of losses and failure drove him to train and prepare much harder than training for the satisfaction of winning matches. Interestingly, Gable was so averse to losses that he won a gold medal in the 1972 Munich games without surrendering a single point.\textsuperscript{38}

\textsuperscript{36} Ibid. pg. 273
\textsuperscript{37} Ibid. pg. 273-274
\textsuperscript{38} Taken from www.dangable.com
2.7.4 Ambiguity Aversion

Lastly, in an interesting study published by Daniel Ellsberg, the concept of ambiguity aversion is discussed. This concept refers to the idea that people are more comfortable with risk and taking risks when they know what the probabilities of outcomes are. This is entirely logical. When faced with the option to bet on a coin flip versus picking a black ball out of an urn filled with a number of balls whose colors he does not know, the average person would chose to bet on the coin flip. The implication to truck driver route choice is that when drivers are in unfamiliar areas, they are less likely to switch to alternate routes than when they are in familiar areas even if the level of information that they have on the alternate routes is the same. This can be simplified into saying that information, in this case the driver’s familiarity with their surroundings, is important in the decision making process.

2.7.5 Prospect Theory as it Relates to this Study

So far, Kahnemen’s and Tversky’s Prospect Theory has been shown to be influential in the decision making process whether it be in terms of risk aversion or in irrational behavior. Prospect theory is important to this study because it provides some explanation of the decision-making processes that individuals use. This thesis is an attempt to examine the basic human decision making processes. Kahneman and Tversky suggest that humans do not always follow rational decision making processes in that they change their preferences based on how the decision will affect their wealth and how the

question is proposed. This is an interesting idea that this thesis will attempt to shed light on.

2.8 Other Scholarly Truck Specific Studies

An enormous number of studies have been done on passenger vehicle route choice. There are however, relatively few studies that have focused solely on truck driver route choice. In addition, many studies have treated all vehicles in the system as the same type of driver. Passenger automobiles are grouped together with long haul truckers. Other studies have entirely neglected the trucks in the system when doing their analysis.

As stated earlier, trucks are not identical to passenger automobiles in terms of the driver’s risk behavior and the type of routes that they pick among other characteristics. Additionally, passenger vehicles have significantly higher flexibility than long haul truckers in terms of destinations. Trucks generally have to be at a certain place at a certain time, and these characteristics are beyond the control of the driver. For passenger vehicles, excluding the typical home to work trip, they have more flexibility on their destinations. For example, an individual is at home and he wants to get some gas for his car. He may have a preferred gas station to go to, but if it is impossible to get there, it is possible that the driver would divert to another gas station. Likewise, passenger vehicles generally have fewer restricted roads than long haul truckers due to height, weight, or hazardous material restrictions on trucks.
The primary reason for so few studies on trucks, according to Robert Stammer et al., is that there has been a lack of reliable Origin-Destination information on trucks.\textsuperscript{40} As seen in the previous chapter, O-D pairs are crucial in the network assignment process as they tell where the vehicle starts from and where it ends. Without this information, it is impossible to do the route determination stage of the network assignment process.

This data set, while lacking in actual O-D pairs as well, still proves very helpful in the analysis that will be done. A heuristic will be developed that will allow for generation of pseudo O-D pairs. This thesis will not look to build a model to incorporate the entire network; rather, case studies of small, representative, sections of the network will be examined. As a result, only trucks that pass through certain points on the network will be examined and those points will be used as the representative Origin-Destination pairs.

3 Explanation of the Data Set

The data set this thesis uses, along with the analysis of it, is the most valuable portion of this study. The data set, after significant data reduction, which will be explained later, consists of over 60,000,000 observations of truck time and location data for over 249,465 unique trucks over a thirteen-day period between August 29, 2002 and September 10, 2002. These observations shed light on the movements of long haul truckers across the United States highway network. By analyzing these movements, this thesis will explore the basic decision making processes that the drivers use. Up until recently, the measuring and modeling of human behavior has been very difficult. Acquiring a sufficient amount of unbiased, complete, and accurate data has been the major shortcoming that has plagued other studies. This thesis is one of the first academic studies that has access to a data set large enough that random sampling and scarcity of observations considerations are only minor factors. While this revealed preference data set does not show many of the conditions that are factors in the utility function, and thus the decision making process, it does allow for many characteristics to be derived. Each observation in the data set is a vector that consists of the following:

- A unique alphanumeric truck identifier (it remains constant throughout the data set.)
- A date field in the format “ddmmyy”
- A time field that measures the number of seconds from midnight of the first day of observations.
- A latitude position in the format “ddd.mm’ss”
A longitude position in the format “\textit{ddd.mm'ss}”

The following sections will discuss ways to collect this type of data, problems associated with using this type of data in determining route choice decision making behavior for the general truck population, and other hurdles that this data set posed.

3.1 Collection of Data

Though the exact method used to collect this data is not known, there are three likely candidates as possible methods of collection for position data: Inertial Navigation Systems, Loran, and Global Positioning System (GPS). The first system relies on very sensitive motion sensing equipment and accelerometers to determine location information.\footnote{Farrell, Jay A. and Matthew Barth. \textit{The Global Positioning System and Inertial Navigation}, United States of America: R. R. Donnelley & Sons, 1999, pg. 13.} Loran and GPS rely on a concept known as Time of Arrival ranging to calculate location information.\footnote{Kaplan, Elliot D. \textit{Understanding GPS: Principles and Applications}, United States of America: Artech House, Inc., 1996 pg. 15} The following sections will discuss the intricacies of Time of Arrival Ranging, Loran, GPS, Inertial Navigation Systems and lastly some of the problems involved with data collection using these systems.

3.1.1 Time of Arrival Ranging

Time of Arrival Ranging, or TOA, is one of the most popular methods of determining one’s position relative to his surroundings. TOA relies on the measurement of the time that it takes for some sort of signal to travel from a known location such as a
radio tower, satellite, or lighthouse to the observer’s location. The observer calculates the time that it takes for the signal to arrive and that time is then multiplied by the speed of the signal that results in the distance to the signal source or transmitter. The speed of the signal is generally either the speed of light or the speed of sound depending on the type of signal generated by the transmitter. After the distance to multiple transmitters has been determined, the user uses the principal of trilateration to determine his exact location.

In determining a latitude and longitude position for a two-dimensional surface like this thesis uses, a number of transmitters are required for exact locations. When the user picks up the signal and determines the range to the transmitter, he knows that he is a specific distance from the known location of the transmitter $T_1$, but he does not know what direction he is. In other words, the user ($P$) knows that he is somewhere on the perimeter of a circle with radius $R_1$ centered at $T_1$.

---

43 Ibid.
44 This example and explanation is adapted from an explanation of GPS found at www.HowStuffWorks.com.
In order to determine the user’s location more accurately, he will need to attain signals from multiple sources. In the following case, the user has picked up a signal from two different transmitters, T1 and T2. The user can narrow down his position from somewhere on the perimeter of circle 1 to one of the two points at the intersection of circles 1 and 2, P.1 or P.2. He can do this because he knows how far he is from both transmitters. The only locations on a flat plane that satisfy the distance locations are P.1 and P.2. Additionally, there is a special case that only requires two transmitters to determine one’s exact position. This only happens though when the user is standing exactly on a line that passes through both T1 and T2.

The exact location can now be determined with the help of a third transmitter as long as the three transmitters are not located in a straight line. With the third transmitter, the location P.1 is eliminated and P is selected as the location of the user, as P.1 is not R3.
units away from T3. Additional transmitters can be used to increase the accuracy of the position information as well as error check the solution.

Loran and the Global Positioning System use this technique of trilateration to generate position information. If we generalize the surface of the surface of the earth to be a flat plane this exact technique can be used to generate position information. Loran outputs only latitude and longitude for a position, not altitude, and the area of coverage is small enough that approximating the surface of the earth to be a flat plane does not reduce the effectiveness of the system a significant amount.

Extending this example of trilateration to GPS is not difficult. Fundamentally, the mechanics of 3-D positioning is the same as 2-D positioning. If one thinks of the circles generated by the transmitters to now be spheres in three-space, the same technique can be
used. If the user determines that he is 25,000 miles from satellite A, he knows that he is somewhere on the surface of a sphere with a radius of 25,000 miles centered at satellite A. Once the user obtains a second signal from satellite B, he can determine that he is on an exact circle at the intersection of the two spheres. With three satellites, the user will know that he is at one of the two points of intersection of the three spheres. The user only needs to get a signal from one additional transmitter to determine his exact altitude relative to the transmitters. However with GPS, users are usually only on the surface of the earth. As a result, the surface of the earth can also act as a sphere so that only three visible transmitters are needed, though four are usually used to improve accuracy of location and the internal quartz clocks in the receiver units.

3.1.2 Loran

_LONG RANGE Navi_gation, or Loran for short, is a system of navigation developed during World War II to help the U.S. Navy to improve their day as well as night navigation.\textsuperscript{45} Today Loran is a federally funded terrestrial system that provides users with positioning as well as timing information. The system has coverage anywhere in the continental United States as well as much of the coastal waters. Currently, the United States Coast guard operates and maintains the Loran system at an annual cost of around $30 million. The current system of Loran, Loran-C, has been in place since 1957.\textsuperscript{46} Additionally, there are other nationalized Loran systems throughout the world in countries like China and Russia.


\textsuperscript{46} http://www.ac11.org/gpsvs.htm
The Loran system uses the principles of TOA ranging to determine locations. Radio transmitters are setup in a group of three or more and are separated by hundreds of miles. Each group has one master transmitter and a series of secondary transmitters. The stations constantly broadcast radio signals with precise timing information. These timing signals are gathered by the receiver and compared with the other timing signals from different master-secondary transmitters. The receiver measures the difference between the signals and calculates the distance to the transmitters. With these distances and the precise locations of the transmitters, the receiver calculates its position in the previously mentioned manner.\(^47\)

In general, the position information generated by Loran systems is very accurate. While accuracy of 5 meters is possible in conjunction with other navigation devices such as GPS, the accuracy in most areas covered by Loran is closer to 100m-300m range of accuracy. With this level of accuracy, compounded by the fact that Loran is available across the entire continental United States, Loran could be used to determine position information for a truck. Additionally, because the system is accurate enough, it is possible to determine what highway a truck is on with a high level of accuracy.\(^48\)

### 3.1.3 Global Positioning System

The Global Positioning System, or GPS, is an all weather, worldwide, continuous coverage, satellite-based radio navigation system that utilizes the principles of TOA to determine the location of GPS receivers anywhere in the world with a clear view of the

\(^{47}\) Ibid.
\(^{48}\) Ibid.
sky. The best part about the system is that it is available to all users at a cost equal to the cost of a receiver. There are no direct annual fees or charges to the users, as the system is funded by federal tax dollars. The Department of Defense started the system in the 1960’s to improve on the accuracy and availability of the Loran system. There are essentially three types of equipment that make up the GPS system: the network of satellites (Figure 3-4), GPS receivers, and land based control units.49

Figure 3-4: Image courtesy of the U.S. Department of Defense

The space-based portion of GPS consists of 24 full time satellites plus 3 extra back up satellites. While in orbit, the satellites function as the transmitters and generate a pseudo random stream of numbers that is used as a timing signal for the receiver portion of the system. Because it takes a certain amount of time for the signal to get to the surface of the earth from space, there is a lag in the data stream. The satellites are

49 Farrell pg. 142.
aligned so that there are 6 orbital planes equally spaced 60° apart in reference to the prime meridian. Each orbit plane contains four satellites that are spaced such that a satellite passes a point in the orbital plane every 6 hours. The radii of the orbital planes are 20,200 km. This organizational pattern ensures that a receiver will be able to pick up a signal from at least 4 different satellites anywhere on the surface of the earth at all times. Since the satellites’ orbits are less than the required distance to make them geosynchronous, they are in constant motion relative to the earth. The ground based control units monitor and update this motion.\textsuperscript{50}

The land-based portion of the GPS system is responsible for monitoring the status of the space-based portion in addition to broadcasting updating signals to the constellation. The terrestrial portion consists of six remote monitoring stations and four ground antennas located around the world in addition to a master control station located at Falcon Air Force Base in Colorado Springs, CO. The master control station is responsible for generating the orbital model and clock correction parameters for each satellite as well as relaying the correction information to the ground antennas so that the updates can be sent to the constellation.\textsuperscript{51}

The receiver portion of the system allows for the signals from the satellites to be collected and decoded so that its location can be determined. The hand held receivers gather at least four signals from different transmitters. From there, they calculate the time delay in the signals. The receiver generates the same pseudo random data stream as the satellites transmit and calculates the time differential between when the signal is received and when it was sent. The problem with this method though is that the satellites

\textsuperscript{50} Ibid.
\textsuperscript{51} Ibid.
use highly accurate atomic clocks to generate their timing signals while the receivers use less accurate quartz clocks. The less accurate clocks result in greater errors in time estimation. To get around this limitation, GPS incorporates a fourth satellite signal that is used to constantly update the clocks in the receivers. Using the time delay information the receiver determines its position relative to the in-view satellites by multiplying the time delay by the speed of light as the timing signals are electromagnetic radio signals that travel at the speed of light. Additionally, the receivers store the locations of the satellites in their memory, so that they always know where the satellites are and the receiver can then generate its position relative to the satellites.\textsuperscript{52}

The limitations to the accuracy of the GPS system on its own are minor, which accounts for its popularity. The errors in the accuracy of GPS come from a few sources. First, as stated earlier, the clocks in the receiver units are only quartz clocks, which are significantly less accurate than atomic clocks and account for a loss of precision in time delay analysis. Next, electromagnetic signals travel through a vacuum at the speed of light. However, to get from space, a relative vacuum, to the earth, the signal must travel through the atmosphere where the signal slows down and varies depending on the medium that it is traveling through. The result is the precision of system is reduced.\textsuperscript{53}

\subsection*{3.1.4 Inertial Navigation Systems}

Inertial Navigation Systems or INS are quite different from GPS and Loran in terms of how they calculate location information. Instead of relying on timing signals

\textsuperscript{52} Ibid.

\textsuperscript{53} Ibid. pg. 261
and expensive monitoring equipment, INS relies on Isaac Newton’s second law of motion:

\[
\text{Net Force} = \text{Mass} \times \text{Acceleration}
\]  \hfill (3-1)

Rewritten:

\[
\text{Acceleration} = \frac{\text{Net Force}}{\text{Mass}}
\]  \hfill (3-2)

In essence, an INS is just a piece of equipment with three highly sensitive accelerometers attached to measure accelerations in the three dimensions. The INS collects and records the outputs of the accelerometers based on the equipment’s set sampling rate. Armed with this information, along with the starting location for the system, the INS calculates the current position.

The INS records the accelerations and the direction that they occurred in. It also records the amount of time that the INS was accelerating. Taking the acceleration in the \( z \) direction as an example, one can multiply the acceleration over time of the INS times the length of time squared to obtain a distance segment in the \( z \) direction. Combining the \( z \) segment with the \( x \) and \( y \) segments, one can determine distance and direction that the INS currently is relative to the starting position.

Like GPS, there are problems with INS systems. First, noise in the sensors limits their ability to estimate parameters with a high level of accuracy. Next, INS system accuracy is limited to the rate at which new information enters the system. For example, if an INS samples acceleration once every 10 seconds, and in between the two sample times a large acceleration occurs which the INS does not pick up, the INS will not know that it has changed direction and will think that it is someplace that it is not. In addition,
the state estimation accuracy is directly related to the motion of the system. Lastly, the reliability of the system is highly dependent on obtaining external position updates.\(^5\)

### 3.1.5 General Comments on Accuracy of Navigation Equipment

The collection of position and navigation information is always an inexact science. While the accuracy of the collection equipment is not exact, for this study, they all have a sufficient level of accuracy. Many of the tools used for navigation rely on sensing of fields that are external to the actual equipment. Compasses rely on the sensing of magnetic fields that are not constant. Also, GPS and Loran sense electrical fields that vary with time and for a variety of reasons are also not always available. INS systems, though they do not require the sensing of fields outside their self-contained equipment, also have fundamental problems relating to accuracy. Going forward however, increasing accuracy and precision of the equipment can be accomplished by augmenting the primary collection method with additional alternative collection methods.

\(^5\) Farrell pg 11.
3.2 Complications and Difficulties with Data Extraction

There are a number of difficulties involved with the collection and use of this data set. A short list of the complications will be discussed below.

3.2.1 Origin-Destination Information

One of the most important factors in analyzing a representative traveler in any study examining routing and travel demand is obtaining origin-destination (OD) information. It is necessary to have the origin and destination information of the drivers, so that one can determine where the truck started and where it stopped on any given trip. This data set did not explicitly specify separate trips, rather, it contained all of the location information for all of the trucks as a single grouping of consecutive observations. In other words, the data came as a single chunk of the entire 13 days for each truck. Additionally, the data set does not divulge anything that occurred between separate location observations. The result is that it is impossible for the analyst to know if a driver was on a single trip, or had multiple trips linked together. Also, because the observations were generally separated by about 45 minutes, it was impossible to tell exactly what route the truck followed between consecutive observations. This will make a large assumption that the driver took the shortest path between separate observations. While this assumption could seriously damage the accuracy of the analysis, it appears as if it is a fairly safe assumption.
3.2.2 Non-Random Sample of Trucks

One important consideration to take into account for this study is the possibility that the sample of trucks in the database is not a totally random sample of trucks in general. This is an important consideration because if the data is not totally random sample of the population of trucks, then the decisions shown in this representative sample could potentially be different from the decisions displayed by the rest of the truck population. For this study, it is likely that this is not a random sampling of trucks because the trucks in the data set had to have special position gathering equipment. The trucking company was required to purchase the equipment, and thus many truckers are excluded from the data collection.

3.2.3 Random Time Spacing of Observations

One other factor that further complicated the analysis for this thesis was the lack of uniformity of time spacing of observations. Each truck had a different number of observations over the two week period that were not equally spaced apart in times. Furthermore, the frequency of reporting of position information was also varying such that a truck could report five observations spaced 20 minutes apart, and then the next observation might be 4 hours later. This made the analysis very difficult to determine some of the route choice decisions being made. As will be shown later, the time spacing was also a large factor when analyzing some of the smaller sections on the US highway network. There was no guarantee that there would be an observation in the study area even if the truck did pass through the respective area. Another assumption was made that further complicated the accuracy of the analysis. This study assumes that the truck
traveled at a constant rate between two observations. This can potentially reduce accuracy of analysis because the interpolated speed of the truck might not be the actual speed of the truck. As a result, this study made an effort to reduce this problem by limiting the number of acceptable trips to those that occurred in a “reasonable” amount of time. The specific amount of time will be covered later in the discussion of the analysis.

3.2.4 Observation Reporting Error and Map Matching Complications

As stated previously, neither INS, GPS, or Loran systems can determine exactly where the truck is. There is always some error. Depending on the system used, this error could be as little as 1 m or as much as 5000 m. However, they do provide a pretty good estimate for the location of a truck. In this study, the position observations were gathered and matched to the U.S. highway network using a complicated algorithm courtesy of ALK associates Co Pilot® navigation system that would “snap” the observation to a specific link on the network. The problem with this method though is the location of the observation was not always on a highway or a road for that matter. Additionally, even if the observation snapped to a major highway, many highways have other roads that run parallel to the highway that trucks do drive on. If the matched location is not the actual location, it is possible for the accumulated errors to not accurately portray the decisions made.
3.2.5 Stop Determination

One last factor that posed problems for the data analysis and data reduction was the determination of stops. Stop determination is very important because it signifies the beginning or the end of a trip. The data set did not contain any information that explicitly stated that the truck had made a stop. However, this thesis used a heuristic to estimate stops. There are a number of areas in the data set where there is a significant time differential in two consecutive time tags and only a small difference in the reported position of the truck. A stop was defined as any point in the data set where two or more consecutive observations were close enough together such that a truck traveling at 4 miles per hour could go between the two points in a 30 minutes or less. This heuristic was used in the data reduction stage of analysis, but it presents a problem. Because of the relative infrequency of observations, it is possible that a truck could have been at a terminal, left to make a delivery, and returned before any position observation was made. This is important because many short trips are left out of the data analysis.
3.3 Additional Data Considerations

As stated previously, there are many factors that contribute to the utility function of drivers. Many of those factors however are difficult to gather. Because most of those factors are not included in the data set, the precision of analysis is reduced. The following sections will cover a few of the more important factors in the decision making process that were not included in the data set.

3.3.1 Observed Traffic on Routes

As stated previously, traffic is a very important factor that effects decision making by truck drivers. It is possible that the drivers had information about either their primary route or an alternate route that effected their route choice decisions. While it would significantly increase the complexity of the data analysis, it would be very helpful to have the traffic information for the area around the trucks. This information could be incorporated into the decision making model and could potentially dramatically increase the precision of the analysis.

3.3.2 Socioeconomic Factors

Socioeconomic factors have been previously demonstrated to dramatically effect decision making. The lack of socioeconomic data in the database is therefore problematic for an analysis such as this. Not only does the lack of socioeconomic data limit the precision of the analysis, but also in applying the results of this study to the general population of truck drivers, it is difficult to make predictions of decision making
even if one had the socioeconomic data for any given driver. For this study, because the data is not available, socioeconomic factors will be entirely ignored and will be grouped in as part of the error term in the analysis.

3.3.3 Characteristics of Specific Trucks

Lastly, the data does not include any special characteristics of the trucks. For example, it is not possible to determine if a representative truck is carrying hazardous waste and is not allowed to follow certain routes. Route choice determination is effected because if a driver is unable to use a certain route, the data does not show this factor. One would not know if a truck driver chose a sub optimal route because he was hauling hazardous waste, or if the driver perceived the given route to be superior to the alternatives available. This study assumes that trucks are allowed to travel freely on any route that allows truck traffic.
4 Methodology

As with any scholarly data analysis, it is important to clearly define the thought, as well as analysis, process used to come to any conclusion. This thesis is focused on the route choice decisions that truck drivers make. It uses an enormous revealed preference data set. However, the data is not formatted in an easy to use format with all of the routes and decisions laid out. There is a large amount of data that is essentially useless to this study, so it is necessary to get into the data and pull out all of the “good” data to be used in the analysis. During this reduction process, it is important to keep in mind that this study wants to reduce the data into a more workable size while at the same time maintaining the information content of the data set.

4.1 Stop Determination

The first area where this thesis looked to remove extraneous data was when the trucks were stopped. The data set contained a large amount of position data for trucks that showed that the trucks were essentially stopped for extended periods of time. This observation is in agreement with reality because truck drivers are only allowed to drive for a set number of hours each day and must stop to rest during each day. This thesis, however, is not concerned with trucks that are stopped. This thesis views a stop as the end of a trip. Determining stops, however, is not as easy as one might think.

The format of the data made it possible to determine stops via a heuristic without any degradation in the information content of the data set. Because the data collection
methods had some sort of error involved with them, it was unlikely that a truck would report itself to be in *exactly* the same position. It was however quite common that the truck would report itself to be quite close to its previous observation. This study decided that if a truck reported itself to be within 4 miles of its previously reported position 30 minutes or longer after the previous observation, then the truck was deemed to be stopped over that time period. The advantage to this method allowed for significant data reduction to be performed and the data analysis was further simplified. Additionally, there were a number of trucks that were stopped for eight or more hours in the same location. The algorithm would search for these instances and remove all of the stop data except for the first and last stopped observations. One possible drawback to this heuristic is the possibility that a truck was stopped at a terminal, made a location observation, then departed to make a delivery, and returned to the terminal before making another observation. While this chain of events is quite possible and would potentially alter the conclusions of this thesis, it is important to keep in mind that this thesis is focused on the basic decision making behavior of long haul truckers rather than single load delivery vehicles.

4.2 Determination of Regions to be Examined

The determination of zones to be examined came from three sources. First and foremost, Todd Burner ’99 set the framework for determining appropriate cities to be examined in his senior thesis. Additionally, Prof. Alain Kornhauser contributed to the list of potential analysis regions. Lastly, I picked the Chicago Skyway case study, as well as the 90/94 case study.
The cases to be examined are as follows:

- Chicago, IL (Chicago Skyway-80/94)
- Cincinnati, OH (I-75/I-275)
- Columbus, OH (I-70/I-270)
- Indianapolis, IN (I-70/I-74)
- Nashville, TN (I-40/I-440)
- Memphis, TN (I-40/I-440)
- Houston, TX (I-10/I-610)
- Oklahoma City, OK (I-40/I-240)
- Richmond, VA (I-95/I-295)
- St. Louis, MO (I-55/I-255)
- San Antonio, TX (I-35/I-410)
- Wilmington, DE (I-95/I-495)
- Interstates 90 and 94 between Tomah, WI, and Hirsch, MT

As was previously stated, the analysis regions were picked based on the layout of the highway network surrounding major cities. Regions that were selected had a major highway that lead up to and through the downtown area of a major city in addition to a bypass route that circumvented the downtown area of the city. These areas were selected because they proved to be fertile ground for performing decision-making analysis. Additionally, the 90/94 case study was chosen because it hopefully would also be a fertile
area for analysis of perceived speeds on alternate routes that were greater than 100 miles in length.

4.3 Trip Determination

Determining trips from the data set proved to be quite a large task. There are a few software packages available that can take raw GPS data as inputs and map match them to the U.S. highway network. This study chose to use ALK Associates’ CoPilot guidance package.

The first attempt at analyzing the data involved a process where 10,000 trucks worth of GPS data would be loaded into the computer. Then, one of the previously determined case studies would be examined. The number of trucks on either the downtown or the bypass route would then be counted manually. This method had many significant drawbacks. First, because there was so much data being shown on any given screen, it was quite difficult to single out any specific truck. Additionally, it was quite difficult to determine exactly what route the truck had taken. For example, it was possible that a truck would enter the downtown route, make an observation, and then immediately exit the downtown route. It was also quite difficult to accurately count the numbers of trucks on each route because of the infrequency of the data collection. Lastly, sorting out “good” truck data from 250,000 trucks for 13 analysis zones and two possible routes is an enormous task to do by hand, so a better method was needed.
4.3.1 Route Determination Heuristic

In order to streamline the data analysis, another heuristic was devised to determine what routes people took. For travel demand modeling, it is important to have Origin-Destination pairs. These pairs are used to generate the trips and routes for the analysis. In looking at the data, one piece of information that has proven to be quite useful is the latitude and longitude information contained in the GPS records. With this information, along with the exact location of the roads, one can determine exactly where the truck is on the map. Additionally, this thesis only wants to examine trucks that pass through the selected analysis region without stopping. This study needs only trucks that pass through the analysis region, because the drivers of those specific trucks are totally free to make their own route choice decisions. They do not need to go to a certain point in the city to drop off a load, thereby forcing the driver to take a specific route. The first idea conceived was to pick all of the links in the CoPilot database that corresponded to the downtown route as well as the bypass route and figure out which trucks “snapped” to those routes. However, the computational intensity of this method is beyond the scope of this project, and a simpler method was devised.

4.3.2 The Box Algorithm

The method devised, henceforth referred to as the Box Algorithm, proved to be a viable heuristic. The premise behind the Box Algorithm is the data is already formatted in a nice numeric format that is easy for a computer to work with. This study leveraged the computing power of three supercomputers to sort the data and extract the useful data. Essentially, the process the algorithm utilized is as follows. First, the user inputs a series
of pairs of GPS coordinates. These pairs were used to set the northwest and southeast corners of a rectangular box. For a list of the input coordinates, please refer to the Appendix. The program takes four pairs of coordinates to generate boxes 1, 2, 3, and 4 respectively. Boxes 1 and 2 were set up so that they capture the trucks that were on the highway leading up to and out of the analysis zone. These boxes are essentially the O-D pairs for the trips. Box 3 was set up to capture the downtown route for most cases and Interstate 90 for the 90/94 case. Box 4 was set up to capture the bypass route for most cases and Interstate 94 for the 90/94 case. Boxes 3 and 4 were used to generate the route decisions that the drivers made. Please refer to Figure 4-1 for a graphic representation.

Figure 4-1: Map of Houston road system with boxes covering Highway 10 leading into and away from Downtown Houston in addition to Box 3, which captures the downtown portion of Highway 10 and Box 4, which captures the bypass portion of Highway 610.
The computer program would then sift through the 250,000 truck records to generate instances where a truck was in any given box. The output file consisted of a stream of data that included the Truck ID, Time of Observation, Latitude, Longitude, and the box number that the truck was found in. The output file was then sorted by Truck ID and Time of Observation. The output file was then used as an input into another program that would sort through the data and output all of the trucks that were found in box combinations that matched one of the following sequences: [1, 3, 2], [1, 4, 2], [2, 3, 1], [2, 4, 1]. The program accomplished this task using a very complicated Finite State Machine or FSM. Because FSMs are quite useful in pattern matching, they are the primary candidates to sort through the data. The simplified method that the FSM used is as follows. First, the data would be read in. Each observation would receive a tag that signified which box the observation was in. Next, the data was examined to see if any of the previously mentioned patterns were found. The four patterns were important because if a truck had position data that was in Box 1, then Box 3, and finally Box 2, that meant that the truck came in from the east, chose to take the downtown route, and exited the city to the west. However, the data was not already in a nice format, so many different error checks were done. For example, there were many cases where the truck would come through Box 2, then the next observation would be in Box 1, and another later observation would be in Box 3. This trip, however, does not help the analysis because it is not the type of trip that this study is looking for. Refer to Figure 4-2 for a graphical representation of the FSM operator. Additionally, C code for the FSM can be found in the appendix.
The output from the FSM program was a file that broke the data down into usable trips. The program had two file outputs: *data3.txt and *data4.txt corresponding to the downtown route data and bypass route data respectively. These two files contained a stream of trips. The format of that data is: Truck ID, Start Time, Start Date, Start Basis Seconds, End Time, End Date, End Basis Seconds, Travel Time, and Average Speed. The average speed calculation was one of the derived statistics that the program output. By taking the location information from Box 1 and Box 2 and then inputting them into a great circle calculation, this thesis was able to derive a good approximation for the total distance traveled and then calculate the average speed by using the distance and travel time. The following is an extract of a great circle algorithm.
\[ p1 = \sin(first\_lat\_d) \times \sin(last\_lat\_d); \]
\[ p2 = \cos(first\_lat\_d) \times \cos(last\_lat\_d); \]
\[ p3 = \cos(abs\_long); \]
\[ \text{distance}_c = \arccos(p1+p2\times p3) \]

An analysis was then done for all of the routes. The summary statistics can be found in the appendix. The output statistics were as follows: Count of Trucks on Downtown and Bypass routes, Average Travel Time, Minimum Travel Time, Variance of Travel Times, Std. Deviation of Travel Time, Mean + 2 Sigma Travel Time, Median Travel Time, and Maximum Travel Time.

4.3.3 Drawbacks Associated with the Box Algorithm

When using any heuristic, there are always going to be certain drawbacks. This thesis made a concerted effort to minimize any biases that would occur while using the Box Algorithm. However, a few problems did appear as the analysis was being performed. First, there are a few analysis regions where there is not enough data to perform an accurate analysis. For example, the Wilmington region has only seven trucks that take the bypass route and zero trucks that take the downtown route. This can be explained by two different possibilities. First, the analysis region is no more than 900 square miles. Included in that area are no more than 70 miles of roads encompassed by boxes that are being analyzed. The trucks in this study only report their locations at random times with \( E[\text{time between observations}] = 45 \) minutes. It is virtually certain that
more than 7 trucks in the data set passed through the Wilmington area, but it is rather unlikely that they would have made a position observation in boxes 1,2, and 3 or 1,2, and 4. The other possibility is that trucks did pass through Wilmington and make observations on the highway leading into the city, out of the city, and either the bypass or downtown routes, but the observation was not included in one of the analysis boxes. If one refers to Figure 4-1, he will notice that neither the entire length of the downtown nor the bypass routes are included in a box. This was done to minimize the number of trucks mistakenly included in the data set that might have been on different highways than the analysis highways.

The other disadvantage to using the Box Algorithm is that for cities that had the major highways passing through the city at some direction other than North, South, East, or West such as Northwest to Southeast, it was quite difficult to find a rectangle that would encompass enough of the highway so that there would be a sufficient number of observations on the route while also minimizing the interference from trucks on totally different highways. Additionally, finding a way to make Box 3 and Box 4 similar in size to each other was quite difficult. This thesis attempted to include the same number of miles of highway in Box 3 as in Box 4, but differences in box size could contribute to a bias in the data towards either the bypass or downtown route.
4.4 Further Data Reduction Using Minimum and Maximum Average Speeds

After analyzing the results of the FSM program, it became apparent that there were still trucks that were included in the data set that should not have been included. This study is only concerned with trucks passing through the city without stopping. The data does not tell why a trucker would stop, either for gas or to make a delivery, so it is impossible to eliminate only select trips. As a result, it is necessary to eliminate all the trips where the analysis leads one to believe that the probability of a stop was rather high. There are a few ways of accomplishing this, but the simplest method involves using minimum average speeds for the trips.

As previously stated, the average speed for the trips were determined using a great circle calculation to find the approximate distance traveled and then using the travel time as the time divisor to find the approximate average speed. There is no theory or rule in place to determine if a truck made a stop on a trip or not based on its average speed, however, another heuristic was developed. This study decided that if a truck went slower than 30 mph for its entire trip, the data should be eliminated from the analysis.

One significant drawback to this method however was that more case cities were eliminated from the analysis. There were a few cities, such as Memphis and Chicago, that either had no trucks on the downtown or on the bypass route where the average speeds for their respective trip was greater than 30 mph. One possible explanation why so much data was eliminated results from the algorithm used to pull out the good data. Because the Box Algorithm only sensed position data and did not reduce data that did not make sense, too much data was selected. A number of the trips had travel times that were
anywhere from ten to thirteen days where the trip was only 30 to 70 miles in length. This can be explained by the possibility that a truck passed through Box 1 early in the week. Later in the week, it happened to pass through Box 3, and the next week through Box 4. The Box Algorithm would record this as a single trip even though it is an impossibility. Therefore, trips with average speeds less than 30 mph were eliminated.

Lastly, there was the issue of trucks traveling too fast on some routes. While it might seem unlikely that the data would show that a truck was traveling too fast on a route, it happened in a few cases. For example, in the Oklahoma City case study, truck 39960 had an average speed of 263 mph. This is impossible and is attributable to “bad data”. Like the minimum average speed case, there is no rule or theory for a maximum speed on a route. As a result, another heuristic was developed such that any truck that had an average speed greater than 75 mph was eliminated from the analysis.

Output statistics for the final data run including truck counts on routes and average speed characteristics can be found in the appendix.
5 Building the Model

The goal of this thesis is to examine the basic human decision making process that long-haul truckers use to formulate and carry out route choice decisions. By analyzing the basic decision making process, this study is attempting to better quantify and more clearly define some of the more important factors in the decision making process. The most important factors to be analyzed are how long-haul truckers trade off between distances and time when faced with multiple routes.

The more general analysis began in earlier chapters with commentary on Utility Functions. Truck drivers were assumed to be rational decision makers and by extension utility maximizers. Some factors that have been shown to contribute to the overall decision making process and thus the utility function include factors such as: length of alternate routes, expected traffic on routes, risk aversion, income and education, and time of day. While some, including Nobel Prize winner Prof. Daniel Kahneman, propose that humans are not necessarily rational decision makers in reality, this thesis will attempt to explain this purported irrationality by proposing that humans base decisions primarily on perceptions.

This study will examine tradeoffs made by truckers between distance and time. The most obvious place to analyze tradeoffs between distance and time are in areas where there are multiple plausible routes to get from point A to point B. Theoretically there are infinite routes between two points, however, plausible routes include routes that are suitably similar to either the time or distance minimized optimal solution. The most common area to have multiple alternate routes is around major cities that have a major
highway passing through the downtown area and a bypass route that skirts the downtown region. These particular case cities are fertile areas for analysis of tradeoffs between distance and time for a few reasons. First, typically, the downtown route follows the most direct path from one side of a city to the other and as a result is shorter than the bypass route. Next, the bypass route typically is designed as a limited access route that is used by travelers who are not planning on stopping in the central business area of the city. Limiting access to a road generally speeds up the flow of traffic because drivers have less of a need to alter their speed. Conversely, downtown routes typically have far more exits and entrances thereby increasing the probability of interrupting the flow of traffic resulting in congestion. Additionally, highway designers implement bypass routes to minimize the potential for excessive congestion caused by through traffic in the downtown area. The results of the differences between the downtown and bypass routes are that, in general, the expected value of the travel time on the downtown route is lower than the travel time for the bypass route, however, the variance in travel time is at the same time greater on the downtown route than the bypass.

5.1 Foundation for Discrete Choice Models

Previous chapters have already covered how the subset of trucks to be analyzed was selected. After implementing the Box Algorithm and performing significant data reduction, this study counted the number of trucks on the bypass route and the downtown route. With this information, this study then calculated the percentage of trucks on the bypass route. Information regarding specific counts and selected data can be found in the appendices. By incorporating the percentage of trucks on the bypass route in conjunction
with the specific characteristics of the downtown route and bypass such as distance of trip, time of trip according to calculations made by CoPilot®, and time of trip on each path assuming 55 mph on the downtown route and 65 mph on the bypass route, one can build a model that predicts the percentage of users that use the bypass route as a function of the perceived speed on the downtown route. This thesis chose to examine the usage of the bypass route because there is a greater chance of variability of travel times on the downtown route, thus perceptions for speeds on the downtown route have a greater variability.

5.2 Utility Maximization

This thesis has proposed that truck drivers are rational decision makers, thus they are utility maximizers. When making a decision on which route to take, the truck driver will either consciously or unconsciously calculate his expectation of utilities for the potential routes based on his perceptions of the routes. Once this calculation is performed, the driver will then pick the route that offers the maximum utility. This study is focused solely on a choice between two options, so the difference in utility can be quantified using the following equation:

\[ \Delta Utility = Utility_{or} - Utility_{bp} \]  \hspace{1cm} (5-1)

Using this equation, the driver will only select the downtown route with certainty if \( \Delta Utility \) is greater than 0. Otherwise, the driver is either indifferent between routes, or prefers the bypass route.
The problem now is how to determine either $\Delta Utility$ or the utilities for the downtown and bypass routes respectively using the travel times and distances for each route in addition to the derived quantity of the percentage of trucks on the bypass. Because this is a discrete choice problem, a Logit model is the likely candidate.

5.3 The Logit Model

The case studies in this study were picked because the truck drivers were essentially only able to choose between two plausible routes. As a result, the decision is a binary choice between two alternatives: take the downtown route or take the bypass route. Logit models are designed so that they predict the probability of choosing one of the routes based on the inputs to the model. The inputs to the Logit model are simply the distance and time characteristics of the respective routes. The output of the model is the probability of trucks that take the bypass route. The model is calibrated using the percentage of trucks on the bypass route, which was derived by sifting through the truck position data.\footnote{Burner. Pg. 70.}

The outcome of the route choice decision (R) will be defined as:

$$R = 1 \text{ iff the bypass route is used}$$

$$R = 0 \text{ iff the downtown route is selected}$$

As was previously stated, the decision to chose a particular route is quite complex. There are theoretically an infinite number of factors that contribute to the decision process. However, when considering only distance and time, the decision process can be modeled as a function defined as:
\[ Y = \text{constant} + \beta_{\text{distance}} \cdot (\text{Distance Factor}) + \beta_{\text{time}} \cdot (\text{Time Factor}) + \text{error term} \quad (5-2) \]

This thesis will set the variable \( Y \) to be the difference in the two utility functions: \( \Delta Utility \). One complication results from this definition. It is now necessary to define one utility function for each route. However, one trick to get around this problem is to define \( \Delta Utility \) to be a function of the Distance Factor and Time factor for both the downtown and bypass routes.\(^{56}\)

One method to model the difference in \( \Delta Utility \) is to use a combination of the difference in the distance and time factors while assuming speeds on the respective routes. However, this simplistic approach is not very helpful in modeling the decision making factors, because it does not capture enough of each factor. Essentially, the results of this method would be the differences in distances and speeds for the two routes. However, when making route decisions, one must consider more than just the distance savings in absolute terms.\(^{57}\) The 90-94 case study in relation to the Wilmington, DE case study is a perfect example of this significance. Each of the routes in the 90-94 case are approximately 1000 miles in length. The Wilmington routes are approximately 10 miles in length. A change in distance of one mile on the Wilmington route should have more significance to the decision process than the 90-94 case because of the percentage change in the Wilmington case is significantly greater than the 90-94 case.

Todd Burner ’99 proposed using factors that show the percentage difference in routes as opposed to the absolute difference in routes. This has two important benefits. First, units of time and distance drop out of the calculation. Secondly, when estimating parameters of the Logit function, multiple case studies can be used to calibrate the model.

\(^{56}\) Ibid.  
\(^{57}\) Ibid.
because the cases are relatively similar. The equation for the ratio that the individual uses to evaluate the distances for the two routes is as follows:

\[
R_{dist} = \frac{(dist_{BP} - dist_{DT})}{\left(\frac{dist_{BP} + dist_{DT}}{2}\right)}
\] (5-3)

The equation for the ratio of times for the two routes is similarly defined using the time considerations for the downtown and the bypass routes instead of the distance considerations. Plugging these distance and time factors into the \(\Delta Utility\) function yields:

\[
\Delta Utility = \alpha + \beta_{distance} R_{dist} + \beta_{time} R_{time}
\] (5-4)

When considering the \(\Delta Utility\) function, it is important to keep in mind that the utility has no meaning across different cases. This concept was covered in Chapter 2. The \(\Delta Utility\) is merely used as a comparison to make a decision. In other words, the \(\Delta Utility\) is not transferable between cases. A new \(\Delta Utility\) must be calculated for each case because each case has a unique ratio of distances and times.

With \(\Delta Utility\) defined, it is now necessary to define the odds of choosing the bypass route for the Logit model.

\[
\frac{P(R = 1)}{P(R = 0)}
\] (5-5)

By defining the natural logarithm of the odds of \(R = 1\) to be \(\Delta Utility\), then:

\[
\ln\left(\frac{P(R = 1)}{P(R = 0)}\right) = \Delta Utility
\] (5-6)

This equation can be rearranged such that:
\[ P(R = 1) = \frac{\exp(\Delta Utility)}{1 + \exp(\Delta Utility)} \]  

(5-7)

Now that these equations have been defined, it is necessary to estimate the parameters of the \( \Delta Utility \). Estimating the parameters used in the Logit function (L) for this study is an exercise in maximum likelihood estimation.

\[
L = \prod_{i=1}^{n} \left( \frac{\exp(\Delta Utility)}{1 + \exp(\Delta Utility)} \right)^{D_i} \left( 1 - \frac{\exp(\Delta Utility)}{1 + \exp(\Delta Utility)} \right)^{(1-D_i)} \]  

(5-8)

In this equation, \( n \) represents the number of case studies used to calibrate the model. The \( D_i \) represents the respective probabilities for case study \((i)\). The \( \pi \) symbol is used to represent the sum of products.

It is now necessary to take the logarithm of equation (5-8) in order to derive the log-likelihood factors. Additionally, using properties of logarithms, equation (5-8) can be rearranged:

\[
L = \sum_{i=1}^{n} \left[ D_i \log \left( \frac{\exp(\Delta Utility)}{1 + \exp(\Delta Utility)} \right) + (1 - D_i) \log \left( 1 - \frac{\exp(\Delta Utility)}{1 + \exp(\Delta Utility)} \right) \right] \]  

(5-9)

Because this is an attempt to maximize the likelihood of the parameters the Excel solver is used. The following inputs were used in the optimization program to solve for the parameters in the utility function:
The two different calculations were done using two different sets of inputs. For the sections dealing with CoPilot, the inputs for the times, and thus derived speeds, were taken from the output of the trip calculation performed by CoPilot. The other section uses an assumption of 55 mph on the downtown route and 65 mph on the bypass route. The
following are the observed number of trucks on the downtown (light green) and bypass (pale blue) for each city used in the calibration of the model:

<table>
<thead>
<tr>
<th>City</th>
<th>90-94</th>
<th>Cincinatti</th>
<th>Columbus</th>
<th>Houston</th>
<th>Indianapools</th>
<th>Memphis</th>
<th>Oklahoma City</th>
<th>Richmond</th>
<th>San Antonio</th>
<th>St Louis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>501 Trucks</td>
<td>35 Trucks</td>
<td>289 Trucks</td>
<td>905 Trucks</td>
<td>160 Trucks</td>
<td>2845 Trucks</td>
<td>2076 Trucks</td>
<td>166 Trucks</td>
<td>499 Trucks</td>
<td>56 Trucks</td>
</tr>
<tr>
<td>Count</td>
<td>128 Trucks</td>
<td>2 Trucks</td>
<td>102 Trucks</td>
<td>140 Trucks</td>
<td>16 Trucks</td>
<td>148 Trucks</td>
<td>251 Trucks</td>
<td>395 Trucks</td>
<td>171 Trucks</td>
<td>168 Trucks</td>
</tr>
</tbody>
</table>
5.4 Derivation of Perceived Speeds

After obtaining the parameters of the $\Delta Utility$ function through the maximum log-likelihood estimation program, it is now possible to determine the perceived speed on the downtown route.

The Logit function can be rearranged to read:

$$ P = F(Z) = \frac{1}{1 + e^{-z}} $$  \hspace{1cm} (5-10)

Where $Z$ is set equal to the $\Delta Utility$ function and $P$ is equal to the probability of choosing the bypass.\(^{58}\) Equation (5-10) is equivalent to function (5-7).

In generating the forecasted probabilities for usage of the bypass route, the following were the results:

<table>
<thead>
<tr>
<th>Case</th>
<th>Calculated Probabilities via Logit Model</th>
<th>Observed P of Bypass</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. Louis, MO Case Study</td>
<td>0.457299292</td>
<td>0.75</td>
</tr>
<tr>
<td>San Antonio, TX Case Study</td>
<td>0.223100997</td>
<td>0.255223881</td>
</tr>
<tr>
<td>Richmond, VA Case Study</td>
<td>0.303040597</td>
<td>0.704099822</td>
</tr>
<tr>
<td>Oklahoma City, OK Case Study</td>
<td>0.192234955</td>
<td>0.107864203</td>
</tr>
<tr>
<td>Memphis, TN Case Study</td>
<td>0.16827417</td>
<td>0.049448714</td>
</tr>
<tr>
<td>Indianapolis, IN Case Study</td>
<td>0.458542529</td>
<td>0.090909091</td>
</tr>
<tr>
<td>Houston, TX Case Study</td>
<td>0.172068118</td>
<td>0.133971292</td>
</tr>
<tr>
<td>Columbus, OH Case Study</td>
<td>0.179096821</td>
<td>0.260869565</td>
</tr>
<tr>
<td>Cincinnati, OH Case Study</td>
<td>0.066642555</td>
<td>0.054054054</td>
</tr>
<tr>
<td>90/94 Case Study</td>
<td>0.389638874</td>
<td>0.203497615</td>
</tr>
</tbody>
</table>

One of the most valuable aspects of the truck data set that this thesis utilizes is that the probability of trucks on the downtown and bypass routes is known. Because of this fact, it is possible to back out the value of the $\Delta Utility$ function. Logit models are generally used to make predictions of the expected probability of usage for the routes.

---

\(^{58}\) Ibid. pg. 73
based on the ΔUtility function. However, because the actual probabilities of usage for the routes are available, it is possible to use the ΔUtility function to derive the perceived speed on the downtown route and by perturbing the input probability for the bypass route, it is possible to generate a curve for the perceived speed for all probabilities.\(^{59}\)

Using the two sets of speed assumptions, CoPilot and 55/65, when looking at the Logit function, the only variable that is free to change with the probability on the bypass route is the time spent on the downtown route. Because the distance of the downtown route is known, it is easy to solve for speed on the downtown route.

\[
Speed = \frac{\text{Distance}}{\text{Time}} \quad \text{(5-11)}
\]

In examining the Logit function in its fully expanded form when solved for speed on the downtown route:

\[
speed_{\text{downtown}} = \frac{\text{Dist}_{\text{downtown}}}{2 - \left( \frac{-\ln\left( \frac{1}{P} - 1 \right) - \alpha - \beta_{\text{distance}} \cdot R_{\text{dist}}}{\beta_{\text{time}}} \right) * \text{Time}_{\text{BP}}} \quad \text{(5-12)}
\]

Equation (5-12) is used to generate the perceived speed curves that will be discussed in detail in the following chapter.

\(^{59}\) Ibid. pg. 74
6 Commentary on Results

The previous chapter covered the intricacies of Logit models and the parameter estimation that is used to develop an appropriate model for this data set. Ten of the thirteen possible case cities were used as inputs in running the optimization model that predicted the maximum log-likelihood estimators for the parameters. The other three cities did not have significant data that would allow for any reasonable estimation of parameters. The results of the optimization can be put into a single formula that can then be applied to all of the case cities to generate the perceived speed curves. All one needs to input are the respective distance ratios and time ratios for the case cities. The format of the equation is as follows:

\[
\frac{1}{1 + e^{-(-0.238 + 8.17\% (R_{\text{dist}}) + 7.028\% (R_{\text{time}}))}} \quad (6-1)
\]

The parameters derived via the optimization model are the result of an aggregation of all the probabilities and distance and time ratios for all of the case cities. Therefore, the perceived speed curves generated will not necessarily be in line with the observations of the percentage of trucks on the bypass route and the assumptions for speeds: both CoPilot speeds and 55/65 speeds. However, the overall qualitative aspects of the perceived speed curves are a good estimate of the actual perceptions that truck drivers have based on their perceptions. The portions that are of significant value are the coefficients for the distance and time ratios. These coefficients represent the tradeoffs that the truck drivers have displayed between distance and time. Discussion of the coefficients will follow the results of the perceived speed curve generation.
Perceived Speed Curves Using CoPilot Calculations

Perceived Speeds on Downtown Route Using Assumed Speeds
6.1 Interpretation of Perceived Speed Curves

The perceived speed curves, while time consuming to generate proved to be quite interesting to analyze. The meaning of the perceived speed curves is as follows. The percentage of trucks that follow the bypass route can be written as a function of the trucker’s perceptions of conditions on the downtown route. The assumption is that the truck drivers know the distances of the downtown and the bypass routes. They do not know the traffic conditions on either route with certainty. However, they might be able to form some sort of view or perception of the conditions on the downtown route. These perceptions are formed in a number of ways. One example of a way to form a perception is if a truck driver has driven on the downtown route before, he might be familiar with the layout of the route and be able to form a guess as to the level of congestion on the route. If when the last time the truck driver chose the downtown route, there was a significant amount of congestion, the driver might suspect that there currently is congestion on the route and his perception of the speed on the route will therefore decrease. When the perceived speed decreases, the probability that the driver chooses the bypass increases.

When examining the perceived speed curves, it is important to note that the curves were generated strictly using observed probabilities of usage, distance ratios and time ratios. As was stated earlier, there are many more factors that contribute to the cost function for the drivers. However, this revealed preference data set did not include any other data besides location and time. As a result, it is impossible to determine the exact values that drivers put on distance and time. However, one can speculate that, based on the relatively flat slope of the curves that drivers value distance and time at very different levels. The flat slope at the point of indifference signifies a very high level of risk.
aversion. In other words, there is a fairly large change in the percentage of people that use the downtown route for an incremental change in perceived speeds. This means that the truckers are very risk averse when picking their routes.

6.2 Points of Indifference

The interesting aspects though are the tradeoffs that the drivers make. The slope of the perceived speed curves at the 50\textsuperscript{th} percentile, or the indifference point (the 50\textsuperscript{th} percentile is the indifference point because there is a 50/50 chance that the driver will take the downtown or the bypass route), is remarkably flat. There is only a 10\% decline in usage for a 4 mile per hour change in the perceived speed. However, when one considers the speed difference when the percent of trucks that use the bypass route in the 90-94 case moves from 80\% to 20\%, the increase in perceived speed is almost 25 mph. If one considers a 25 mph change in travel speed, over a 1000-mile trip, the savings of time is enormous.

It is also interesting to examine the plot of the perceived speed in relation to the ratio of distances on the bypass route and the downtown route at the point of indifference. The following is a chart of perceived speed vs. the distance ratio at the point of indifference between the bypass and downtown routes for the truckers. The plot uses the perceived speeds generated using the 55/65 assumption.
Indifference Speed as a function of the ratio of the distance of the bypass route to the downtown route

\[ y = -44.978x + 61.935 \]

\[ R^2 = 0.9965 \]

There are a few interesting aspects of this graph. First, the intercept of the regression is approximately equal to 62 mph. This is an interesting quality because the perceived speeds were generated using an assumption of 65 mph on the bypass route. The Y-intercept corresponds to the point where the X value is equal to zero. For this graph, the meaning of X equaling zero is the ratio of distances for the two routes equals zero. One would expect the Y-intercept to equal 65 mph. Additionally, it appears as if the value for the perceived speed drops off rather precipitously as the distance ratio increases only a relatively small amount.
6.3 Interpretation of the Parameters

When analyzing the results of the optimization program, it was interesting to contemplate the meaning of the coefficients of the R variables. The $\Delta Utility$ function is essentially the gain in utility for an incremental change in either time or distance in the trip. Based on the parameters it appears as if the magnitude for the time parameter is significantly higher than for the distance parameter. This would suggest what this study suspected all along is true: truck drivers are first and foremost time minimizers before anything else. Truck drivers look for the shorter routes, but only take them because they are also the minimum time routes.

6.4 Pitfalls in the Analysis

After performing the analysis of the data and performing the analysis of the perceived speed curves, it became evident that this analysis is not entirely perfect and there are a number of areas where the analysis could be improved. Some of the specific areas that could be improved on are: small sample size for certain cases, infrequency of data collection, inaccuracy of stop definition, and inaccuracy in selection of trucks on both the bypass and downtown route.

6.4.1 Small Sample Size for Specific Cases

After performing the analysis, it became evident that certain cases were not as fertile as this study had hoped. There are a few reasons why there was not a significant amount of data for a given case. One reason, which will be covered shortly, relates to the
heuristic used for the counting of trucks on routes. Another explanation relates to the small area for some of the cases. The Wilmington, DE case was especially difficult. Each of the routes, both bypass and downtown, were approximately 9 miles in length. If one assumes that all of the trucks that could have potentially been on the route were pulled from the data using the Box Algorithm (this is quite an ambitious assumption), then of those trucks selected, assuming an average interval of data collection of 45 minutes, the probability that the truck reported its location while on one of the routes is only around 20%. This percentage is would result in the maximum number of trucks pulled from the data set. Wilmington was not the only case where infrequency of reporting was a problem.

6.4.2 Infrequency of Data Collection

The previous section alluded to some of the problems relating to the infrequency of data collection. This was especially problematic for most of the cities. Because of the infrequency of data collection, it was significantly harder to determine stops. For example, if a truck came into a city, chose the downtown route, then got of the road to make a quick delivery, and returned soon after the delivery, the truck should have been removed from the data set, however, the truck probably was grouped into the slower end of the 30-75 mph selection heuristic.
6.4.3 Inaccuracy of Stop Definition

Another area of concern is in the stop definition. Because of previously mentioned problems with infrequency of data collection, it was quite difficult to determine stops. It was entirely possible, in fact likely, that a truck made a stop between observations, but the stop selection heuristic showed that the truck was still moving. It is possible to argue that the inability to discern stops was equally detrimental to the bypass data as the downtown data, however, the strength of this argument is fairly limited. It is much more likely that a truck made a stop on a downtown route than on the bypass route. There are far more potential stops that a truck could make that are closer to the downtown area than the bypass zone.

6.4.4 Problems Associated with the Box Algorithm

One of the most pressing issues that should be addressed in future analysis is in regards to the Box Algorithm. The Box Algorithm is a heuristic used to simplify the collection of truck data from each of the routes. The problem that this thesis encountered though was when the bypass and downtown routes were significantly close to one another. It was very difficult to find a box that would maximize the length of road in the box while at the same time no include any portion of the alternate route or another major highway. Additionally, it was very difficult to ensure that the length of road selected by the downtown box was equal to the length of road selected by the bypass box. Because of this, the count of trucks on either one of the routes could be skewed towards the route with the larger box. Lastly, the boxes did not include the entire bypass or the entire downtown route. Because of this, it is even more likely that the count of trucks on the
routes was biased downwards. Also, in picking the boxes, it was harder to find a box to pick the bypass route because of the fact that the bypass route generally had much more curvature because it skirted the city whereas the downtown route was generally a straight shot through the city. When the road was straight, it was much easier to select data.

6.4.5 Overall Impact on the Model

All things considered, a majority of the previously mentioned pitfalls in the analysis should only have a limited impact on the analysis process. Because the model is should be used in qualitative applications, the soundness of the model should remain unchanged. However, the problems associated with the Box Algorithm were significant. Possible solutions to the problem will be covered in the following chapter.
7 Conclusion

The goal of this study was to do an empirical analysis of long-haul trucker route choice decision-making as they navigate the U.S. highway network. This study looked at the basic decision making process that the drivers demonstrated. This study examined areas in the U.S. highway network where truckers were faced with a decision between two different routes: a downtown route that was passed through, or very close to, the center of the city, and a bypass route the skirted the edge of the city. By examining the decisions shown by the truck drivers it is possible to determine the manner at which drivers trade off distance and time. This was accomplished by building a Logit model that was calibrated with data from almost 250,000 trucks over a 13-day period. The logit model predicted the percentage of trucks that used the bypass route as a function of the perceived speed on the downtown route.

Using the perceived speed curves, this thesis attempted to explain the decision making process in terms of rationality of decision-making. This thesis showed in the previous chapter that truck drivers are primarily time minimizers. (If the drivers were distance minimizers, then the route choice decisions would be made for them and there would be no need for further study.) Supposing that drivers are time minimizers, it is now interesting to examine the decision-making process. Kahneman would say that the drivers are irrational because they pick routes at random. They should know what they speed is on each route before making a decision if they are rational drivers. Since they have no basis for the actual conditions on either route, any decision that they make is made irrationally. However, this thesis proposes that the decision making process is more scientific than a simple 50/50 coin flip when drivers are faced with uncertainty.
This thesis proposes that truck drivers usually do not just make a guess about which route to take when faced with similar routes. Instead, the drivers make route choice decisions based on their respective perception of each route. Their perceptions are generated from a number of factors including past experience on the route, time of day, current traffic conditions, and knowledge of the route to name a few. The driver puts all of these factors together, along with other non-quantifiable factors to generate an idea for the expected time of the trip. With this information, the driver then derives the perceived speed on the downtown route. With this perceived speed, the driver then subconsciously compares the result with the perceived speed on the alternate route. With these speeds, the driver then computes the time spent on each route. The minimum of these two quantities results in the time minimized, utility maximized, optimal solution. The explanation for the need for a probability of usage on the bypass route results in the fact that different people have different perceptions of the route in front of them. One driver may have taken both routes previously with one route being significantly better than the other route. A different truck driver may have had an entirely different experience, or no experience previously. As a result, the second driver is going to most likely have different perceptions of the route in front of his truck.

7.1 Interpretation of the Results

The results of this study do not conclusively show that truck drivers make rational decisions with respect to their perceptions. The data does however show that time is a significant factor in the decision making process. It is impossible to tell exactly how much more significant time is in the decision making process then distance is due to the
dimensionless nature of utility, however, it is likely that time plays a more important role than distance.

7.2 Areas for Further Study

While this thesis did not provide a strong proof to explain the decision making process of long-haul truckers, it did shed some more light on the more important factors in the decision making process. It is also safe to say that there is more to be done with respect to this study to come to more meaningful conclusions. The three most obvious areas that would benefit from further study are the determination of stops, the frequency of data collection, and improving on the Box Algorithm.

Todd Burner faced many of the same problems associated with stop determination. It is really difficult to determine, from this data set, if a truck is stopped. Stop determination is very important because it is likely that there are trucks in the study that should not be analyzed. The goal of this study was to examine route choice decision making of trucks that were passing through a city without stopping. Having these extraneous trucks in the sample could potentially alter the results of the study.

Along the same lines of stop determination, if the frequency of data collection were increased, there would be fewer problems with “bad data” making it into the group to be analyzed. Having more frequent location observations would allow for easier stop determination in addition to simplifying the counting of the trucks on each route. The results would be more precision in model calibration stages and a more accurate picture in general.
Lastly, the Box Algorithm must be updated. Currently, the Box Algorithm misses far too much data. Additionally, the Box Algorithm returns slightly biased data. Because of these facts, it is vital that for further analysis the next analyst should improve on the technique of selecting the subset of trucks used in the study.
Bibliography


Appendix 1: Summary Statistics of Trucks on Routes:

First Attempt

<table>
<thead>
<tr>
<th></th>
<th>90-94</th>
<th></th>
<th>Cincinatti</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>1303 Trucks</td>
<td>404 Trucks</td>
<td>307 Trucks</td>
<td>23 Trucks</td>
</tr>
<tr>
<td>Average Travel time</td>
<td>178787.8358</td>
<td>224681.0124</td>
<td>301181.7459</td>
<td>304563.3043</td>
</tr>
<tr>
<td>Min Travel Time</td>
<td>55951</td>
<td>51495</td>
<td>1728</td>
<td>5460</td>
</tr>
<tr>
<td>Variance of Travel Time</td>
<td>5996875.99</td>
<td>10151688.18</td>
<td>21401441.67</td>
<td>38232325.09</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>40.81</td>
<td>53.10</td>
<td>77.10</td>
<td>103.05</td>
</tr>
<tr>
<td>Median Travel Time</td>
<td>141676.00</td>
<td>158905.00</td>
<td>239885.00</td>
<td>90431.00</td>
</tr>
<tr>
<td>Max Travel Time</td>
<td>1101912</td>
<td>1106632</td>
<td>1113067</td>
<td>1067931</td>
</tr>
<tr>
<td></td>
<td>306.0866667</td>
<td>307.3977778</td>
<td>309.1852778</td>
<td>296.6475</td>
</tr>
<tr>
<td></td>
<td>Columbus</td>
<td>Houston</td>
<td>Memphis</td>
<td>Indianapolis</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Count</td>
<td>1198 Trucks</td>
<td>1295 Trucks</td>
<td>Count</td>
<td>790 Trucks</td>
</tr>
<tr>
<td></td>
<td>804 Trucks</td>
<td>634 Trucks</td>
<td></td>
<td>376 Trucks</td>
</tr>
<tr>
<td>Average Travel time</td>
<td>170212.621</td>
<td>36933.27645</td>
<td>Average Travel time</td>
<td>227602.2342</td>
</tr>
<tr>
<td></td>
<td>201392.25</td>
<td>136848.3486</td>
<td></td>
<td>228402.609</td>
</tr>
<tr>
<td></td>
<td>47.28</td>
<td>10.26</td>
<td></td>
<td>63.22</td>
</tr>
<tr>
<td></td>
<td>55.94</td>
<td>38.01</td>
<td></td>
<td>63.45</td>
</tr>
<tr>
<td>Min Travel Time</td>
<td>1126</td>
<td>1720</td>
<td>Min Travel Time</td>
<td>912</td>
</tr>
<tr>
<td></td>
<td>1394</td>
<td>3616</td>
<td></td>
<td>2105</td>
</tr>
<tr>
<td></td>
<td>0.31</td>
<td>0.48</td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>0.39</td>
<td>1.00</td>
<td></td>
<td>0.58</td>
</tr>
<tr>
<td>Variance of Travel Time</td>
<td>17536248.19</td>
<td>3427861.65</td>
<td>Variance of Travel Time</td>
<td>21514892.86</td>
</tr>
<tr>
<td></td>
<td>19590797.30</td>
<td>12511177.16</td>
<td></td>
<td>21149359.85</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>69.79</td>
<td>30.86</td>
<td>Std. Dev.</td>
<td>77.31</td>
</tr>
<tr>
<td></td>
<td>73.77</td>
<td>58.95</td>
<td></td>
<td>76.65</td>
</tr>
<tr>
<td>2 Std. Dev Time</td>
<td>139.90</td>
<td>147.93</td>
<td>2 Std. Dev Time</td>
<td>154.87</td>
</tr>
<tr>
<td></td>
<td>147.93</td>
<td>118.91</td>
<td></td>
<td>153.88</td>
</tr>
<tr>
<td>Median Travel Time</td>
<td>37673.00</td>
<td>7323.00</td>
<td>Median Travel Time</td>
<td>89523.50</td>
</tr>
<tr>
<td></td>
<td>76172.50</td>
<td>41801.00</td>
<td></td>
<td>96696.00</td>
</tr>
<tr>
<td></td>
<td>10.46</td>
<td>2.03</td>
<td></td>
<td>24.87</td>
</tr>
<tr>
<td></td>
<td>21.16</td>
<td>11.61</td>
<td></td>
<td>26.86</td>
</tr>
<tr>
<td>Max Travel Time</td>
<td>1162194</td>
<td>1102786</td>
<td>Max Travel Time</td>
<td>1166988</td>
</tr>
<tr>
<td></td>
<td>1188356</td>
<td>1059287</td>
<td></td>
<td>1148479</td>
</tr>
<tr>
<td></td>
<td>322.8316667</td>
<td>306.3294444</td>
<td></td>
<td>324.1633333</td>
</tr>
<tr>
<td></td>
<td>330.0988889</td>
<td>294.2463889</td>
<td></td>
<td>319.0219444</td>
</tr>
<tr>
<td>City</td>
<td>Count</td>
<td>Count</td>
<td>Average Travel time</td>
<td>Average Travel time</td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td>--------</td>
<td>---------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td><strong>Nashville</strong></td>
<td>38</td>
<td>15</td>
<td>261098.0526</td>
<td>275998.9333</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>72.53</td>
<td>76.67</td>
</tr>
<tr>
<td>Min Travel Time</td>
<td></td>
<td></td>
<td>1484</td>
<td>20761</td>
</tr>
<tr>
<td>Variance of Travel Time</td>
<td></td>
<td></td>
<td>31273185.47</td>
<td>13049735.97</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>93.20</td>
<td></td>
<td>60.21</td>
<td></td>
</tr>
<tr>
<td>2 Std. Dev Time</td>
<td>186.82</td>
<td>126.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Travel Time</td>
<td></td>
<td></td>
<td>83524.50</td>
<td>186367.00</td>
</tr>
<tr>
<td>Max Travel Time</td>
<td>1071437</td>
<td>608479</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Oklahoma City</strong></td>
<td>3102</td>
<td>379</td>
<td>42241.68665</td>
<td>69073.34037</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>72.53</td>
<td>76.67</td>
</tr>
<tr>
<td>Min Travel Time</td>
<td></td>
<td></td>
<td>985</td>
<td>1584</td>
</tr>
<tr>
<td>Variance of Travel Time</td>
<td></td>
<td></td>
<td>4384652.34</td>
<td>7430893.05</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>34.90</td>
<td></td>
<td>45.43</td>
<td></td>
</tr>
<tr>
<td>2 Std. Dev Time</td>
<td>70.07</td>
<td>91.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Travel Time</td>
<td></td>
<td></td>
<td>7238.00</td>
<td>7243.00</td>
</tr>
<tr>
<td>Max Travel Time</td>
<td>1101739</td>
<td>1050492</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Richmond</strong></td>
<td>678</td>
<td>1225</td>
<td>185294.1475</td>
<td>186377.409</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>51.47</td>
<td>51.77</td>
</tr>
<tr>
<td>Min Travel Time</td>
<td></td>
<td></td>
<td>1282</td>
<td>2810</td>
</tr>
<tr>
<td>Variance of Travel Time</td>
<td></td>
<td></td>
<td>16485569.32</td>
<td>15318390.74</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>67.67</td>
<td></td>
<td>65.23</td>
<td></td>
</tr>
<tr>
<td>2 Std. Dev Time</td>
<td>135.70</td>
<td>131.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Travel Time</td>
<td></td>
<td></td>
<td>69153.50</td>
<td>78067.00</td>
</tr>
<tr>
<td>Max Travel Time</td>
<td>1175109</td>
<td>1127068</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>San Antonio</strong></td>
<td>853</td>
<td>281</td>
<td>83436.99062</td>
<td>77789.59786</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>23.18</td>
<td>21.61</td>
</tr>
<tr>
<td>Min Travel Time</td>
<td></td>
<td></td>
<td>2329</td>
<td>3723</td>
</tr>
<tr>
<td>Variance of Travel Time</td>
<td></td>
<td></td>
<td>8142566.85</td>
<td>8578370.79</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>47.56</td>
<td></td>
<td>48.81</td>
<td></td>
</tr>
<tr>
<td>2 Std. Dev Time</td>
<td>95.76</td>
<td>98.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Travel Time</td>
<td></td>
<td></td>
<td>7747.00</td>
<td>8676.00</td>
</tr>
<tr>
<td>Max Travel Time</td>
<td>1078262</td>
<td>1019478</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>St Louis</strong></td>
<td></td>
<td><strong>Wilmington</strong></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------</td>
<td>----------------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td><strong>Count</strong></td>
<td>233 Trucks</td>
<td>334 Trucks</td>
<td>1 Trucks</td>
<td></td>
</tr>
<tr>
<td><strong>Average Travel time</strong></td>
<td>135101.8155</td>
<td>103577.3922</td>
<td>1411</td>
<td></td>
</tr>
<tr>
<td></td>
<td>37.53</td>
<td>28.77</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td><strong>Min Travel Time</strong></td>
<td>2139</td>
<td>1433</td>
<td>1411</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.59</td>
<td>0.40</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td><strong>Variance of Travel Time</strong></td>
<td>10716761.19</td>
<td>11802742.94</td>
<td>#DIV/0!</td>
<td></td>
</tr>
<tr>
<td></td>
<td>54.56</td>
<td>57.26</td>
<td>89.50</td>
<td></td>
</tr>
<tr>
<td><strong>Median Travel Time</strong></td>
<td>38904.00</td>
<td>7340.50</td>
<td>1411.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.81</td>
<td>2.04</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td><strong>Max Travel Time</strong></td>
<td>1041039</td>
<td>1044961</td>
<td>1411</td>
<td></td>
</tr>
<tr>
<td></td>
<td>289.1775</td>
<td>290.2669444</td>
<td>0.391944444</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>238.1630556</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 2: Summary Statistics of Trucks on Routes:
Second Attempt (Post Data Reduction)

<table>
<thead>
<tr>
<th>90-94</th>
<th>90-94</th>
<th>Cincinatti</th>
<th>Cincinatti</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>Count</td>
<td>Count</td>
<td>Count</td>
</tr>
<tr>
<td>501 Trucks</td>
<td>128 Trucks</td>
<td>35 Trucks</td>
<td>2 Trucks</td>
</tr>
<tr>
<td>Average Travel time</td>
<td>Average Travel time</td>
<td>Average Travel time</td>
<td>Average Travel time</td>
</tr>
<tr>
<td>82077.31138</td>
<td>91017.07813</td>
<td>2475.914286</td>
<td>6354</td>
</tr>
<tr>
<td>22.80</td>
<td>25.28</td>
<td>0.69</td>
<td>1.77</td>
</tr>
<tr>
<td>Min Travel Time</td>
<td>Min Travel Time</td>
<td>Min Travel Time</td>
<td>Min Travel Time</td>
</tr>
<tr>
<td>55951</td>
<td>51495</td>
<td>1728</td>
<td>5460</td>
</tr>
<tr>
<td>15.54</td>
<td>14.30</td>
<td>0.48</td>
<td>1.52</td>
</tr>
<tr>
<td>Variance of Travel Time</td>
<td>Variance of Travel Time</td>
<td>Variance of Travel Time</td>
<td>Variance of Travel Time</td>
</tr>
<tr>
<td>62165.91</td>
<td>75495.89</td>
<td>78.97</td>
<td>444.02</td>
</tr>
<tr>
<td>4.16</td>
<td>4.58</td>
<td>0.15</td>
<td>0.35</td>
</tr>
<tr>
<td>2 Std. Dev Time</td>
<td>2 Std. Dev Time</td>
<td>2 Std. Dev Time</td>
<td>2 Std. Dev Time</td>
</tr>
<tr>
<td>23.85</td>
<td>23.46</td>
<td>0.78</td>
<td>2.22</td>
</tr>
<tr>
<td>Median Travel Time</td>
<td>Median Travel Time</td>
<td>Median Travel Time</td>
<td>Median Travel Time</td>
</tr>
<tr>
<td>79493.00</td>
<td>92622.50</td>
<td>2352.00</td>
<td>6354.00</td>
</tr>
<tr>
<td>22.08</td>
<td>25.73</td>
<td>0.65</td>
<td>1.77</td>
</tr>
<tr>
<td>Max Travel Time</td>
<td>Max Travel Time</td>
<td>Max Travel Time</td>
<td>Max Travel Time</td>
</tr>
<tr>
<td>115432</td>
<td>118802</td>
<td>4008</td>
<td>7248</td>
</tr>
<tr>
<td>32.064444444</td>
<td>33.000555556</td>
<td>1.113333333</td>
<td>2.013333333</td>
</tr>
<tr>
<td>Min Avg Speed</td>
<td>Min Avg Speed</td>
<td>Min Avg Speed</td>
<td>Min Avg Speed</td>
</tr>
<tr>
<td>30.066673</td>
<td>30.003455</td>
<td>30.26799</td>
<td>26.221161</td>
</tr>
<tr>
<td>Max Avg. Speed</td>
<td>Max Avg. Speed</td>
<td>Max Avg. Speed</td>
<td>Max Avg. Speed</td>
</tr>
<tr>
<td>60.041853</td>
<td>66.118858</td>
<td>61.341051</td>
<td>30.445638</td>
</tr>
<tr>
<td>Median Avg. Speed</td>
<td>Median Avg. Speed</td>
<td>Median Avg. Speed</td>
<td>Median Avg. Speed</td>
</tr>
<tr>
<td>42.233006</td>
<td>38.062255</td>
<td>46.308082</td>
<td>28.3333995</td>
</tr>
<tr>
<td></td>
<td>Columbus</td>
<td>Houston</td>
<td>Indianapolis</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------</td>
<td>------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Count</td>
<td>289 Trucks</td>
<td>905 Trucks</td>
<td>160 Trucks</td>
</tr>
<tr>
<td></td>
<td>102 Trucks</td>
<td>140 Trucks</td>
<td>16 Trucks</td>
</tr>
<tr>
<td>Average</td>
<td>Average Travel Time</td>
<td>Average Travel Time</td>
<td>Average Travel Time</td>
</tr>
<tr>
<td>Travel</td>
<td>3805.858131</td>
<td>4623.66667</td>
<td>2657.5625</td>
</tr>
<tr>
<td>time</td>
<td>1.06</td>
<td>1.28</td>
<td>0.74</td>
</tr>
<tr>
<td>Min Travel Time</td>
<td>1372</td>
<td>1619</td>
<td>1203</td>
</tr>
<tr>
<td>time</td>
<td>0.38</td>
<td>0.45</td>
<td>0.33</td>
</tr>
<tr>
<td>Variance</td>
<td>Variance of Travel Time</td>
<td>Variance of Travel Time</td>
<td>Variance of Travel</td>
</tr>
<tr>
<td>of Travel</td>
<td>641.27</td>
<td>748.55</td>
<td>231.77</td>
</tr>
<tr>
<td></td>
<td>0.42</td>
<td>0.46</td>
<td>0.25</td>
</tr>
<tr>
<td>2 Std.</td>
<td>1.23</td>
<td>1.36</td>
<td>0.84</td>
</tr>
<tr>
<td>Dev.</td>
<td>Median Travel Time</td>
<td>Median Travel Time</td>
<td>Median Travel Time</td>
</tr>
<tr>
<td></td>
<td>3571.00</td>
<td>4691.50</td>
<td>2490.50</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>1.30</td>
<td>0.69</td>
</tr>
<tr>
<td>Max Travel Time</td>
<td>7383</td>
<td>8636</td>
<td>5341</td>
</tr>
<tr>
<td></td>
<td>2.0508333333</td>
<td>2.398888889</td>
<td>1.483611111</td>
</tr>
<tr>
<td>Min Avg Speed</td>
<td>Min Avg Speed</td>
<td>Min Avg Speed</td>
<td>Min Avg Speed</td>
</tr>
<tr>
<td></td>
<td>30.126765</td>
<td>30.232489</td>
<td>30.006676</td>
</tr>
<tr>
<td></td>
<td>30.590571</td>
<td>32.339033</td>
<td>30.00003</td>
</tr>
<tr>
<td></td>
<td>71.362467</td>
<td>70.610499</td>
<td>75.817036</td>
</tr>
<tr>
<td></td>
<td>52.370938</td>
<td>36.7895865</td>
<td>49.994506</td>
</tr>
<tr>
<td></td>
<td>5.3.6.2</td>
<td>5.3.6.2</td>
<td>5.3.6.2</td>
</tr>
<tr>
<td>City</td>
<td>Count</td>
<td>Count</td>
<td>Average Travel Time</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>--------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Oklahoma City</td>
<td>2076 Trucks</td>
<td>251 Trucks</td>
<td>6672.673892</td>
</tr>
<tr>
<td></td>
<td>1.85</td>
<td>1.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min Travel Time</td>
<td>Min Travel Time</td>
<td>1319</td>
</tr>
<tr>
<td></td>
<td>0.37</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Variance of Travel Time</td>
<td>Variance of Travel Time</td>
<td>836.65</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>Std. Dev.</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>2 Std. Dev Time</td>
<td>2 Std. Dev Time</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>Median Travel Time</td>
<td>Median Travel Time</td>
<td>7233.00</td>
</tr>
<tr>
<td></td>
<td>2.01</td>
<td>2.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max Travel Time</td>
<td>Max Travel Time</td>
<td>15261</td>
</tr>
<tr>
<td></td>
<td>4.239166667</td>
<td>4.0186111111</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min Avg Speed</td>
<td>Min Avg Speed</td>
<td>30.022938</td>
</tr>
<tr>
<td></td>
<td>Max Avg. Speed</td>
<td>Max Avg. Speed</td>
<td>75.006916</td>
</tr>
<tr>
<td></td>
<td>Median Avg. Speed</td>
<td>Median Avg. Speed</td>
<td>51.24547</td>
</tr>
<tr>
<td>Richmond</td>
<td>166 Trucks</td>
<td>395 Trucks</td>
<td>5574.662651</td>
</tr>
<tr>
<td></td>
<td>1.55</td>
<td>1.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min Travel Time</td>
<td>Min Travel Time</td>
<td>2614</td>
</tr>
<tr>
<td></td>
<td>0.73</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Variance of Travel Time</td>
<td>Variance of Travel Time</td>
<td>424.15</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>Std. Dev.</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>2 Std. Dev Time</td>
<td>2 Std. Dev Time</td>
<td>1.41</td>
</tr>
<tr>
<td></td>
<td>Median Travel Time</td>
<td>Median Travel Time</td>
<td>5439.50</td>
</tr>
<tr>
<td></td>
<td>1.51</td>
<td>1.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max Travel Time</td>
<td>Max Travel Time</td>
<td>8538</td>
</tr>
<tr>
<td></td>
<td>2.37166667</td>
<td>2.8711111111</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min Avg Speed</td>
<td>Min Avg Speed</td>
<td>30.280048</td>
</tr>
<tr>
<td></td>
<td>Max Avg. Speed</td>
<td>Max Avg. Speed</td>
<td>61.140095</td>
</tr>
<tr>
<td></td>
<td>Median Avg. Speed</td>
<td>Median Avg. Speed</td>
<td>43.230688</td>
</tr>
<tr>
<td>San Antonio</td>
<td>499 Trucks</td>
<td>171 Trucks</td>
<td>7149.220441</td>
</tr>
<tr>
<td></td>
<td>1.99</td>
<td>2.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min Travel Time</td>
<td>Min Travel Time</td>
<td>2331</td>
</tr>
<tr>
<td></td>
<td>0.65</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Variance of Travel Time</td>
<td>Variance of Travel Time</td>
<td>651.38</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>Std. Dev.</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>2 Std. Dev Time</td>
<td>2 Std. Dev Time</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td>Median Travel Time</td>
<td>Median Travel Time</td>
<td>7234.00</td>
</tr>
<tr>
<td></td>
<td>2.01</td>
<td>2.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max Travel Time</td>
<td>Max Travel Time</td>
<td>14023</td>
</tr>
<tr>
<td></td>
<td>3.895277778</td>
<td>3.6638888889</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min Avg Speed</td>
<td>Min Avg Speed</td>
<td>30.289082</td>
</tr>
<tr>
<td></td>
<td>Max Avg. Speed</td>
<td>Max Avg. Speed</td>
<td>71.446431</td>
</tr>
<tr>
<td></td>
<td>Median Avg. Speed</td>
<td>Median Avg. Speed</td>
<td>51.385469</td>
</tr>
<tr>
<td>St Louis</td>
<td>56 Trucks</td>
<td>168 Trucks</td>
<td>5817.839286</td>
</tr>
<tr>
<td></td>
<td>1.62</td>
<td>1.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min Travel Time</td>
<td>Min Travel Time</td>
<td>2139</td>
</tr>
<tr>
<td></td>
<td>0.59</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Variance of Travel Time</td>
<td>Variance of Travel Time</td>
<td>868.60</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>Std. Dev.</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>2 Std. Dev Time</td>
<td>2 Std. Dev Time</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td>Median Travel Time</td>
<td>Median Travel Time</td>
<td>6157.50</td>
</tr>
<tr>
<td></td>
<td>1.71</td>
<td>1.73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max Travel Time</td>
<td>Max Travel Time</td>
<td>8315</td>
</tr>
<tr>
<td></td>
<td>2.309722222</td>
<td>2.9344444444</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min Avg Speed</td>
<td>Min Avg Speed</td>
<td>30.973379</td>
</tr>
<tr>
<td></td>
<td>Max Avg. Speed</td>
<td>Max Avg. Speed</td>
<td>55.546389</td>
</tr>
<tr>
<td></td>
<td>Median Avg. Speed</td>
<td>Median Avg. Speed</td>
<td>41.6407905</td>
</tr>
</tbody>
</table>
Appendix 3: C Code for Data Analysis

The following C code was written by Christopher Chou ’03 to be used by John Knorring ’03.

// knorring_travel.cpp : Defines the entry point for the console application.
// by ctchou 3-28-03
// Program parses a list and selects certain combinations using an FSM
// Make sure that the end of file is at the last line!!!

#include System.Math
#include "stdafx.h"
#include <string>
#include <fstream>
#include <iostream>
#include <vector>
#include <windows.h>
#include <mmsystem.h>
#include <sys/types.h>
#include <sys/stat.h>
#include <io.h>
#include <stdio.h>
#include <stdlib.h>
#include <cmath>

using namespace std;
FILE *stream3;
FILE *stream4;

// converts a char to an int
int chartoint(char input) {
    switch(input) {
        case '0': return 0;
        case '1': return 1;
        case '2': return 2;
        case '3': return 3;
        case '4': return 4;
        case '5': return 5;
        case '6': return 6;
        case '7': return 7;
        case '8': return 8;
        case '9': return 9;
    }
}
int main(int argc, char* argv[])  
{  
    ifstream infile;  
    int line_count=0;  
    int length;  
    int state = 0;  
    int i,j;  
    int total_travel_time = 0;  
    int flag_3 = 0;  // went through 3  
    int flag_4 = 0;  // went through 4  
    char truck[6] = {0};  // old truck id  
    char new_truck[6];  // current truck id  
    char first_timer[13];  // starting time  
    char end_timer[13];  // ending time  
    char first_lat[25] = {0};  
    char first_long[25] = {0};  
    char last_lat[25] = {0};  
    char last_long[25] = {0};  
    string input;  
    basic_string<char>::size_type truck_index;  
    basic_string<char>::size_type truck_index_2;  
    int FD3;  // file descriptor for file containing route 3  
    int FD4;  // file descriptor for file containing route 4  
    int test_count = 0;  
    double distance_diff;  
    double RAD = .01745566;  

    // change data destination... 4 spots  
    FD3 = _creat( "Cincidata3.txt", _S_IREAD | _S_IWRITE );  
    if(FD3 == -1)  
        perror( "Couldn't create data file" );  
    else  
    {  
        printf( "Created data file.\n" );  
        _close(FD3);  
    }  

    FD4 = _creat( "Cincidata4.txt", _S_IREAD | _S_IWRITE );  
    if(FD4 == -1)  
        perror( "Couldn't create data file" );  
    else  
    {  
        printf( "Created data file.\n" );  
        _close(FD4);  
    }  
    stream3 = fopen("Cincidata3.txt","w");  
    stream4 = fopen("Cincidata4.txt","w");
do  //main loop does everything
{
    getline(infile, input);  // Read one line
    length = input.length();

    // Create FSM for finding patterns
    // Patterns: 132,231 and 142, 241
    route=input[length-1];

    // find the truck number
    truck_index = input.find_first_of(‘-‘,0);
    input.copy(new_truck, 6, truck_index+1);  // truck is the truck number (in chars)

    for(i=0;i<6;i++)  //compare trucks
    {
        if(truck[i] != new_truck[i])  //it's a new truck!
            state = 0;
        for(j=0;j<6;j++)  //set new truck
            truck[j] = new_truck[j];
        break;
    }
    //printf("\n");

    // test to see if first long first lat works
    *
    if(test_count < 10) {
        truck_index = input.find_first_of("gps",0);

        // record starting coord
        truck_index += 18;  //starts where long coord begin
        truck_index_2 = input.find_first_of(“ “, truck_index);  // till end of long
        input.copy(first_lat, (truck_index_2 - truck_index), truck_index);
        first_lat[truck_index_2 - truck_index] = ‘\0’;  //null end
        truck_index = truck_index_2+1;

        truck_index_2 = input.find_first_of(“ “, truck_index);  // till end of lat
        input.copy(first_long, (truck_index_2 - truck_index), truck_index);
        first_long[truck_index_2 - truck_index] = ‘\0’;  //null end
        printf("%s %s %s", first_lat, first_long);  //print long lat
    }
    test_count++;
*/

//FSM
switch(state) {
    case 0:  //start
        flag_3 = 0;

    case 1:
        //do something
    case 2:
        //do something
    case 3:
        //do something
    case 4:
        //do something
    case 5:
        //do something
    case 6:
        //do something
    case 7:
        //do something
    case 8:
        //do something
    case 9:
        //do something
    case 10:
        //do something
    case 11:
        //do something
    case 12:
        //do something
    case 13:
        //do something
    case 14:
        //do something
    case 15:
        //do something
    default:
        //do something
}
flag_4 = 0;
if(route == '1') { //matches 1-3-2 or 1-4-2
  state = 1;
  // record starting time
  truck_index = input.find_first_of("gps", 0);
  input.copy(first_timer, 13, truck_index + 4);
  // record starting coord
  truck_index += 18; // starts where long coord begin
  truck_index_2 = input.find_first_of(" ", truck_index); // till end of long
  input.copy(first_lat, (truck_index_2 - truck_index), truck_index);
  first_lat[truck_index_2 - truck_index] = '0'; // null end
  truck_index = truck_index_2 + 1;
  truck_index_2 = input.find_first_of(" ", truck_index); // till end of lat
  input.copy(first_long, (truck_index_2 - truck_index), truck_index);
  first_long[truck_index_2 - truck_index] = '0'; // null end
  break;
}
if(route == '2') { //matches 2-3-1 or 2-4-1
  state = 2;
  // record starting time
  truck_index = input.find_first_of("gps", 0);
  input.copy(first_timer, 13, truck_index + 4);
  // record starting coord
  truck_index += 18; // starts where long coord begin
  truck_index_2 = input.find_first_of(" ", truck_index); // till end of long
  input.copy(first_lat, (truck_index_2 - truck_index), truck_index);
  first_lat[truck_index_2 - truck_index] = '0'; // null end
  truck_index = truck_index_2 + 1;
  truck_index_2 = input.find_first_of(" ", truck_index); // till end of lat
  input.copy(first_long, (truck_index_2 - truck_index), truck_index);
  first_long[truck_index_2 - truck_index] = '0'; // null end
  break;
} else { // it begins with 3 or 4
  state = 0;
  break;
}
case 1: // matches 1-3-2 or 1-4-2
  if(route == '1') { // stay in this state, recopy truck index
    truck_index = input.find_first_of("gps", 0);
    input.copy(first_timer, 13, truck_index + 4);
    // record starting coord
    truck_index += 18; // starts where long coord begin
    truck_index_2 = input.find_first_of(" ", truck_index); // till end of long
    input.copy(first_lat, (truck_index_2 - truck_index), truck_index);
    first_lat[truck_index_2 - truck_index] = '0'; // null end
    truck_index = truck_index_2 + 1;
    truck_index_2 = input.find_first_of(" ", truck_index); // till end of lat
    input.copy(first_long, (truck_index_2 - truck_index), truck_index);
    first_long[truck_index_2 - truck_index] = '0'; // null end
    break;
  }
  if(route == '2') { // go back to 2
    truck_index = input.find_first_of("gps", 0);
    input.copy(first_timer, 13, truck_index + 4);
// record starting coord
truck_index += 18;  // starts where long coord begin
truck_index_2 = input.find_first_of(" ", truck_index); // till end of long
input.copy(first_lat, (truck_index_2 - truck_index), truck_index);
first_lat[truck_index_2 - truck_index] = ' \0'; // null end
truck_index = truck_index_2 + 1;
truck_index_2 = input.find_first_of(" ", truck_index); // till end of lat
input.copy(first_long, (truck_index_2 - truck_index), truck_index);
first_long[truck_index_2 - truck_index] = ' \0'; // null end
state = 2;
break;
}
if(route == '3') { //matches 1-3-2???
    flag_3 = 1;
    state = 3;
    break;
}
if(route == '4') { //matches 1-4-2??
    flag_4 = 1;
    state = 5;
    break;
} case 2: // matches 2-3-1 or 2-4-1
if(route == '2') { // stay in this state, recopy truck index
    truck_index = input.find_first_of("gps", 0);
    input.copy(first_timer, 13, truck_index+4);
    // record starting coord
    truck_index += 18;  // starts where long coord begin
    truck_index_2 = input.find_first_of(" ", truck_index); // till end of long
    input.copy(first_lat, (truck_index_2 - truck_index), truck_index);
    first_lat[truck_index_2 - truck_index] = ' \0'; // null end
    truck_index = truck_index_2 + 1;
    truck_index_2 = input.find_first_of(" ", truck_index); // till end of lat
    input.copy(first_long, (truck_index_2 - truck_index), truck_index);
    first_long[truck_index_2 - truck_index] = ' \0'; // null end
    break;
}
if(route == '1') { //go back to 1
    truck_index = input.find_first_of("gps", 0);
    input.copy(first_timer, 13, truck_index+4);
    // record starting coord
    truck_index += 18;  // starts where long coord begin
    truck_index_2 = input.find_first_of(" ", truck_index); // till end of long
    input.copy(first_lat, (truck_index_2 - truck_index), truck_index);
    first_lat[truck_index_2 - truck_index] = ' \0'; // null end
    truck_index = truck_index_2 + 1;
    truck_index_2 = input.find_first_of(" ", truck_index); // till end of lat
    input.copy(first_long, (truck_index_2 - truck_index), truck_index);
    first_long[truck_index_2 - truck_index] = ' \0'; // null end
    state = 1;
    break;
}
if(route == '3') { //matches 1-3-2???
    flag_3 = 1;
    state = 4;
    break;
if(route == '4') { //matches 1-4-2??
    flag_4 = 1;
    state = 6;
    break;
}
case 3: // matches 1-3-2
    if(route == '3') // stay in this state
        break;
    if(route == '2') { // we have 1-3-2! end
        state = 7;
        break;
    }
    if(route == '1') { // go back to state 1
        truck_index = input.find_first_of("gps", 0);
        input.copy(first_timer, 13, truck_index + 4);
        // record starting coord
        truck_index += 18; // starts where long coord begin
        truck_index_2 = input.find_first_of(" ", truck_index); // till end of long
        input.copy(first_lat, (truck_index_2 - truck_index), truck_index);
        first_lat[truck_index_2 - truck_index] = '\0'; // null end
        truck_index = truck_index_2 + 1;
        truck_index_2 = input.find_first_of(" ", truck_index); // till end of lat
        input.copy(first_long, (truck_index_2 - truck_index), truck_index);
        first_long[truck_index_2 - truck_index] = '\0'; // null end
        state = 1;
        break;
    } else // not a match it's 4
        state = 0;
    break;

case 4:// matches 2-3-1
    if(route == '3') // stay in this state
        break;
    if(route == '1') { // we have 2-3-1! end
        state = 7;
        break;
    }
    if(route == '2') { // go back to state 2
        truck_index = input.find_first_of("gps", 0);
        input.copy(first_timer, 13, truck_index + 4);
        // record starting coord
        truck_index += 18; // starts where long coord begin
        truck_index_2 = input.find_first_of(" ", truck_index); // till end of long
        input.copy(first_lat, (truck_index_2 - truck_index), truck_index);
        first_lat[truck_index_2 - truck_index] = '\0'; // null end
        truck_index = truck_index_2 + 1;
        truck_index_2 = input.find_first_of(" ", truck_index); // till end of lat
        input.copy(first_long, (truck_index_2 - truck_index), truck_index);
        first_long[truck_index_2 - truck_index] = '\0'; // null end
        state = 2;
        break;
    } else // not a match it's 4
state = 0;
break;

case 5: // matches 1-4-2
    if(route == '4') // stay in this state
        break;
    if(route == '2') { // we have 1-4-2! end
        state = 7;
        break;
    }
    if(route == '1') { // go back to state 1
        truck_index = input.find_first_of("gps", 0);
        input.copy(first_timer, 13, truck_index+4);
        // record starting coord
        truck_index += 18; // starts where long coord begin
        truck_index_2 = input.find_first_of(" ", truck_index); // till end of long
        input.copy(first_lat, truck_index_2 - truck_index, truck_index);
        first_lat[truck_index_2 - truck_index] = '0'; // null end
        truck_index = truck_index_2+1;
        truck_index_2 = input.find_first_of(" ", truck_index); // till end of lat
        input.copy(first_long, truck_index_2 - truck_index, truck_index);
        first_long[truck_index_2 - truck_index] = '0'; // null end
        state = 1;
        break;
    }
    else // not a match it's 3
        state = 0;
break;

case 6: // matches 2-4-1
    if(route == '4') // stay in this state
        break;
    if(route == '1') { // we have 2-4-1! end
        state = 7;
        break;
    }
    if(route == '2') { // go back to state 2
        truck_index = input.find_first_of("gps", 0);
        input.copy(first_timer, 13, truck_index+4);
        // record starting coord
        truck_index += 18; // starts where long coord begin
        truck_index_2 = input.find_first_of(" ", truck_index); // till end of long
        input.copy(first_lat, truck_index_2 - truck_index, truck_index);
        first_lat[truck_index_2 - truck_index] = '0'; // null end
        truck_index = truck_index_2+1;
        truck_index_2 = input.find_first_of(" ", truck_index); // till end of lat
        input.copy(first_long, truck_index_2 - truck_index, truck_index);
        first_long[truck_index_2 - truck_index] = '0'; // null end
        state = 1;
        break;
    }
    else // not a match it's 3
        state = 0;
break;

default:
    cout << "error! exiting program 1" << endl;
exit(1);
}

if(state == 7) { // found a pattern! get end time
    // set up next state!!!
    if(route == '1')
        state = 1; // matches 1-3-2 or 1-4-2
    if(route == '2')
        state = 2; // matches 2-3-1 or 2-4-1

    truck_index = input.find_first_of("gps", 0); // copy end time
    input.copy(end_timer, 13, truck_index+4); // record ending coord
    truck_index += 18; // starts where long coord begin
    truck_index_2 = input.find_first_of(" ", truck_index); // till end of long
    last_lat[truck_index_2 - truck_index] = '\0'; // null end
    truck_index = truck_index_2+1;
    truck_index_2 = input.find_first_of(" ", truck_index); // till end of lat
    input.copy(last_long, (truck_index_2 - truck_index), truck_index);
    last_long[truck_index_2 - truck_index] = '\0'; // null end

    // calculate distance C
    double distance_c = 0;
    // first convert from hundreds of degrees to degrees
    double first_lat_d, first_long_d, last_lat_d, last_long_d;
    double factor_x; // mult factor
    double distance_min;
    // re-initialize
    first_lat_d = 0;
    first_long_d = 0;
    last_lat_d = 0;
    last_long_d = 0;

    factor_x = 10;
    distance_min = 0;
    i=0;
    while(first_lat[i] != '.')
        i++;

    if(i==4) { // 4 digits
        for(i=0;i<2;i++) { // calc deg
            first_lat_d += factor_x * chartoint(first_lat[i]);
            factor_x /= 10;
        }
    }
    else { // 5 digits
        factor_x = 100;
        for(i=0;i<3;i++) { // calc deg
            first_lat_d += factor_x * chartoint(first_lat[i]);
            factor_x /= 10;
        }
    }

    factor_x = 10;
    while(first_lat[i] != '\0') { // calc deg plus minutes
if(first_lat[i] != '.') {
    distance_min += factor_x * chartoint(first_lat[i]);
    factor_x /= 10;
}
i++;

first_lat_d += (distance_min/60);
first_lat_d = first_lat_d * RAD;
factor_x = 10;
distance_min = 0;
i = 0;
while(first_long[i] != '.
    i++;

if(i==4) { // 4 digits
    for(i=0;i<2;i++) { // calc deg
        first_long_d += factor_x * chartoint(first_long[i]);
        factor_x /= 10;
    }
} else { // 5 digits
    factor_x = 100;
    for(i=0;i<3;i++) { // calc deg
        first_long_d += factor_x * chartoint(first_long[i]);
        factor_x /= 10;
    }
}

factor_x = 10;
while(first_long[i] != '0') { // calc deg plus minutes
    if(first_long[i] != '.') {
        distance_min += factor_x * chartoint(first_long[i]);
        factor_x /= 10;
    }
    i++;
}

first_long_d += (distance_min/60);
first_long_d = first_long_d * RAD;

factor_x = 10;
distance_min = 0;

i = 0;
while(last_lat[i] != '.
    i++;

if(i==4) { // 4 digits
    for(i=0;i<2;i++) { // calc deg
        last_lat_d += factor_x * chartoint(last_lat[i]);
        factor_x /= 10;
    }
} else { // 5 digits
    factor_x = 100;
    for(i=0;i<3;i++) { // calc deg
        last_lat_d += factor_x * chartoint(last_lat[i]);
        factor_x /= 10;
    }
}
factor_x=10;
while(last_lat[i] != '0') { // calc deg plus minutes
    if(last_lat[i] != '.') {
        distance_min += factor_x * chartoint(last_lat[i]);
        factor_x /= 10;
    }
    i++;
}
last_lat_d += (distance_min/60);
last_lat_d = last_lat_d * RAD;
factor_x=10;
distance_min = 0;
i=0;
while(last_long[i] != '.)
    i++;;
if(i==4) { // 4 digits
    for(i=0;i<2;i++) { // calc deg
        last_long_d += factor_x * chartoint(last_long[i]);
        factor_x /= 10;
    }
}
else { // 5 digits
    factor_x = 100;
    for(i=0;i<3;i++) { // calc deg
        last_long_d += factor_x * chartoint(last_long[i]);
        factor_x /= 10;
    }
}
factor_x=10;
while(last_long[i] != '0') { // calc deg plus minutes
    if(last_long[i] != '.') {
        distance_min += factor_x * chartoint(last_long[i]);
        factor_x /= 10;
    }
    i++;
}
last_long_d += (distance_min/60);
last_long_d = last_long_d * RAD;

// now we have long and lat in decimal form
// ***use great circle distance formula***
double abs_long, p1, p2, p3;
// abs takes in int, so we have to compare and choose
if(first_long_d>last_long_d) // do first minus last
    abs_long = first_long_d - last_long_d;
else
    abs_long = last_long_d - first_long_d;

//
    distance_c = 60* (acos((sin(first_lat_d)*sin(last_lat_d)) +
                     (cos(first_lat_d)*cos(last_lat_d)*cos(abs_long))));
p1 = sin(first_lat_d) * sin(last_lat_d);
p2 = cos(first_lat_d) * cos(last_lat_d);
// p3 = first_long_d + last_long_d;
// check to see if p3 > 180 if it is 360 - p3
    //
    // if(abs_long > 180)
    //
    // abs_long = 360 - abs_long;
    p3 = cos(abs_long);
distance_c = acos(p1 + p2 * p3);
distance_c = distance_c * 3956.316186;

// this is in nautical miles convert to miles
// not in naut miles anymore?
// distance_c = distance_c / 1.15;

// now convert times to seconds from 8/29/02
int hours = 0;
int minutes = 0;
int seconds = 0;
int day = 0;
int total_seconds_1 = 0, total_seconds_2 = 0;
int start_timer_sec = 0;
int finish_timer_sec = 0;

// convert start first
hours = 10 * chartoint(first_timer[0]);
hours += chartoint(first_timer[1]);
minutes = 10 * chartoint(first_timer[2]);
minutes += chartoint(first_timer[3]);
seconds = 10 * chartoint(first_timer[4]);
seconds += chartoint(first_timer[5]);
total_seconds_1 = (3600 * hours) + (60 * minutes) + seconds;
// find total seconds 2
if (first_timer[10] == '8') { // if month is 8
    day = 10 * chartoint(first_timer[7]);
day += chartoint(first_timer[8]);
total_seconds_2 = (day - 29) * 86400;
} else if (first_timer[10] == '9') { // month is 9
    day = 10 * chartoint(first_timer[7]);
day += chartoint(first_timer[8]);
total_seconds_2 = (day * 86400) + 259200;
} else { // error!
    cout << first_timer[10] << endl;
    cout << "error! Exiting 2!" << endl;
    exit(1);
}

start_timer_sec = total_seconds_1 + total_seconds_2;

// convert finish time
hours = 10 * chartoint(end_timer[0]);
hours += chartoint(end_timer[1]);
minutes = 10 * chartoint(end_timer[2]);
minutes += chartoint(end_timer[3]);
seconds = 10 * chartoint(end_timer[4]);
seconds += chartoint(end_timer[5]);
total_seconds_1 = (3600 * hours) + (60 * minutes) + seconds;
// find total seconds 2
if(end_timer[10] == '8') {// if month is 8
    day = 10*chartoint(end_timer[7]);
    day += chartoint(end_timer[8]);
    total_seconds_2 = (day - 29) * 86400;
}
else if(end_timer[10] == '9') { // month is 9
    day = 10*chartoint(end_timer[7]);
    day += chartoint(end_timer[8]);
    total_seconds_2 = (day * 86400) + 259200;
} else { // error!
    cout << "error! Exiting 3!";
    exit(1);
}

finish_timer_sec = total_seconds_1 + total_seconds_2;

total_travel_time = finish_timer_sec - start_timer_sec; // calculate total travel

// PRINT OUT DATA
// Format:
// Truck ID Start StartinSec End EndinSec traveltime avg_speed TAG

double avg_speed;

if(flag_3) { // stick it in file data3.txt
    // calculate avg speed
    avg_speed = distance_c/total_travel_time; // mps
    avg_speed = avg_speed*3600; // mph

    // write out truck ID
    for(i=0;i<6;i++)
        fprintf(stream3,"%c",truck[i]);
    fprintf(stream3," "); // formatting
    // print out start time
    fprintf(stream3,"%c",first_timer[0]);
    fprintf(stream3,"%c",first_timer[1]);
    fprintf(stream3,"%c",first_timer[2]);
    fprintf(stream3,"%c",first_timer[3]);
    fprintf(stream3,"%c",first_timer[4]);
    fprintf(stream3,"%c",first_timer[5]);
    fprintf(stream3,"%c",first_timer[6]); // space
    fprintf(stream3,"%c",first_timer[7]);
    fprintf(stream3,"%c",first_timer[8]);
    fprintf(stream3,"%c",first_timer[9]);
    fprintf(stream3,"%c",first_timer[10]);
    fprintf(stream3,"%c",first_timer[11]);
    fprintf(stream3,"%c",first_timer[12]);
    fprintf(stream3," "); // formatting
    // print out start timer in sec
    fprintf(stream3, "%d", start_timer_sec);
    fprintf(stream3," "); // formatting
// print out end time
fprintf(stream3, "%c", end_timer[0]);
fprintf(stream3, "%c", end_timer[1]);
fprintf(stream3, "%c");
fprintf(stream3, "%c", end_timer[2]);
fprintf(stream3, "%c", end_timer[3]);
fprintf(stream3, "%c");
fprintf(stream3, "%c", end_timer[4]);
fprintf(stream3, "%c", end_timer[5]);
fprintf(stream3, "%c", end_timer[6]); // space
fprintf(stream3, "%c", end_timer[7]);
fprintf(stream3, "%c", end_timer[8]);
fprintf(stream3, "%c");
fprintf(stream3, "%c", end_timer[9]);
fprintf(stream3, "%c", end_timer[10]);
fprintf(stream3, "%c");
fprintf(stream3, "%c", end_timer[11]);
fprintf(stream3, "%c", end_timer[12]);
fprintf(stream3, " "); // formatting
// print out end timer in sec
fprintf(stream3, "%d", finish_timer_sec);
fprintf(stream3, " "); // formatting
// print out travel time
fprintf(stream3, "%d", total_travel_time);
fprintf(stream3, " "); // formatting
// print out long lat coord
/*
fprintf(stream3, "%f", first_lat_d);
fprintf(stream3, " "); // formatting
fprintf(stream3, "%f", first_long_d);
fprintf(stream3, " "); // formatting
fprintf(stream3, "%f", last_lat_d);
fprintf(stream3, " "); // formatting
fprintf(stream3, "%f", last_long_d);
fprintf(stream3, " "); // formatting
// print out average speed
*/
fprintf(stream3, "%f", avg_speed);
fprintf(stream3, " "); // formatting
// print out distance c
// fprintf(stream3, "%f", distance_c);
// fprintf(stream3, " ");
// print out route
if(state == 1) // then it was 2-3-1
    fprintf(stream3, "231 
");
else
    fprintf(stream3, "132 
");
} else if(flag_4) { // stick it in file data4.txt
    // calculate avg speed
    // total distance = c+b-a
distance_c += distance_diff;
    avg_speed = distance_c/total_travel_time; // mps
    avg_speed = avg_speed*3600; // mph
    // write out truck ID
    for(i=0;i<6;i++)
        fprintf(stream4, "%c", truck[i]);
fprintf(stream4, " "); // formatting
// print out start time
fprintf(stream4, "%c", first_timer[0]);
fprintf(stream4, "%c", first_timer[1]);
fprintf(stream4, ":");
fprintf(stream4, "%c", first_timer[2]);
fprintf(stream4, "%c", first_timer[3]);
fprintf(stream4, ":");
fprintf(stream4, "%c", first_timer[4]);
fprintf(stream4, "%c", first_timer[5]);
fprintf(stream4, "%c", first_timer[6]); // space
fprintf(stream4, "%c", first_timer[7]);
fprintf(stream4, "%c", first_timer[8]);
fprintf(stream4, ":");
fprintf(stream4, "%c", first_timer[9]);
fprintf(stream4, "%c", first_timer[10]);
fprintf(stream4, ":");
fprintf(stream4, "%c", first_timer[11]);
fprintf(stream4, "%c", first_timer[12]);
fprintf(stream4, " "); // formatting
// print out start timer in sec
fprintf(stream4, "%d");
fprintf(stream4, " "); // formatting
// print out end time
fprintf(stream4, "%c", end_timer[0]);
fprintf(stream4, "%c", end_timer[1]);
fprintf(stream4, ":");
fprintf(stream4, "%c", end_timer[2]);
fprintf(stream4, "%c", end_timer[3]);
fprintf(stream4, ":");
fprintf(stream4, "%c", end_timer[4]);
fprintf(stream4, "%c", end_timer[5]);
fprintf(stream4, "%c", end_timer[6]); // space
fprintf(stream4, "%c", end_timer[7]);
fprintf(stream4, "%c", end_timer[8]);
fprintf(stream4, ":");
fprintf(stream4, "%c", end_timer[9]);
fprintf(stream4, "%c", end_timer[10]);
fprintf(stream4, ":");
fprintf(stream4, "%c", end_timer[11]);
fprintf(stream4, "%c", end_timer[12]);
fprintf(stream4, " "); // formatting
// print out end timer in sec
fprintf(stream4, "%d");
fprintf(stream4, " "); // formatting
// print out travel time
fprintf(stream4, "%d"); // print out total travel time
fprintf(stream4, ") "); // formatting
// print out long lat coord

/*
fprintf(stream4, "f", first_lat_d);
fprintf(stream4, "); // formatting
fprintf(stream4, "f", first_long_d);
fprintf(stream4, "); // formatting
fprintf(stream4, "f", last_lat_d);
*/
fprintf(stream4," "); // formatting
fprintf(stream4,"%f", last_long_d);
fprintf(stream4," "); // formatting
*/   // print out avg speed
fprintf(stream4,"%f", avg_speed);
fprintf(stream4," "); // formatting
//print out distance c
fprintf(stream4,"%f", distance_c);
//    fprintf(stream4," ");
//print out route
if(state == 1) // then it was 2-4-1
    fprintf(stream4,"241\n");
else
    fprintf(stream4,"142\n");
}
else {   //error
    cout << "error! Exit 4." << endl;
    exit(1);
}
flag_3 = 0; // reset flags
flag_4 = 0; //reset flags

}
line_count++;
//cout << line_count; // if you want, output total, may be problems, change to double
}while(infile.eof()==0);

infile.close();
fclose(stream3);
fclose(stream4);
//End Code for reading files

return 0;
}
Appendix 4: GPS Box Coordinates

### 90/94 Case Study

<table>
<thead>
<tr>
<th>Billings, MT</th>
<th>Tomah, WI</th>
<th>North Route 94</th>
<th>South Route 90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box 1</td>
<td>Box 2</td>
<td>Box 3</td>
<td>Box 4</td>
</tr>
<tr>
<td>West</td>
<td>Box 1</td>
<td>Box 2</td>
<td>Box 3</td>
</tr>
<tr>
<td>110.093046</td>
<td>108.507887</td>
<td>90.440054</td>
<td>104.000000</td>
</tr>
<tr>
<td>North</td>
<td>45.890805</td>
<td>45.575411</td>
<td>44.013855</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>1005.7 Miles</td>
<td>1039.5 Miles</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>15:29 hours</td>
<td>16:01 hours</td>
<td></td>
</tr>
</tbody>
</table>

| DISTANCE 1005.7 Miles | 1039.5 Miles |
| Time 15:29 hours | 16:01 hours |

### Chicago Skyway/ Dan Ryan Case Study

<table>
<thead>
<tr>
<th>North</th>
<th>South</th>
<th>80/94</th>
<th>Skyway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box 1</td>
<td>Box 2</td>
<td>Box 3</td>
<td>Box 4</td>
</tr>
<tr>
<td>West</td>
<td>87.632482</td>
<td>87.629815</td>
<td>87.200758</td>
</tr>
<tr>
<td>North</td>
<td>41.787799</td>
<td>41.784951</td>
<td>41.599407</td>
</tr>
<tr>
<td>DISTANCE 34.4 Miles</td>
<td>30.7 Miles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 33 minutes</td>
<td>30 minutes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Cincinnati, OH Case Study

<table>
<thead>
<tr>
<th>North</th>
<th>South</th>
<th>Downtown (I-75)</th>
<th>Bypass (I-275)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box 1</td>
<td>Box 2</td>
<td>Box 3</td>
<td>Box 4</td>
</tr>
<tr>
<td>West</td>
<td>84.439012</td>
<td>84.356588</td>
<td>84.659528</td>
</tr>
<tr>
<td>DISTANCE 24.6 Miles</td>
<td>41 Miles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 26 minutes</td>
<td>39 minutes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Columbus, OH Case Study

<table>
<thead>
<tr>
<th>East</th>
<th>West</th>
<th>Downtown (I-70)</th>
<th>Bypass (I-275)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box 1</td>
<td>Box 2</td>
<td>Box 3</td>
<td>Box 4</td>
</tr>
<tr>
<td>West</td>
<td>82.713078</td>
<td>82.567703</td>
<td>83.291066</td>
</tr>
<tr>
<td>DISTANCE 17.4 Miles</td>
<td>22.1 Miles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 17 minutes</td>
<td>21 minutes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Houston, TX Case Study

<table>
<thead>
<tr>
<th>East</th>
<th>West</th>
<th>Downtown (I-10)</th>
<th>Bypass (I-610)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box 1</td>
<td>Box 2</td>
<td>Box 3</td>
<td>Box 4</td>
</tr>
<tr>
<td>West</td>
<td>95.246656</td>
<td>95.167986</td>
<td>95.677268</td>
</tr>
<tr>
<td>North</td>
<td>29.789795</td>
<td>29.763296</td>
<td>29.814638</td>
</tr>
<tr>
<td>DISTANCE 13.2 Miles</td>
<td>15.3 Miles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 13 minutes</td>
<td>16 minutes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Indianapolis, IN Case Study

<table>
<thead>
<tr>
<th>Direction</th>
<th>East</th>
<th>West</th>
<th>Downtown (I-70)</th>
<th>Bypass (I-74)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### West
- Distance: 22.6 Miles
- Time: 22 minutes

#### North
- Distance: 22.6 Miles
- Time: 22 minutes

### Memphis, TN Case Study

<table>
<thead>
<tr>
<th>Direction</th>
<th>East</th>
<th>West</th>
<th>Downtown (I-40)</th>
<th>Bypass (I-240)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### West
- Distance: 26 Miles
- Time: 25 minutes

#### North
- Distance: 26 Miles
- Time: 25 minutes

### Nashville, TN Case Study

<table>
<thead>
<tr>
<th>Direction</th>
<th>East</th>
<th>West</th>
<th>Downtown (I-40)</th>
<th>Bypass (I-440)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### West
- Distance: 9.2 Miles
- Time: 9 minutes

#### North
- Distance: 9.2 Miles
- Time: 9 minutes

### Oklahoma City, OK Case Study

<table>
<thead>
<tr>
<th>Direction</th>
<th>East</th>
<th>West</th>
<th>Downtown (I-40)</th>
<th>Bypass (I-240)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### West
- Distance: 22.1 Miles
- Time: 23 minutes

#### North
- Distance: 22.1 Miles
- Time: 23 minutes

### Richmond, VA Case Study

<table>
<thead>
<tr>
<th>Direction</th>
<th>North</th>
<th>South</th>
<th>Downtown (I-95)</th>
<th>Bypass (I-295)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### West
- Distance: 44.8 Miles
- Time: 43 minutes

#### North
- Distance: 44.8 Miles
- Time: 43 minutes
### San Antonio, TX Case Study

<table>
<thead>
<tr>
<th>Box 1</th>
<th>Box 2</th>
<th>Box 3</th>
<th>Box 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>West</td>
<td>98.382277</td>
<td>98.338681</td>
<td>98.767167</td>
</tr>
<tr>
<td>North</td>
<td>29.567254</td>
<td>29.536573</td>
<td>29.283164</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>22.4 Miles</td>
<td>27.3 Miles</td>
<td>22 minutes</td>
</tr>
</tbody>
</table>

### St. Louis, MO Case Study

<table>
<thead>
<tr>
<th>Box 1</th>
<th>Box 2</th>
<th>Box 3</th>
<th>Box 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>West</td>
<td>89.924548</td>
<td>89.805880</td>
<td>90.403872</td>
</tr>
<tr>
<td>North</td>
<td>38.929531</td>
<td>38.781778</td>
<td>38.481619</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>23.5 Miles</td>
<td>25.5 Miles</td>
<td>24 minutes</td>
</tr>
</tbody>
</table>

### Wilmington Delaware Case Study

<table>
<thead>
<tr>
<th>Box 1</th>
<th>Box 2</th>
<th>Box 3</th>
<th>Box 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>West</td>
<td>75.699796</td>
<td>75.597327</td>
<td>75.434093</td>
</tr>
<tr>
<td>North</td>
<td>39.703682</td>
<td>39.652517</td>
<td>39.886265</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>14.0 Miles</td>
<td>14.3 Miles</td>
<td>14 minutes</td>
</tr>
</tbody>
</table>
Appendix 5: Route Maps

Refer to Appendix 4 for route lengths and times.

7.2.1 90-94
7.2.2 Cincinnati, OH
7.2.3 Columbus, OH
7.2.4 Houston, TX
7.2.5 Indianapolis, IN
7.2.6 Memphis, TN
7.2.7 Nashville, TN
7.2.8 Oklahoma City, OK
7.2.9 Richmond, VA
7.2.10 San Antonio, TX
7.2.11 St. Louis, MO
7.2.12 Wilmington, DE