ESTIMATION OF TRAVEL TIME DISTRIBUTION
AND TRAVEL TIME DERIVATIVES

KE WAN

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
OPERATIONS RESEARCH AND FINANCIAL ENGINEERING
ADVISER: ALAIN L. KORNAUSER

FEB 2013
Abstract

With the complexity of transportation systems, generating optimal routing decisions is a critical issue. This thesis focuses on how routing decisions can be made by considering the distribution of travel time and the associated risk. More specifically, the routing decision process is modeled in a way that explicitly considers the dependence between travel time among different links and the risk associated with the volatility of travel time. Furthermore, the computation of this volatility allows for the development of the travel time derivative, which is the financial derivative based on travel time. It serves as a value pricing (also known as congestion pricing) scheme that is based on not only the level of congestion but also its uncertainties. In addition to the introduction (Chapter 1), the literature review (Chapter 2), and the conclusion (Chapter 6), the thesis consists of two major parts:

In the first part (Chapters 3 and 4), the travel time distribution for transportation links and paths, conditioned on the latest observations, is estimated to enable routing decisions based on risk. Chapter 3 sets up the basic decision framework by modeling the dependent structure between the travel time distributions for nearby links using the copula method. In Chapter 4, the framework is generalized to estimate the travel time distribution for a given path using Gaussian copula mixture models (GCMM). To explore more information from fundamental traffic conditions, a scenario-based GCMM is studied: A distribution of the path scenario (systematic factor) is defined first; the dependent structure between constructing links in the path is modeled as a Gaussian copula for each scenario, and the scenario-wise path travel time distribution can be obtained based on this copula. The final estimates are calculated by integrating the scenario-wise path travel time distributions over the distribution of the path scenario. In a discrete setting, it is a weighted sum of these conditional travel time distributions. The property of the scenario-based GCMM is studied when the number of scenarios changes. Furthermore, general GCMM are introduced for better finite scenario performance, and extended expectation-maximum algorithms are designed to estimate the model parameters, which introduces an innovative copula-based machine learning method.
In the second part (Chapter 5), travel time derivatives are introduced as financial derivatives based on road travel time - a non-tradable underlying asset - and a more flexible alternative for value pricing. The chapter addresses (a) the necessity of introducing such derivatives (that is, the demand for hedging), (b) the market and design of the product, and (c) the pricing schemes. The pricing schemes are designed based on the travel time data from loop detectors, which are modeled as mean reverting processes driven by diffusion processes. The no-arbitrage principle is used to generate the price.
Acknowledgements

Thanks to my adviser Professor Kornhauser. He guided me into the topic of this dissertation and shared with me all his expertise in transportation and decision theory. He also shared with me his insight about research and life. "He who teaches me will be my father-figure for life."

Thanks to my parents Xueqing Wan and Mingyu Deng for your understanding and encouragement. This dissertation is presented to you.

Many thanks to all of the anonymous references for their helpful suggestions which help to make this thesis.
## Contents

Abstract .................................................................................................................. iii

Acknowledgements ................................................................................................. v

1 Introduction ........................................................................................................... 9

   1.1 Motivation ......................................................................................................... 9

   1.2 Problem statement and thesis objective ......................................................... 14

      1.2.1 Transportation network and travel time measurement ......................... 15

      1.2.2 Travel time ............................................................................................... 16

      1.2.3 Decision framework .................................................................................. 18

      1.2.4 Major issues of interest ............................................................................. 19

   1.3 Thesis outline .................................................................................................. 25

2 Literature Review .................................................................................................. 26

   2.1 Travel time estimation ..................................................................................... 26

      2.1.1 Link travel time estimation ...................................................................... 27

      2.1.2 Path travel time estimation ..................................................................... 30
2.2 Dependent structure and copula theory ........................................... 32
  2.2.1 Definitions .............................................................................. 32
  2.2.2 Fundamental theorem of copula .............................................. 34
  2.2.3 Tail dependence of the copula .................................................. 35
  2.2.4 Nonparametric density estimation and conditional density estimator . 39
2.3 Derivative pricing ................................................................. 42
  2.3.1 Derivative fundamentals .......................................................... 43
  2.3.2 Probability settings of derivative pricing ................................. 46

3 Link Travel Time Estimation and Routing Decisioning through Copula

Methods .................................................................................................. 50
  3.1 Profile description of link travel time data ..................................... 51
  3.2 Copula based travel time distribution generation ............................ 52
    3.2.1 Copula and two step maximum likelihood estimation ................. 52
  3.3 Routing decision based on the estimated travel time distribution .......... 56
  3.4 Numerical analysis ....................................................................... 63
  3.5 Further development: reliable estimates and similarity-based analysis . 75
    3.5.1 Generation of reliable estimation ............................................. 76
    3.5.2 Similarity-based copula reconstruction .................................... 78

4 Path Travel Time Distribution Estimation through Gaussian Copula Mix-
ture Models .......................................................................................... 81
4.1 The Gaussian Copula Mixture Model (GCMM) ........................................ 82
  4.1.1 Definition of a Gaussian copula mixture model ............................ 82
  4.1.2 Estimation of the path travel time distribution based on a given GCMM model ................................................................. 84

4.2 A Gaussian copula mixture model based on predefined link travel time distributions ........................................................................................................... 85
  4.2.1 Scenario decomposition and summation ....................................... 88
  4.2.2 Scenario-specific estimation of conditional path travel time ............ 91
  4.2.3 Properties of the scenario-based GCMM model ............................ 97

4.3 A general GCMM and the extended expectation maximum algorithms .... 110
  4.3.1 Fixed-marginal-distribution GCMM ............................................ 111
  4.3.2 Varying-marginal-distribution GCMM ........................................ 115

4.4 Numerical analysis ........................................................................... 121
  4.4.1 Experiments for scenario-based GCMM models ......................... 121
  4.4.2 Experiment for general GCMM models ...................................... 125

5 Travel Time Derivatives: Market Analysis and Pricing ............................ 132
  5.1 Initiation and necessity analysis ..................................................... 132
    5.1.1 Derivative and weather derivative as a hedging tool .................... 132
    5.1.2 Travel time derivatives is a flexible congestion pricing scheme ...... 134
    5.1.3 Travel time derivatives can diversify risk for financial market ...... 139
6 Conclusions and Future Work
# List of Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\Omega, F, P)$</td>
<td>A standard probability space</td>
<td>9</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>The sample space</td>
<td>9</td>
</tr>
<tr>
<td>$F$</td>
<td>The $\sigma$ algebra</td>
<td>9</td>
</tr>
<tr>
<td>$P$</td>
<td>The ordinary probability measure</td>
<td>9</td>
</tr>
<tr>
<td>$t$</td>
<td>Time or the index set.</td>
<td>9</td>
</tr>
<tr>
<td>${F_t}_{t&gt;0}$</td>
<td>The filtration</td>
<td>9</td>
</tr>
<tr>
<td>$r.v$</td>
<td>A random variable</td>
<td>9</td>
</tr>
<tr>
<td>$X(\omega)$</td>
<td>The general notation for a random variable</td>
<td>9</td>
</tr>
<tr>
<td>$x$</td>
<td>A specific value that the random variables takes.</td>
<td>9</td>
</tr>
<tr>
<td>$X_t$</td>
<td>The general notation for a stochastic process</td>
<td>9</td>
</tr>
<tr>
<td>$E(X)$</td>
<td>The expectation of a random variable $X$</td>
<td>9</td>
</tr>
<tr>
<td>$Var(X)$</td>
<td>The variance of a random variable $X$</td>
<td>9</td>
</tr>
<tr>
<td>$ATT$</td>
<td>Average Travel Time</td>
<td>??</td>
</tr>
<tr>
<td>$n$</td>
<td>Serial number, end with its capital case</td>
<td>??</td>
</tr>
<tr>
<td>$T_t$</td>
<td>Travel time for a link given a time $t$</td>
<td>??</td>
</tr>
<tr>
<td>$T_m$</td>
<td>The time instant when a vehicle passes a given monument</td>
<td>65</td>
</tr>
<tr>
<td>$T_c$</td>
<td>The time instant when a vehicle passes the center of a link</td>
<td>66</td>
</tr>
<tr>
<td>$\mu$</td>
<td>The mean value of a random variable</td>
<td>31</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>The standard deviation of a r.v and a volatility of a process</td>
<td>31</td>
</tr>
<tr>
<td>$CV$</td>
<td>Coefficient of variation</td>
<td>31</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>$p$</td>
<td>A path</td>
<td></td>
</tr>
<tr>
<td>$l$</td>
<td>A link</td>
<td></td>
</tr>
<tr>
<td>$L(p)$</td>
<td>The links which forms a given path $p$</td>
<td></td>
</tr>
<tr>
<td>$H$</td>
<td>A general univariate/multivariate functions</td>
<td></td>
</tr>
<tr>
<td>$V$</td>
<td>A value function</td>
<td></td>
</tr>
<tr>
<td>$C'$</td>
<td>A sub copula function</td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td>A copula function</td>
<td></td>
</tr>
<tr>
<td>$F(x)$</td>
<td>A univariate cumulative probability function for a r.v, similar to $G(x)$</td>
<td></td>
</tr>
<tr>
<td>$\lambda_U$</td>
<td>The upper tail dependence of a copula</td>
<td></td>
</tr>
<tr>
<td>$\lambda_L$</td>
<td>The lower tail dependence of a copula</td>
<td></td>
</tr>
<tr>
<td>$f$</td>
<td>The density function for the random variables.</td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>Spearman’s $\rho$, a dependent measure</td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>Kendall’s $\tau$, a dependent measure</td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>Parameters of copulas, the same applies to $\theta, \delta, \gamma, \kappa$</td>
<td></td>
</tr>
<tr>
<td>$K(x)$</td>
<td>A univariate kernel function.</td>
<td></td>
</tr>
<tr>
<td>$W(x)$</td>
<td>A univariate kernel function.</td>
<td></td>
</tr>
<tr>
<td>$A$</td>
<td>The transition matrix of a hidden markov chain.</td>
<td></td>
</tr>
<tr>
<td>$B$</td>
<td>The observation matrix of a hidden markov chain</td>
<td></td>
</tr>
<tr>
<td>$\Pi$</td>
<td>The initial probability distribution of a hidden markov chain</td>
<td></td>
</tr>
<tr>
<td>$\pi$</td>
<td>The weight of mixture model</td>
<td></td>
</tr>
<tr>
<td>$\Psi$</td>
<td>The cumulative probability function of standard normal distribution.</td>
<td></td>
</tr>
<tr>
<td>$\psi$</td>
<td>The density function of standard normal distribution</td>
<td></td>
</tr>
<tr>
<td>$Q$</td>
<td>The risk neutral measure</td>
<td></td>
</tr>
<tr>
<td>$S_t$</td>
<td>The stock price</td>
<td></td>
</tr>
<tr>
<td>$N_t$</td>
<td>The jump measure/poisson random measure</td>
<td></td>
</tr>
<tr>
<td>$L_t$</td>
<td>The levy process</td>
<td></td>
</tr>
<tr>
<td>$\mu(x,y)$</td>
<td>The measure for a pair of random variables</td>
<td></td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>$K(t, dT)$</td>
<td>the transitional kernel/conditional probability density of $T$, given $t$</td>
<td></td>
</tr>
<tr>
<td>$B_t^H$</td>
<td>The fractional Brownian motion</td>
<td></td>
</tr>
<tr>
<td>$B_t$</td>
<td>The Brownian motion</td>
<td></td>
</tr>
<tr>
<td>$\rho(X)$</td>
<td>The transportation risk measure</td>
<td></td>
</tr>
<tr>
<td>$U(x)$</td>
<td>The utility function for a given value $x$</td>
<td></td>
</tr>
<tr>
<td>$\text{VAR}(X)$</td>
<td>Value of risk, given a random variable</td>
<td></td>
</tr>
<tr>
<td>$L$</td>
<td>Log likelihood function</td>
<td></td>
</tr>
<tr>
<td>$P$</td>
<td>The dependent parameter for a Gaussian copula</td>
<td></td>
</tr>
<tr>
<td>$S$</td>
<td>The current scenario for a path</td>
<td></td>
</tr>
<tr>
<td>$\hat{\pi}$</td>
<td>The scenario frequency after consider the pseudo observations</td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>The lagrange multiplier</td>
<td></td>
</tr>
<tr>
<td>$G$</td>
<td>Payoff when buying travel time derivatives</td>
<td></td>
</tr>
<tr>
<td>$P$</td>
<td>The price of travel time derivatives</td>
<td></td>
</tr>
<tr>
<td>$K$</td>
<td>The strike of options</td>
<td></td>
</tr>
<tr>
<td>$T_{us}$</td>
<td>The standard country travel time index for USA</td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>The market value of risk</td>
<td></td>
</tr>
<tr>
<td>$v$</td>
<td>The value function or the value of a claim</td>
<td></td>
</tr>
</tbody>
</table>
List of Figures

1.1 Predict the conditional distribution of travel time on downstream links . . . 20

1.2 The empirical path travel time distribution is different from the estimation based on independence assumption (red: estimated distribution under the assumption of independence; green: empirical distribution) . . . . . . . . . . 21

1.3 Assembly the path travel time distribution from link data . . . . . . . . . . . 22

1.4 Price the travel time derivatives and make routing decisions according to price 24

3.1 There are more data records on link AB than any path that goes through it(ABC or ABD) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52

3.2 Test network N95-US 195, the red frames are the targeted monuments . . . . 64

3.3 Fit travel time distribution on AB . . . . . . . . . . . . . . . . . . . . . . . 68

3.4 Empirical distribution v.s. simulated data based on different copulas . . . . . 69

3.5 The comparison of tails of the joint structure: Red lower tail, Blue upper tail 71

3.6 Estimation of copula and conditional probability function of AB based on bb1 copula . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 73
3.7 Estimated conditional density (Red) v.s. Empirical conditional density (Blue: histogram at 1.2 miles/hour, Black: kernel smoothing) given current observation 67 mile/hour per hour at a loop detector ........................................ 75

3.8 Combined Conditional probability density function (Green) compare to two original estimation by T copula (Red) and by bb1 copula (Blue) ............... 77

3.9 Difference in estimation between the borrowed copula (red) and the original copula (blue)) Empirical pdf (blue); Estimated pdf (red) ..................... 79

4.1 Definition of path scenarios ...................................................... 86

4.2 Estimate the copula in each path scenario ................................. 87

4.3 Aggregate path scenarios according to their historical frequency .......... 87

4.4 Generate pseudo observations .................................................. 94

4.5 The experimental network in New Jersey .................................... 121

4.6 Change of estimation as $\Delta_i$ changes (Empirical: red dots; $0.5\sigma_i$: cyan; $\sigma_i$: red; $2\sigma_i$: blue; $3\sigma_i$: black) ..................................................... 122

4.7 Change of estimation as the Lasso penalty changes. (Empirical: red dot; Independent: yellow; $v=0.001$: Cyan; $v=0.01$: red; $v=0.05$: Blue; $v=0.1$: Black; $v=0.5$: Green) ..................................................... 123

4.8 Estimated cdf when number of scenarios change: 1: yellow; 2: red; 4: cyan; 8: green; 16: blue; empirical: black; independent: black dots; Left figure displays the lower tail, Right figure displays the upper tail ................. 123

4.9 Left: 2 scenarios per link; Right: 8 scenarios per link; Estimated pdf: red; empirical: green; scenario-specific cdfs: blue ................................. 124

4.10 12 Path travel time based on the approximation; red: Path 1; blue: Path 2 . 125
4.11 Comparison of the Marginal scenario decomposition for the two links . . . . 129

4.12 Comparison of the copula structure for the two links . . . . . . . . . . . . . 130

4.13 Comparison of path travel time distributions estimated through different models: Black: empirical; Red: Model 1; Green: Model 2; Magenta: Model 3; Blue Model 4; Cyan: Model 5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 131

5.1 Travelers and travel time(QOS) protection - good scenario . . . . . . . . 135

5.2 Travelers and travel time(QOS) protection - bad scenario . . . . . . . . . . 135

5.3 The price of travel time derivatives predicts short term traffic conditions . . 137

5.4 The price of travel time derivatives predicts long term traffic conditions . . 138

5.5 Alternative risk transfer between transportation related industry and financial industry . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 139

5.6 Experiment Network Graph . . . . . . . . . . . . . . . . . . . . . . . . . . 153

5.7 Remove the trend, weekly, daily part of 80E data series . . . . . . . . . . 154

5.8 80E ARIMA model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 155
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Definition of copula functions</td>
<td>37</td>
</tr>
<tr>
<td>2.2</td>
<td>Tail dependence for typical copulas</td>
<td>37</td>
</tr>
<tr>
<td>3.1</td>
<td>Definition of distance measures</td>
<td>55</td>
</tr>
<tr>
<td>3.2</td>
<td>Tail dependence for typical copulas</td>
<td>55</td>
</tr>
<tr>
<td>3.3</td>
<td>Definition of monument links</td>
<td>64</td>
</tr>
<tr>
<td>3.4</td>
<td>Distance measures for parametric estimators</td>
<td>67</td>
</tr>
<tr>
<td>3.5</td>
<td>Goodness-of-fit for Copulas</td>
<td>68</td>
</tr>
<tr>
<td>3.6</td>
<td>Decision statistics for different rules</td>
<td>73</td>
</tr>
<tr>
<td>3.7</td>
<td>Deviation of estimates in cross validation</td>
<td>74</td>
</tr>
<tr>
<td>3.8</td>
<td>L2 distance of estimation with empirical conditional distribution</td>
<td>74</td>
</tr>
<tr>
<td>3.9</td>
<td>Parameter estimates for T copula</td>
<td>77</td>
</tr>
<tr>
<td>3.10</td>
<td>Decision statistics for mean-variance rule on AB by different estimation pro-</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>cedures</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Decision statistics for different rules for the two paths</td>
<td>125</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.1</td>
<td>Payoff of typical weather derivatives</td>
<td>134</td>
</tr>
<tr>
<td>5.2</td>
<td>Payoff of traditional road toll, $P$ denotes the toll amount</td>
<td>136</td>
</tr>
<tr>
<td>5.3</td>
<td>Payoff of dynamic congestion pricing, $P$ denotes the toll amount</td>
<td>136</td>
</tr>
<tr>
<td>5.4</td>
<td>Payoff of a travel time derivatives, $P$ denotes its price</td>
<td>136</td>
</tr>
<tr>
<td>5.5</td>
<td>Comparison between travel time insurance and travel time derivatives</td>
<td>139</td>
</tr>
<tr>
<td>5.6</td>
<td>Different participants</td>
<td>141</td>
</tr>
<tr>
<td>5.7</td>
<td>Seasonal effects in the 80E data</td>
<td>153</td>
</tr>
<tr>
<td>5.8</td>
<td>ARIMA modeling of the residuals</td>
<td>153</td>
</tr>
<tr>
<td>5.9</td>
<td>Model selection for 80E data</td>
<td>154</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Motivation

Transportation systems are complex structures with substantially varied performance, much of which results from the cumulative effects of many individuals’ behaviors. When individuals compete for the use of limited traffic flow of varying capacity, their interactions yield the variability in the performance of the traffic system, which in turn yields uncertainties in travel time.

In everyday life, uncertainties in travel time have a substantial impact on travelers, especially when many people try to reach their destinations at the same time, usually during peak hours. Drivers are accustomed to everyday congestion, and so they plan for it. Most leave home early enough to reach their destinations on time. Unexpected congestion, however, is more likely to trouble travelers. When the expected amount of travel time takes much longer, people complain. Thus, the reliability of travel time is of most concern. Travelers should make decisions based on not only expectations of travel time but also its uncertainties. Many do not like to be late, and they tend to be risk reverse. Consequently, their perceptions of travel time reliability, which identifies the individuality of travel time, are of
particular concern. Certain questions require answering to address the risk in traffic systems, such as "How bad can the traffic condition be?" and "What is the stressed travel time scenario that occurs once ‘every thousand’ days?” To answer such questions is the objective of this research. In order to find ways to reduce uncertainties, it is necessary to characterize and to quantify the magnitude of the uncertainties. Such research can deliver substantial benefits, which include but are not limited to the following: (a) It characterizes the dynamics of travelers’ experiences; (b) it helps to generate routing suggestions that can minimize the risk for travelers; and (c) it leads to innovative value pricing schemes, which link road tolls directly to travelers’ expectations of service quality. All of these potential benefits can affect travelers’ behaviors, and if properly implemented, can effectively improve the performance of traffic system as well.

This thesis focuses on the modeling and prediction of travel time of vehicles in a transportation network. Specific reference is made to imitate vehicles traveling in real roadway networks. The goal is to be able to compute and disseminate travel time predictions continuously and in real time, so as to allow individual drivers to make dynamic route choices that improve travel performance. The thesis focuses on improving the travel time for any feasible route and does not address the system-optimal problem of improving the overall travel time of all travelers. A second objective of this thesis is to focus on the volatility of travel time and its implication for the development of external pricing of related financial derivatives with appropriate hedging strategies. For this purpose, the concept of a travel time derivative is developed. The thesis is presented in two parts.

In the first part of the thesis, a suitable methodology is designed to describe and predict uncertainties in travel time. The probability distribution function describing these uncertainties, especially its tail, is a fundamental characterization of risk. To estimate uncertainties in travel time, it is critical to assess the extent to which the uncertainties are related not only to what is happening on a section of the network but also to its extensions in neighboring sections. One approach based on the copula method is developed in this thesis. With cop-
ulas, the dependent structure between links of a transportation network can be described, and observations of real links can be used to estimate the inter-dependence of travel times on different links. The travel time distribution on a specific link can be then estimated using the travel time observations on nearby links based on estimated copula structures. Furthermore, copulas can help solve data insufficiency problems caused by uneven utilization of transportation resources in different network spots. A copula can be calibrated with the abundant data available at one intersection and applied to another intersection with similar physical conditions but insufficient data.

The copula method is also helpful as a method by which individual link travel time distributions can be aggregated to estimate a path travel time distribution. To conduct such an aggregation, one can assume the travel times between different links are independent. However, this approach is unrealistic, for a transportation network as the dependent structure between the travel times of different links cannot be ignored. In this thesis, copulas are used to model dependent structure given the traffic conditions along the path. More specifically, it is assumed there are several path scenarios of the overall path traffic conditions. For each path scenario, a copula is used to model the dependence between the links in the path. Using the copula, a conditional path travel time distribution is estimated for each scenario. The unconditional path travel time distribution can be obtained by taking weighted sum of these conditional path travel time distributions.

The research on travel time estimation using copula models is a timely topic because information and data systems that support analytical approaches and address uncertainties in travel time are becoming available through a growing focus on what has become known as intelligent transportation systems (ITS). The increasing reliance on ITS has heightened the need to estimate and predict travel times accurately and reliably. Turn-by-turn route guidance has been enabled, which allows road users to make more informed route decisions (pre-trip and en-route), and that planning can potentially yield more stable and less congested traffic conditions.
In the second part of the thesis, the economic coupling between travel time and travelers is studied.

Traditionally, value pricing has been used as an incentive to change usual consumption of transportation resources by pricing the use of those resources according to change in demand from over-capacity regime to under-capacity regime, so as to improve the overall utility of the transportation system. There are many value pricing schemes in the literature? Vickrey (1992)?, including static pricing, dynamic pricing and system-optimal pricing. In static pricing, the price of traversing a link is set as a constant, which can be related to the long term average performance of the link; in dynamic pricing, the price of traversing a link may change according to time of day or daily pattern; in system-optimal pricing, a set of prices for all links is calculated by optimizing a system goal. However, these pricing schemes have two drawbacks:

1. The prices paid by travelers do not directly associate with the uncertainty of travelers’ experienced travel times. In static pricing, the constant price cannot change accordingly to address this concern; in typical dynamic pricing schemes, a higher price could be charged during rush hour, but this price is usually fixed and does not change if experienced travel time contains large deviations from expected levels on a day-to-day basis for each driver.

2. The traditional road pricing scheme only charges a premium without providing economic payoff/protection to travelers according to their experienced travel times. The charge is the access fee to the resource, which is irrelevant to the quality of service received. Therefore, the charge amount does not contain information about expected traffic conditions in future. Furthermore, without providing pay-off to individual travelers, the amount of payment/premium is generally small and not flexible according to changes in travelers’ experienced travel times. Therefore, the potential of traditional road pricing for changing travelers’ behaviors is limited.
Hypothetically, there could be a new type of value pricing scheme that can provide payoff/protection against adverse changes in the experienced travel time on a link for a traveler, and the traveler’s toll payment could fluctuate according to the changes of the value of such protection. In this scenario, travelers can pay for the real protection that they obtained in the trip. The payoff and value of the protection is determined by travelers’ experienced travel times: The price is the expectation of the payoff. As a result, travelers pay tolls associated with the expected quality of transportation service that they enjoy. Furthermore, given the tight linkage between current pricing and experienced travel time in future, travelers would alter their traveling behaviors according to the changes in the price. These behavioral changes would reflect people’s general concerns with future performance of the transportation system. Such a new pricing plan would potentially lead to innovative routing schemes. Considering such possibilities, travel time derivatives are introduced in this thesis.

As a financial derivative such as options can be structured based on underlying stock, so can a similar derivative in this context be based on travel time. To buy this derivative, a traveler pays a premium - the toll for that path - in exchange for protection against uncertainties of travel time in future. The payoff of such a derivative is a function of the travel time actually experienced in future. The relationship between the prices of such derivatives and the payoff to the travelers can be determined using financial asset pricing.

Financial asset pricing refers to the general methodology that yields prices of traded assets in financial markets. Usually, such assets include options, futures and other financial derivativesHull (2009). A financial derivative is a financial instrument whose value is derived from one or more underlying assets, market securities or indices, which are called the underlying assets. For example, the value of stock options is determined by the price of the corresponding stock. Asset pricing theory yields a price linked not only to the expectation but also to the volatility of future prices of underlying financial assets. To apply these concepts to the context of travel time, travel time derivatives are financial instruments whose
value is derived from travel time, which is the underlying asset. Travel time is treated as a non-tradable financial asset, and financial asset pricing yields a price of the financial derivative that is linked not only to the expectation but also to the volatility of future travel time. The higher the expectation of future travel time, or the more volatile future travel time is, the more travelers should have the option to obtain protection. Correspondingly, travelers should pay more to purchase travel time derivatives in exchange for adequate protection, as the protection itself is more valuable. Here volatility is a statistical measure of the dispersion of returns for the travel time on a link over time. It can be measured by computing the standard deviation of the one-step ratios of the historical travel time series. By measuring such dynamic structure over time, the price of travel time derivatives can be reasonably computed. The introduction of travel time derivatives is based on a sound foundation, as the market-making and pricing methods for financial derivatives, especially for derivatives based on non-tradable assets, have been extensively studied Hull & White (1990). Based on the no-arbitrage principle and risk neutral pricing methods, derivatives with reasonable payoff functions can be priced using analytical solutions, numerical schemes or Monte Carlo simulation.

With abundant technology and well-defined asset pricing methods, travel time distributions can be modeled and predicted, and innovative value pricing schemes can be designed and implemented. Travel time has been chosen as the fundamental measure of interest in this research because it is arguably the most basic and readily observable measure. While travel time is the measure of focus, the findings in this thesis are extendable to other measures.

1.2 Problem statement and thesis objective

To address the concerns above, a suitable research framework should be defined. This section introduces the framework for this research, including the transportation network setting, the definition of travel time and major issues of interest.
1.2.1 Transportation network and travel time measurement

To study the transportation system, a rigorous definition of the transportation network is given below:

A physical transportation network can be modeled using a set of nodes and links. A physical intersection is defined as a node. The section formed by the same physical road between two adjacent intersections is defined as a link. A traveler may travel through many links from origin to destination. A group of links studied together for their dependent structure is called a link set. A path is a special link set in which the links are ordered by the sequence in which they are visited by a traveler in a trip.

By the definition above, transportation systems can be modeled as graphs composed of nodes connected by links. In such graphs, traffic flowing through the network can branch from one link to another. Travel is then described as a sequence of these links following a path from node to node. Path travel time is the sum of the travel time across each link and each node traversed in the path from origin to destination.

Traditionally, links are considered the physical segments spanning intersections, and nodes are physical points of intersection. However, practical observations suggest not only that nodes have travel time associated with them but also that the travel time is not single-valued. Its value depends on which downstream link is being traversed. Travel times on paths going straight differ from those on paths turning left or turning right. This is an important practical compulsion. However, if link travel time is measured on virtual links, which start from the midpoint of a physical link to the midpoint of a neighboring downstream physical link, then the travel time through the virtual node at the midpoint of the physical link tends to be zero because there is no ambiguity in the assignment of travel time to one virtual link, as all vehicles tend to traverse the node at an automated speed. Thus, these definitions of node and link automatically reduce the measurement uncertainties of travel time.
Following this logic, monuments are defined as reference points (nodes) to and from which travel times are measured. Usually, a monument is set at the middle point of a link, and the shortest path between two monuments is defined as a monument-to-monument (m2m) link. These m2m links span intersections along a path and encapsulate the delay of the intervening intersections and turning movements. A route in a m2m network is simply the sequence of m2m segments chained to form a path from origin to destination. In the analysis of travel time distribution in this thesis (Chapters 3 and 4), all the links are monument-to-monument links. Hence in this framework, routing decisions are made at the m2m link level (i.e., travelers go from m2m link to m2m link rather than from node to node).

1.2.2 Travel time

There are different ways of defining road travel time, and the empirical travel time data show characteristics distinct to those different definitions. In this research, two definitions of travel time over a specified segment are considered: (a) the travel time of an individual traveler and (b) some assembled measures over an aggregation of individual travel time (e.g., average travel time).

For the first part of research, individual travel time is studied. By definition, the travel time experienced by an individual traveler is termed as individual travel time, and it is the difference in the arrival time between two specific points along a specific trip.

Historical observations of individual travel time are usually collected directly from the field using methods that include test vehicles with GPS devices, automatic license plate recognition, automatic vehicle identification and electronic distance measuring devices. GPS devices carried on vehicles are the data sources for individual travel time in this part of the study. Most of the data used here were recorded every three seconds along each trip. As the data were collected only when some cars with GPS devices traveled, the travel time observations in the data set can be sparse in time and difficult to analyze by usual time series
method. However, such observations of actual travel time at different times can form the empirical distribution of travel time, which is a good representation of the traffic conditions on that link. The estimation of such distribution enables travelers to make routing decisions by minimizing the risk due to uncertainties in travel time.

For the second part of this research, travel time derivatives are designed based on the average travel time of a link at time $t$. For a link, an average travel time at time $t$ is defined as the average of the experienced travel time by all travelers in the link who enter the link in time interval $t$. It is an overall description of the traffic conditions on that link at the given time. Ideally, such measurement can be obtained by taking a small time interval $\delta t$ and we calculate the travel time experienced by the vehicles that enter the link within $[t - \delta t/2, t + \delta t/2]$ assuming traffic condition is stationary within such a short time interval. Because this average travel time is an overall description of the traffic status on the link, it is often estimated using standard formulations of other measured parameters such as speed, volume and occupancy. Sources for indirect travel time include intrusive and non-intrusive detection sensors such as inductance loops, video detection, microwave radar, infrared, ultrasonic, passive acoustic array, and magnetic technologies. Loop detectors are the data sources for average travel time in this research.

These data can be measured periodically and are dense in time. The data also show the instant traffic conditions on that segment. Because average link travel time is the aggregate measure of all travelers’ behaviors, it is more difficult for an individual to manipulate average link travel time than experienced individual travel time. Hence, compared to derivatives based on individual travel time, travel time derivatives based on the average link travel time should be better tools to provide protection against travel time risk on that segment, and the price should be more stable.
1.2.3 Decision framework

Using the terminology above, the decision framework can be described as follows: A traveler traverses from one virtual link to another during a trip. At a monument, a path is chosen by comparing the estimated travel time distributions associated with all alternative paths. To make the routing choice, a risk measure is calculated for each alternative path, and the path with the optimum travel time distribution in terms of this risk measure is selected. Other settings for this decision process include the following:

1. Travel time observations are indexed by the entering times of travelers. As travelers base their routing decisions on the travel time predictions at time of entrance to a link, travel time observations are sorted by the time that the traveler enters the link. Such forward-looking data address concerns of instantaneous routing decisions.

2. Routing decisions are based on forecasts of future travel times, which, in turn, are predicted based on past observations. Historical observations lead to parameter estimation for the copula through a calibration process.

3. Only link travel time observations and the time at which these observations were made are stored. If the number of nodes is $N$ in a given network, its description requires $O(N)$ data elements. As a comparison, suppose only the best paths between any two spots in the network were stored, then the number of data records stored for the network would be of the same magnitude as $O(N^2)$. Furthermore, there are infinite numbers of arbitrary paths from any two given nodes. As a result, it is impossible to store the data for all the paths. These data constraints result from the large scale of the network and from limits in storage.

4. The decision should be made by estimating the practical travel time distribution without assuming that the travel times on different links are each subject to independent normal distributions. The empirical marginal distribution of travel time and the de-
The measurement, estimation and decision-making process focus on certain distributions of travel time. To estimate such distributions, this thesis introduces the copula model, previously used in finance, to the study of transportation. For a selected link set, a copula model can be calibrated. This model describes the dependence as calibrated using the historical travel time observations on all related links together. Then, the model predicts the downstream travel time distributions based on the copula, using the experienced upstream travel time as an input. In this way, the model takes into consideration the instantaneous changes of the traffic conditions along the path.

1.2.4 Major issues of interest

This decision framework poses substantial challenges to the research from different aspects. To focus the topic and reduce those challenges, this thesis emphasizes the uncertainties of travel time and the dependence aspect of the problem given a limited number of alternative paths. This assumption is suitable for a practical road network on which the number of alternative paths between current position and destination is limited. Three main topics of interest are introduced as follows:

1. Prediction: How is the unknown travel time distribution for a downstream link predicted during a trip?

In a trip, a traveler needs to make minimum-risk routing decisions based on current conditions of the transportation network. To estimate the risk to the unfinished part of trip, the traveler should be able to predict the distribution of travel time for the downstream network, given the experienced travel time on the current link and all other travel information up to the decisioning time. To enable this calculation, a model should be built to generate the travel time distribution for each downstream
link, given necessary historical observations and the experienced travel time up to the decision node in the trip. The model should capture the dependence between the travel time of traversed links and those of downstream links. Moreover, when data are not available to study such dependence at the current decision node, a calibrated model from a similar node should be used.

This topic is illustrated in Figure 1.1

![Figure 1.1: Predict the conditional distribution of travel time on downstream links](image)

The simple case is to estimate travel time distribution of a downstream link, given a state of the two links. To address this issue, copula methods are used to estimate the link travel time distributions conditioned on current observations. The dependent structure between links is modeled by only one copula function, and the marginal distribution of travel time for a given link is modeled by kernel methods. Conditional density is calculated using the estimated copula, and risk measures are calculated using the conditional density.

2. Assembly: How are the link travel time data assembled to compute an estimate of the path travel time distribution?

An important issue related to the paths is the challenge of creating route travel times by aggregating the segment travel times through a generally summing process. It
would be ideal if the sample size for each path (origin-destination pair) was sufficient to develop a travel time distribution for every possible departure time under every possible condition, but this expectation is unreasonable. Hence, it is necessary to assemble segment travel times to create path-level travel time distributions. Such aggregation is not trivial because correlation exists among drivers and segments. Many of the same vehicle drivers who create the travel times on one segment are also involved in creating travel times on others, so the travel times for segments nearby are clearly related, the empirical evidence for which is shown in Figure 1.2. Clearly, the true path travel time distribution has much heavier tails than the distribution estimated, which assumes links are independent.

Figure 1.2: The empirical path travel time distribution is different from the estimation based on independence assumption (red: estimated distribution under the assumption of independence; green: empirical distribution)

To conduct the aggregation, the joint distribution of travel time on the links that constitute a given path is estimated, and then the path travel time distribution is generated as the distribution of the assemblage of dependent random variables. This topic is illustrated below and in Figure 1.3.

Fundamentally, the model assumes a path state is uniquely specified by a series of factors, including time of day, weather, events on the path, and so forth. The path state varies as the system factors change. Once specified, the path state determines
the vector of link travel time states on the virtual links and the dependence between them. Within this state, the copula structure is stationary; thus a Gaussian copula can be used to model this dependent structure. A path state specifies a joint distribution for the travel time of all links that constitute the path; hence, it represents the specific traffic conditions along the path.

For example, in Figure 1.3, there are three links in the path. For a given sequence of link travel time scenarios, a specific Gaussian copula is used to model the dependent structure between the constituting links. This path state can be pre-defined according to traffic context. Alternatively, it can be determined by using maximum likelihood algorithms introduced in this thesis.

The overall path travel time distribution is approximated by a weighted sum of a finite number of path travel time distributions corresponding to different path states. Copula models are used to model the dependence within each path scenario. The weights are the frequency number of the traffic scenarios.

The thesis further generalizes the mathematic form of the model introduced above. A Gaussian copula mixture model (GCMM) is defined as a weighted sum of finite joint distributions, each of which contains a Gaussian copula. The model above belongs to a specific type of GCMM, in which each Gaussian copula corresponds to a predefined scenario, and observations are classified to these scenarios before they are used for
For more general GCMMs, each Gaussian copula may represent a statistical scenario that can be updated during certain recursive learning schemes; the observations can belong to different scenarios with different probabilities. This GCMM can be fitted to the empirical travel time data that describes the dependent structure between links. To conduct this fitting process for a general GCMM, a two-step maximum likelihood estimator is proposed, and corresponding algorithms are designed.

3. Pricing: Can travel time, with its uncertainties and volatility, be treated as an underlying asset similar to that on which financial derivatives are based? How should the market be made, and how should such travel time derivatives be priced?

Value pricing has been an important research topic, as pricing will change travelers’ behaviors and improve the performance of traffic systems. Traditional value pricing schemes mainly include static value pricing and dynamic value pricing. To effectively provide insurance against the risk of travel time measured in standard deviation or volatility and to improve travelers’ behaviors, the derivatives based on experienced travel time are proposed in this thesis as a more precise stochastic value pricing scheme.

Travel time derivatives are financial derivatives based on experienced travel time. Travelers purchase derivative contracts before their trips start and obtain a payoff when they finish their trips. The payoffs are directly linked to the performance of the path travel time that they experienced, and the price of the contracts is determined not only by the level of congestion but also by the anticipated volatility of future congestion before they started the trip.

Travel time derivatives provide travelers effective insurance against uncertainty in their daily travel time. They also help travelers to plan their trips, as the trading price of such derivatives reflect the overall market view on the performance of the link in future. Introducing travel time derivatives is also motivated by the need to hedge traffic-related
risk for other participants/industries, such as logistic companies and various delivery services, whose welfare is related to the transportation system.

To clarify further, a typical travel time derivative contract is demonstrated below1.4. The payoff of the derivatives is a function of the experienced travel time in the future, and the price of the contract is the conditional expectation of the payoff. Therefore, the price/road toll corresponding to each link shows the general market view of the future traffic conditions on that link. If the contract is a put option, the payoff is larger when the future travel time is smaller, so that a higher price implies people generally expect better traffic conditions on that link. The traveler may select routes based on the prices of all alternative links. The payoff to the traveler receives at end of the trip compensates the risk due to unexpected traffic conditions. To enable such a mechanism, market making, product designing and pricing are concerns that should be addressed, and they are all discussed in detail in this thesis.

More specifically, mean reverting processes driven by diffusion and integrated diffusion processes are used, and derivative prices are derived based on the no-arbitrage pricing principle and the risk neutral pricing principle under incomplete market conditions.
In all, the interaction and similarity between traffic systems and financial systems are highlighted in this thesis. The copula methods and stochastic models from the financial math literature are used to model travel time. Meanwhile, travel time derivatives are introduced as financial assets. Implementation of these concepts can effectively improve the utilization of the traffic network and help travelers to hedge transportation-related risk.

1.3 Thesis outline

The rest of the thesis is organized as follows. Chapter 2 provides background and a literature review for travel time estimation methods and related mathematic theory. Chapter 3 describes the link travel time estimation through copula models. Chapter 4 describes the estimation for path travel time distribution. Gaussian copula mixture models (GCMM) are used to model the dependence between the traffic states of links. Both a state-based GCMM and general GCMMs are discussed. Chapter 5 introduces the travel time derivatives and their pricing method. Chapter 6 concludes the thesis and describes potential future work.
Chapter 2

Literature Review

This chapter reviews the research on travel time estimation, dependent structure and derivative pricing method. In this thesis, dependent structure models, more specifically, copulas, are used to model travel time for generating routing guidance, and asset pricing models are used to price travel time derivatives. The literature on these topics provides the theoretical background of the thesis.

2.1 Travel time estimation

As noted in Chapter 1, road travel time estimation has been an important topic in transportation research. Researchers have examined methods for accurate prediction of travel time on roads. In previous research, authors have introduced various models to describe the dependent structure of the travel time processes; however, general dependent structure models (copulas) have not been used in this area.
2.1.1 Link travel time estimation

Because links are the fundamental segments of road networks, the measurement characteristics and forecast of the travel time on links are essential steps for this research. In this section, different methods of calculating link travel time estimation are reviewed, with a focus on the dependence model between different links.

1. Time series methods: In Ho & Lee (2004) and VEMURI et al. (1998), the problem of short-term forecasting of traffic delay is formulated as a time-series evolution problem. Delays are predicted under a regression framework. Zhang & Rice (2003) proposes a method to predict freeway travel times using a linear model in which the coefficients vary as smooth functions of the departure time. A cross-validation procedure of mean percentage prediction error is also given. In Wu et al. (2004), the support vector regression method is used to model daily changes in travel time. It significantly reduces the root-mean-squared errors of predicted travel times.

2. Kalman filtering: Chu et al. (2005) uses the Kalman filter to analyze travel time changes. The system model is described with a state equation and an observation equation based on the traditional traffic flow theory. Cathey & Dailey (2003) identifies a three-component model with a tracker, a filter, and a predictor, in which the Kalman filter is used as the filter. This model uses automatic vehicle location (AVL) data to position a vehicle in space and time and then predict the arrival/departure at a selected location.

4. Flow theory model: The models presented in Van Grol et al. (1998) and Petty et al. (1998) are based on the macroscopic hydrodynamic traffic flow theory. Nie & Zhang (2005) proposes the cell transmission model (CTM), which assumes that all cars travel at exactly the free flow speed from the gate cell to the sink cell. The dependence between different links is implicitly modeled in the inflow and the outflow. In Carey & McCartney (2002) and Carey et al. (2003), link travel time is approximated by a whole-link model, in which for a vehicle entering a link at time $t$, the link travel time is expressed as a weighted average of the inflow rate at the time the vehicle enters and the outflow rate at the time it exits.

5. Non-parametrical method: In Pattanamekar et al. (2003), Gaussian kernels are used to estimate the continuous mean travel time at a particular point in time $t$. A local three-point polynomial approximation is used to estimate the mean link travel time as a function of time of day, and a two-factor model, in which stochastic travel time is caused by a systematic error and a vehicle error, is used for error decomposition.

Although different models have been applied in previous literature to estimate link travel times, the underlying model of the dependence between links shares some common assumptions. First, most research assumes that the travel times between different links are independent, and the focus is on predicting the future change of travel time on a single link over time. Second, it is also assumed that the distributions of travel time are Gaussian, but the general dependent structures between travel times between different links or between different entering times have not been explicitly studied. For research that has studied the dependent structure, the usual assumption has been that travel time of different links is subject to a joint normal distribution, such as Dailey et al. (2000). This assumption makes the result tractable, but a single joint normal distribution cannot fully describe the real dependence between the travel times of different links in a realistic transportation network. A new model that can effectively describe more flexible dependent structure should be developed.
Furthermore, travelers tend to care more about the extreme values of travel time. They need to know the occurrence of congestion in downstream links, given the traffic status up to decision time. In other words, the dependence on the tail of travel time distributions should be studied to describe travel time risk. The new model should be able to describe such dependence in the tails of travel time distribution.

Besides the difference in mathematics, the estimation method for link travel time also varies according to data source. In general, link travel time can be directly measured by devices or calculated through other measurements, and the estimation should be conducted in different ways.

1. Site-based measurement:

Site-based measurement mainly refers to data from loop detectors. The measurements come from fixed sites, and they usually include speed, occupancy, etc. Average travel time (ATT) on the link can be calculated using these measurements. Usually, the data points are obtained periodically; hence, the data are suitable for regular time series analysis.

2. Vehicle-based measurement:

Floating car data and GPS-based data are typical vehicle-based measurements. Usually, the device onboard sends back the location of the vehicle and the corresponding time. Travel time can be calculated by taking the difference of such recorded times. The observations obtained in this way are recorded at random time points, and, hence, they are not suitable for time series analysis unless aggregated into discrete time periods. In this thesis, copula models are selected to analyze these data, and this choice is justified in detail in Chapter 3.
2.1.2 Path travel time estimation

After the link travel time distribution is characterized, the further challenge is how to assemble link travel time distribution to determine a characterization of the travel time distribution for a path. This characterization should be created for each feasible path, and a routing choice is then generated by choosing the best path from all feasible paths according to certain risk measures. To generate such path travel time distribution, spatial dependence and temporal dependence should be considered. Here, spatial dependence means the change of travel time on a current link may depend on the travel time on upstream links. Temporal dependence means the change of travel time on current links may depend on overall traffic system conditions, which change mainly according to time of day. This section reviews related literature as follows:

Fu & Rilett (1998) consider the shortest path problem in dynamic and stochastic networks. Taylor expansions are used to generate the conditional mean and variance for path travel time. The assumption is that travel times on individual links at a particular point in time are statistically independent. It is then argued the probability distributions of the link travel times are modeled as functions of the time of day: the dependence between links in a path only exists through the entering time. This approach ignores the second category of dependent structure between links in a path: If the travel time in the upstream link is high, then the travel time is likely to be high in the downstream link as well.

In Waller & Ziliaskopoulos (2002) the assumption is that there is one-step spatial dependence between successive links and limited temporal dependence on each link. The authors then derive algorithms to generate the shortest paths based on known probabilistic transition matrices that describe the transition from a specific state of the upstream link to one specific state of the downstream link. In contrast, how to estimate such probability distribution of travel time by considering the underlying dependent structure between upstream and downstream links is the focus of this thesis.
In Pattanemkar et al. (2003), it is mentioned that to estimate the conditional mean and variance of one link given the observation of the other, the joint probability density function is needed. However, the author suggests it is not practical to estimate the this function and uses a more straight forward method- a three-point polynomial approximation to estimate mean travel time instead of studying the joint probability density function in detail.

Rakha et al. (2006) uses coefficient of variation, defined below, to estimate the variance of path travel time.

$$CV = \frac{\mu}{\sigma}$$

The author shows several methods for the estimation of path travel-time variance from its component segment travel-time variances:

$$\sigma_p^2 = \frac{(\sum \mu_j)^2}{m^2} \sum_{l \in L(p)} \frac{\mu_j}{\sigma_j}$$

$$\sigma_p^2 = (\sum \mu_j)^2 med_{l \in L(p)} \frac{\mu_j}{\sigma_j}$$

$$\sigma_p^2 = (\sum \mu_j)^2 \frac{1}{2} (\max_{l \in L(p)} \frac{\mu_j}{\sigma_j} + \min_{l \in L(p)} \frac{\mu_j}{\sigma_j})$$

The three estimators above are derived by considering the trip CV as the mean/median/mean of the CV over all segment realizations in the path. The underlying assumption is that the travel times of different links are independent from each other.

Sherali et al. (2006) derives another formulation for estimating the path travel-time variance. The maximum and minimum segment travel-time CVs are used to construct bounds on the path CV given that the CV is independent of the length of the segment. Then the path variance is estimated by:

$$\sigma_p^2 = \frac{(\sum \mu_j)^2}{\sum \mu_j^2} \sum_{l \in L(p)} \sigma^2$$

It is a modification of Rakha’s estimation.
The joint normal assumption is the usual assumption of most studies with respect to correlated link travel time estimation prediction of Dailey et al. (2000). Gao & Chabini (2006) considers the simplified correlation model over time but not between different links. In all, the research on the real time dependence of link travel time is far from complete, and the dependence between non-normal distributed link travel time should be addressed as the next improvement.

2.2 Dependent structure and copula theory

Because the dependence of non-normally distributed travel time is studied in this thesis, it is necessary to introduce some theoretical tools to model such dependence. For this purpose, the theory of dependent structure and copula functions are introduced below:

2.2.1 Definitions

Dependent structure is the dependent relationship between random variables. Mathematically, it is expressed in copula functions. The approach to formulate a multivariate distribution using a copula is based on the idea that a simple transformation can be conducted for each marginal variable so that each transformed marginal variable has a uniform distribution. (The transformation is to plug the random variable into its own cumulative probability function.) After this transformation, the dependence structure can be expressed as a multivariate distribution on the obtained uniforms, and a copula is precisely a multivariate distribution of these marginally uniform random variables. A more rigorous definition of copula is given below based on definition of the 2-increasing functions, grounded functions, and sub-copulas, which are defined in Nelsen (2006) and U Cherubini (2004).
Definition 2.2.1 2-increasing functions Given a two-dimensional function

\[ V_H(B) = H(x_2, y_2) - H(x_1, y_2) - H(x_2, y_1) + H(x_1, y_1) \]

defined on the product space \( B = [x_1, y_1] \times [x_2, y_2] \), a 2-place real function \( H \) is 2-increasing if \( V_H(B) \geq 0 \) for all rectangles \( B \) whose vertices lie in \( \text{Dom } H \).

Definition 2.2.2 Grounded functions

Let \( A_1 \) and \( A_2 \) be two subsets of the full set \( I \), denote \( a_1 \) and \( a_2 \) the least elements of \( A_1 \) and \( A_2 \) respectively. The function \( H \) is grounded if for every \( H(v, z) \) of \( A_1 \times A_2 \) where, \( H(a_1, z) = 0 = H(v, a_2) \).

Definition 2.2.3 Sub-copula A two-dimensional sub-copula is a function \( C' \) with the following properties:

1. \( \text{Dom } C' = S_1 \times S_2, \) where \( S_1 \) and \( S_2 \) are subsets of \( I = [0, 1] \), containing 0 and 1;
2. \( C' \) is grounded and 2-increasing
3. For every \( u \) in \( S_1 \) and every \( v \) in \( S_2 \), \( C'(u, 1) = u \) and \( C'(1, v) = v \).

Definition 2.2.4 A two-dimensional copula A two-dimensional copula is a two-dimensional sub-copula with \( S_1 = S_2 = I \)

The definitions above can be extended to a multivariate case. A multivariate copula is a multivariate distribution function defined on the unit cube \([0, 1]^n\), with uniformly distributed marginal distributions.
2.2.2 Fundamental theorem of copula

Based on the definitions of copula function, Sklar (1959) shows that a \( N \)-dimensional joint distribution function may be decomposed into its \( N \) separate marginal distributions and a copula function, which completely describes the dependent structure between the \( N \) random variables. The fundamental mathematical theorem is as follows:

**Theorem 2.2.1** Sklar’s theorem of copula Let \( H \) be a joint distribution function with marginal cumulative probability functions \( F_X \) and \( F_Y \). Then there exists a copula \( C \) such that for all \( x,y \in \bar{\mathbb{R}} \),

\[
H(x, y) = C(F_X(x), F_Y(y))
\]

if \( F_X \) and \( F_Y \) are continuous, then \( C \) is unique; otherwise, \( C \) is uniquely determined on \( \text{Ran} F_X \times \text{Ran} F_Y \). Conversely, if \( C \) is a copula and \( F_X \) and \( F_Y \) are distribution functions, then the function \( H \) defined above is a joint distribution function with margins \( F_X \) and \( F_Y \).

A direct application of this theorem yields the following conclusion:

**Theorem 2.2.2** \( N \) random variables are subject to a joint Gaussian distribution, if and only if the \( N \) marginal distributions are Gaussian and the dependent structure between them is a Gaussian copula function.

The following theorem defines a conditional copula:

**Theorem 2.2.3** Sklar’s theorem of conditional copula: Sklar’s theorem for continuous conditional distributions Given two random variables \( X \) and \( Y \), let \( H \) be the conditional bivariate distribution function with continuous marginal cumulative distribution function \( F_X \) and \( F_Y \), and let \( F \) be some conditioning set or \( \sigma \)-algebra, then there exists a unique conditional copula \( C \) which is defined on the product space \([0, 1] \times [0, 1] \), such that
\[ H(x, y|F) = C(F_X(x|F), F_Y(y|F)|F), \forall x, y \in R \]

Conversely, if \( C \) is a conditional copula and \( F_X \) and \( F_Y \) are the conditional distribution functions of the two random variables \( X \) and \( Y \), then the function \( H \) defined by the above equation is a bivariate conditional distribution function with margins \( F_X \) and \( F_Y \).

### 2.2.3 Tail dependence of the copula

One important concept in copula theory is the tail dependence, which provides a measure for extreme co-movements in the lower and upper tail of \( F_{X,Y}(x, y) \). Fischer & Klein (n.d.)

**Definition 2.2.5** The upper tail dependence coefficient \( \lambda_U \):

\[
\lambda_U = \lim_{u \to 1} -P(Y > F_Y^{-1}(u)|X > F_X^{-1}(u)) = \lim_{u \to 1} \frac{1 - 2u + C(u, u)}{1 - u} \in [0, 1]
\]

**Definition 2.2.6** The lower tail dependent coefficient \( \lambda_L \):

\[
\lambda_L = \lim_{u \to 0^+} P(Y \leq F_Y^{-1}(u)|X \leq F_X^{-1}(u)) = \lim_{u \to 0^+} \frac{C(u, u)}{u} \in [0, 1]
\]

More background about the copula theory is given below:

1. The density of the conditional probability is calculated as:

\[
f(x, y) = \frac{\partial C}{\partial u \partial v} \frac{\partial F_X(x)}{\partial x} \frac{\partial F_Y(y)}{\partial y}
\]

2. The usual dependency measures, such as Pearson’s correlation \( \rho \), Spearman’s \( \rho \), and Kendall’s \( \tau \), can be calculated based on a specified copula function. In this way different
correlation measures are integrated into a common framework under the copula theory Nelsen (1995) and Fredricks & Nelsen (2007).

3. To estimate the copula model, the usual method is a two-step maximum likelihood estimator in which the marginal distribution is estimated first and copula is estimated given the estimated marginal distribution. The basic theory about these estimators can be found in White (1994), including the construction of the estimators, the estimation algorithms and the corresponding asymptotic properties. Copulas are later used in Patton (n.d.) to construct the dynamic covariance model.

There are five major advantages of the copula-based approach:

1. with copula models, more flexible dependent structures can be studied than with joint Gaussian distribution. This increased scope means the model is capable of describing flexible marginal distribution and dependence structures.

2. a copula model can be used to estimate the dependence between extremely high travel times. This use shows the probability of simultaneous occurrence of extreme travel time on link pairs. This detail may potentially yield precise predictions of congestion and make a great impact on routing decisions.

3. copula models can model the marginal distributions (i.e., the univariate travel time distributions) and the dependent structure between links (the copula) separately. The former can be estimated with a larger volume of data because there are usually relatively more observations on a single link. To estimate copula functions, however, a pair of observations on both links that satisfies defined conditions should be found, and usually the number of such pairs is smaller than the number of observations on each link in the link set.

4. a copula model can be estimated through maximum likelihood method based on time-sparse data. If the correspondences between link observations are defined, a vector of
such observations can be collected over time, and parameters can be estimated based on such data set.

5. the parameters of copula represent the dependent structure between a set of links. The copula type and parameters can be used for similar link sets, in which observations are too limited to estimate dependent structure.

Compared to the characteristics of GPS data, the copula methods is suitable for conducting analysis of sparse data sets generated by GPS devices. The copulas used in this thesis are defined in Table 2.1:

<table>
<thead>
<tr>
<th>Name</th>
<th>CDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bivariate Normal</td>
<td>$C_{\mu, \rho}(u, v) = \int_{-\infty}^{v} \Phi^{-1}(u) \int_{-\infty}^{v} \Phi^{-1}(v) \frac{1}{2\pi(1-\rho^2)^{1/2}} e^{\frac{s^2-2\rho st+t^2}{2(1-\rho^2)}} dsdt$</td>
</tr>
<tr>
<td>T Copula</td>
<td>$C_{\nu, \rho}(u, v) = \int_{-\infty}^{v} \Gamma^{-1}(u) \int_{-\infty}^{v} \Gamma^{-1}(v) \frac{1}{2\pi(1-\rho^2)^{1/2}} \left{ 1 + \frac{s^2-2\rho st+t^2}{\nu(1-\rho^2)} \right}^{-1/(\nu+2)/2} dsdt$</td>
</tr>
<tr>
<td>Clayton</td>
<td>$C(\kappa, \gamma)(u, v) = 1 - [1 - (1 - u)^{\kappa}]^{-\gamma} + [1 - (1 - v)^{\kappa}]^{-\gamma} - 1^{-1/\gamma^{1/\kappa}}$</td>
</tr>
<tr>
<td>Gumbel</td>
<td>$C_{\theta}(u, v) = \exp\left(-[(-\log u_1)^{\theta} + (-\log u_2)^{\theta}]^{1/\theta}\right)$</td>
</tr>
<tr>
<td>Frank</td>
<td>$C_{\theta}(u, v) = -\frac{1}{\theta} \log \left(1 + \frac{(e^{\theta u_1}-1)(e^{\theta u_2}-1)}{e^{\theta}-1}\right)$</td>
</tr>
<tr>
<td>BB1</td>
<td>$C_{\theta, \delta}(x, y) = \left(1 + \left[(u^{\theta} - 1)^{\delta} + (v^{\theta} - 1)^{\delta}\right]^{1/\delta}\right)^{-1/\theta}$</td>
</tr>
</tbody>
</table>

To compare different copulas, their theoretical tail dependence properties are listed in Table 3.2.
Finally, when working with empirical data, appropriate copula functions are selected by comparing the likelihood values when fitting the data to different copulas. The theoretical tail dependence of different copulas is also compared with the empirical tail statistics. The empirical tail statistics (Lower tail $L(z)$ and Upper tail $R(z)$) are defined as follows:

**Definition 2.2.7 Empirical tail dependence of two-dimensional observations**

\[ L(z) = \frac{N(z, z)}{z} \]

\[ R(z) = \frac{(1 - 2z + N(z, z))}{(1 - z)} \]

where $N(z, z)$ is the number of pairs of data in which the observations on two links are both smaller than the real number $z$.

As Gaussian distribution and Gaussian copula are used in this thesis, other covariance matrix estimation methods can be applied beyond the usual maximum likelihood method. The most straightforward method is to estimate the pairwise correlations for the focused random variables after some suitable transformations; by recording these correlations, the parameters for a Gaussian copula can be obtained. This method, however, cannot in general guarantee the correlation matrix is positive semi-definite; without being positive semi-definite, such a parameter matrix leads to defaults in calculation. As a solution, the lasso method is employed to estimate the covariance matrix in this thesis: The lasso algorithm is first proposed in Tibshirani (1996). It minimizes the sum of square subject to the sum of the absolute value of the coefficient being less than a constant.

\[
(\alpha, \bar{\beta}) = \arg\min_{(\alpha, \beta)} \sum_{i=1}^{n} \left(y_i - \alpha \sum_j \beta_j x_{ij}\right)^2
\]

s.t. $\sum |\beta_j| \leq t$ The Lasso solution is obtained by conducting an iterative procedure that starts from an overall least square estimate and solves a constrained least square problem.
in each step. In Friedman et al. (2008), lasso is used to estimate the inverse of an unknown covariance matrix. The penalized constraint removes unnecessary terms in the covariance matrix while keeping it positive semi-definite. The problem of inverse covariance matrix estimates is then set as

$$\arg\min_{\Omega: \Omega \succeq 0} \log\det\Omega + \text{tr}(S\Omega) + \rho|\Omega|_1$$

where $S$ is the sample variance matrix and $\rho$ is the weight. Its dual problem is

$$\min_{\beta} \frac{1}{2} W^{1/2} 1 1^T \beta - b_2 + \rho|\beta|_1$$

where $\beta$ can be solved by lasso.

### 2.2.4 Nonparametric density estimation and conditional density estimator

In this section, some non-parametric estimation methods for joint distribution are reviewed. After applying these methods, the conditional distribution of random variables can be estimated accordingly.

The density at $x$ by the kernel smoothing method for univariate data is given as follows with properties discussed in Bowman et al. (1998):

$$\hat{f} = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

The multivariate kernel estimator is given as:
\[ f(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K_{\frac{1}{h}}(x - X_i) \]

The normal kernel is defined as:

\[ K(x) = (2\pi)^{-d/2} \exp(-\frac{1}{2} x^T x) \]

Based on these definitions, the parametric methods and nonparametric methods to fit a joint distribution can be introduced:

1. Estimate the joint distribution as an integrated structure.
   
   (a) Conduct parametric estimation for the joint distribution: A parametric family of joint distribution can be selected and fitted to the empirical data. There are two drawbacks to this method. First, the empirical distribution can hardly be fully described by a parametric joint distribution family; second, it is costly to estimate parameters if the dimension is high.

   (b) Conduct non-parametric kernel estimation for joint distribution. High dimensional kernel estimation can be conducted to approximate the joint distribution. This method is denoted as (Method 1). Hansen (2004) suggests both a one-step estimator and a two-step estimator. The corresponding properties are discussed.

2. Separate the estimation of marginal distributions from that of the dependent structure.
   
   The dependent structure is estimated using copula methods, and marginal distributions can be estimated using parametric/non-parametric methods. Considering the complex nature of the marginal travel time, nonparametric estimation for the marginal distribution is selected, as it is more flexible to fit the real distributions. Then there are two alternatives, as described below:
(a) Estimate marginal distributions through non-parametric ways and dependent structure through parametric ways. After the non-parametric estimation of the marginal distribution, the multivariate dependent structure can be estimated through certain parametric copula families. If one single parametric copula does not fit the data well, a Gaussian copula mixture model is introduced, which is the weighted sum of several Gaussian copulas. This flexible parametric model is a powerful estimator of complex dependent structures. This method is denoted as (Method 2).

(b) Estimate marginal distributions through non-parametric ways and dependent structure through non-parametric ways. High dimensional kernels can be applied to estimate the copula function. The challenge, however, is that because a copula and its density are defined on a compact cube $[0, 1]^d$, continuous kernel estimators suffer from boundary noise, Chen & Huang (2007) and Chen & Huang (2007). This method is denoted as (Method 3).

3. Estimate conditional density directly through nonparametric methods. This method is to estimate the derivatives of copula function by kernel methods. This method is denoted as (Method 4), and the details are given below:

Fan et al. (1996) proposes a double-kernel local linear regression approach. The authors treat the conditional density function as a regression model given as follows:

$$E(K_{h_2}(Y - y)|x = x) = g(y|x)$$
The estimator is then calculated by a Taylor expansion:

\[
E(K_{h2}(Y - y)|x = z) = g(y|x) = g(x) + g'(x)^T(z - x) + \frac{1}{2}g''(x)(z - x)'(z - x)
\]

Then the parameters can be estimated by minimizing the following objective, which is a typical weighted least square regression:

\[
\sum_{i}^{n}(K_{h2}(Y_i - y) - \beta_0 + \beta_1T(X_i - x) + \beta_2vec((X_i - x)'(X_i - x)))^2W_{h2}(X_i - x)
\]

In this thesis, Method 2 is selected as the major modeling methodology, and the distribution of travel time on each link is estimated through kernel methods.

2.3 Derivative pricing

In this thesis, new financial derivatives based on travel time are introduced to hedge potential risk resulting from the uncertainties of travel time. To introduce a legitimate framework and some appropriate pricing methods that enable market making and trading for travel time derivatives, it is necessary to review major topics in financial asset pricing theory. For this purpose, this section reviews fundamental concepts and methods in the literature of financial asset pricing.
2.3.1 Derivative fundamentals

To understand financial asset pricing, it is necessary to clarify the definitions of financial assets and related concepts about the financial markets. As an introduction, the following concepts define the financial terms on which financial asset pricing theory is based: A financial asset is an intangible asset that derives value because of a contractual claim. A financial instrument is then a tradable financial asset of any kind, including cash; evidence of an ownership interest in an entity; or a contractual right to receive or deliver cash or another financial instrument. A financial derivative is a financial instrument whose price is derived from the value of something else (known as the underlying asset). Any stochastic changing element that will generate changes of cash flow and relate to economic life can serve as underlying asset here. Therefore, the underlying asset on which a derivative is based can be the price of an asset (e.g., commodities, equities [stock], residential mortgages, commercial real estate, loans, bonds), the value of an index (e.g., interest rates, exchange rates, stock market indices, consumer price index [CPI]), or other items. The main types of derivatives are forwards, futures, options, and swaps, John (2000).

The underlying assets of derivatives can be classified as tradable and non-tradable. Stocks, commodities, and other existing trading instruments are tradable as they are instruments that people can trade in the market. Traditionally, equity options, commodity futures, and interest rate swaps are financial derivatives, the values of which are based on corresponding tradable assets. Black-Scholes theory introduces classical methods to price the financial derivatives based on tradable assets. Compared to the derivatives above, weather derivatives are one of the major financial derivatives of which the values are based on non-tradable assets. Typically, temperature changes at a location are not traded on the financial market, but financial derivatives can be derived based on the temperatures. These derivatives may provide protections to farmers around that location against adverse future weather conditions that may reduce their crop production Banks (2002). Below, the review will focus
on financial derivatives based on non-tradable assets, as travel time is not a tradable asset on the financial market.

The weather derivative was first introduced to the market in late 1990s. In 1999, the Chicago Mercantile Exchange introduced weather futures contracts, the payoffs of which are based on average temperatures at specified locations. According to Stewart (2002), in a wide variety of industries from property management to natural gas retailing, firms face the possibility of significant earnings declines or advances because of unpredictable weather patterns. That fluctuation is a strong incentive to initiate the weather derivative market. According to Banks (2002), the business model in the weather derivative market is formed in response to the need for hedging: Industries subject to weather risk participate first and take roles on either the buy side or the sell side. Speculators then come in to provide additional liquidity to the market. The trading and capital activities enable a beneficial mechanism for all parties involved.

To price such derivatives, there are several major pricing models, including actuarial, no-arbitrage (market pricing), indifferent, risk neutral/replication, and so forth. Some of them are reviewed below and used to price travel time derivatives in this thesis.

1. **Actuarial pricing** The actuarial pricing principle states that the price of a financial instrument is the discounted sum of its expected future return and the additional price for risk determined by the contract properties based on the current price and position Banks (2002). The use of this model should be based on the statistics of the contract payout and their relation to the current holding position. The results are subject to how the expected return and risk are modeled.

2. **Risk neutral pricing** Originally, the risk neutral pricing principle assumes the underlying asset is tradable, and dynamic hedging can be used to replicate the payoff of the derivative using the underlying assets Steele (2001). After scaling the return of every instrument with its risk, every instrument can be compared and priced. The
Black-Scholes pricing model is one of the major theories following this logic, and Black Scholes method can be extended to price financial derivatives based on non-tradable assetsLuenberger (2004).

3. **No-arbitrage pricing** The no-arbitrage pricing principle states that ”If two investments yield the same payoff in all scenarios, they must yield the same market price, otherwise you will get arbitrage by buying the cheap and sell it high”? It has been used to price interest rate derivatives Sundaresan (2009). This methodology can be used for travel time derivativesDelbaen & Schachermayer (1997) and Hull & White (1990).

4. **Indifferent pricing** The indifferent pricing principle states that the utility an investor receives by investing the underlying investment alone should be unchanged after introducing a new derivative based on it. The utility indifference buy (or bid) price $p$ is the price at which the investor is indifferent (in the sense that the expected utility under optimal trading is unchanged) between paying nothing and not having the derivative and paying $p$ now to receive the payoff of the derivative at time $T$. This approach is also an effective way to price weather derivatives in Carmona & Diko (2005) and Xu et al. (2008).

Of interest is the similar stochastic nature between temperature changes at a given location and the travel time along a given path. Hence, travel time derivatives will be developed in a similar fashion to that used in the creation and pricing of weather derivatives. The characteristics of the travel time changes are described with suitable stochastic model in later chapters.
2.3.2 Probability settings of derivative pricing

After financial derivatives are introduced, suitable methods should be developed to price them. In the context of travel time derivatives, the price of using the link should be calculated based on expected traffic conditions on the link. To conduct such calculation, necessary probabilistic settings are introduced below: Let \((\Omega, F, \{F_t\}_{t>0}, P)\) be a complete filtration probability space. A random variable is a mapping \(X: \omega \rightarrow R\) if it is \(F\)-measurable, whereas a family of random variables depending on time \(t\), \(\{F_t\}_{t>0}\) is defined as a stochastic process. A process \(X_t\) is \(F\)-adapted if every \(X_t\) is measurable with respect to the \(\sigma\)-algebra \(F_t\). If the paths \(t \rightarrow X(t, \omega)\) are right continuous with left limits everywhere with probability one, then the stochastic process is called cadlag.

A stopping time \(\tau\) is a random variable with values in \([0, \infty]\) and with the property \(\omega \in \Omega \mid \tau(\omega) \leq t \in F_t\), for every \(t > 0\). An adapted cadlag stochastic process \(M_t\) is called a martingale if it is in \(L^1(P)\) for all \(t > 0\) and for every \(t \geq s \geq 0\) : \(E[M_t \mid F_s] = M_s\), \(M_t\) is a local martingale if there exists a sequence of stopping times \(\tau_n < \infty\) where \(\tau_n < \infty\), such that \(M_{t \wedge \tau_n}\) is a martingale. Consider a finite time horizon \(t \in [0, T]\) and let \(Q\) be a probability measure equivalent to \(P\).

Based on the probability settings, some typical stochastic processes are introduced below: A dynamic of the spot price evolution is desirable, first because modeling uncertainty in spot prices is of interest to traders and second because spot prices are used as reference points for settlement of forward and future contracts. In mathematical finance, the traditional models are based on Brownian Motion \(B_t\), also called the Wiener process. The most common model for the price dynamics \(S_t\) of a financial asset is the exponential of a drift Brownian motion, known as the geometric Brownian motion,

\[
S_t = S_0 \exp(mt + sB_t)
\]
Mean reverting processes are another special class of stochastic processes. In this thesis, mean reverting processes are used to model the changes of travel time when pricing travel time derivatives. The general mathematical form is as follows:

\[ dX_t = \mu_t + a_t(\mu_t - X_t)dt + dy_t \]

The descriptions of each part are as follows:

1. \( \mu_t \) is the mean value of the process. The realization of the stochastic process tends to revert to this mean value as the fluctuation part is negative when the process is above its mean value and positive when the process is below its mean value.

2. \( a_t(\mu_t - X_t)dt \) is the fluctuation part of the process. It is negative when the process is above its mean value and positive when the process is below its mean value. This part describes the fluctuation phenomenon by which the travel time may be less or greater than the usual mean value at a given time instant but always fluctuates around its statistical mean value.

3. \( dy_t \) is the stochastic driven process, and different \( y_t \) processes lead to the different \( \varepsilon_t \). In classical time series models, the noise term is subject to independent identical normal distribution, the corresponding continuous driven process to which is a standard Brownian motion. A more complex driving process can be introduced that leads to different properties for \( X_t \).

The processes above are continuous, and under such models, there is no jump in value for the asset prices. To model the possible discontinuous changes in prices, more advanced models such as the Levy process are introduced.

Consider a finite time horizon \( t \in [0, T] \) and let \( Q \) be a probability measure equivalent to
Let \( Z_t \) be the density process of the Random Nikodym derivative so that:

\[
Z_t = \frac{dQ}{dP} |_{F_t}
\]

The Levy-Kintchine decomposition of the process gives the connection to semi-martingales:

\[
L_t = g_t + M_t + \int_0^t \int_{|z|<1} \tilde{N}_i(ds, dz) + \int_0^t \int_{|z|>1} N_i(ds, dz)
\]

where \( M_t \) is a local square integrable continuous martingale with quadratic variation equal to \( Ct \). \( N \) denotes the random jump measure, \( \tilde{N} = N - l \) stands for the compensated random jump measure. An adapted cadlag stochastic process \( S_t \) is a semi-martingale if it has the Levy-Kintchine representation

\[
S_t = S_0 + A_t + M_t + \int_0^t \int_{\mathbb{R}\setminus\{0\}} X_{1(t,z)} \tilde{N}(ds, dz) + \int_0^t \int_{\mathbb{R}\setminus\{0\}} X_{2(t,z)} N(ds, dz)
\]

where \( A_t \) is an adapted continuous stochastic process having paths of finite variation on finite time intervals, \( M_t \) is a continuous square integrable local martingale, \( S_0 \) is an \( F_0 \)-measurable random variable, \( X_{1(t,z,w)}, X_{2(t,z,w)} \) are predictable random variables defined on \([0, \infty) \times \mathbb{R} \times \mathcal{W}\) with \( X_{1(t,z,q)} X_{2(t,z,w)}(t, z, q) = 0 \), satisfying \( \int_0^t \int_{\mathbb{R}\setminus\{0\}} |X_{1(s,z)}| l(ds, dz) < \infty \) and \( \int_0^t \int_{\mathbb{R}\setminus\{0\}} |X_{2(s,z)}| N(ds, dz) < \infty \), a.s.

As introduced in Todorov (2007), Levy measure specifies the properties of the Levy processes. Every other Levy process is a combination of these two parts: the continuous part of every Levy process is the Brownian motion, which has unbounded variation and quadratic variation proportional to time; the pure jump part of every Levy process is of finite activity when \( \mu(R_d^0) < \infty \), and it is of infinite activity when \( \mu(R_d^0) = \infty \). Further, the set of infinitely active pure jump processes can be subdivided into those with finite variation or infinite variation. For a pure jump process to be of finite variation, it is necessary and sufficient that \( \int_{|y|<1} |y| \mu(dy) < \infty \).
This chapter reviews fundamental theories in transportation research and financial engineering. It is shown that the need of modeling travel time distribution and related value pricing schemes lead to demand of innovative modeling methods. Copulas and asset pricing theory which have been developed and applied first in the financial have been identified as suitable candidates for such modeling effort. As a first demonstration, copulas are used for modeling the dependent structure between travel times of different links and estimating travel time distributions in next chapter.
Chapter 3

Link Travel Time Estimation and Routing Decisioning through Copula Methods

Based on the literature review in Chapter 2, this chapter studies the travel time distribution across links using copula models. The profile description of travel time data is discussed first in Section 3.1. The definitions of Copula models are introduced and a two step maximum likelihood estimator is used to estimate the model in Section 3.2: First, parametric and non-parametric estimators are compared to model link travel time distribution. Second, appropriate conditional copulas are used to model the dependent structure between travel times of different links. After conditional link travel time distributions are estimated, transportation risk measures are introduced to generate different routing decisions in Section 3.3, which set up the basic routing decisioning framework for this thesis. Numerical examples are used to demonstrate the whole procedure in Section 3.4.

Two further applications are demonstrated based on the framework: first, reliable prediction of travel time distribution is generated by the aggregation of the predictions from
different adjacent link sets; second, a method is developed by which similarity principles are used to compute copulas for intersections without sufficient data from copulas for intersections with sufficient data. These demonstrations show the potential of the developed methods in practical transportation applications.

### 3.1 Profile description of link travel time data

To model link travel time properly, the major characteristics of link travel time should be addressed, which determine model type, model structure, and parameterization and calibration methods. For these purposes, the following characteristics of link travel time data are considered:

1. Data are not lined up physically. As vehicles show up randomly in a given link, the travel time observations from onboard GPS devices are collected at random time points and can be sparse in time. Hence, it is difficult to find consecutive travel time observations with fixed time intervals in between, a data point which is required for standard time series analysis.

2. The data set is composed of the speed based on trips of individual vehicles. For a given path, the available data records (each record a vector of travel time observations for the constituent links) are equal to the number of travelers who traversed the whole path. On the other hand, one link in this given path can be part of many other paths, so the number of available data records for a given link is equal to or more than that for any path that goes through the link. To take advantage of such difference, it is preferable to use models that can estimate the marginal link travel time distribution and the dependent structure between different links separately. This idea is illustrated in Figure 3.1.

3. The data are unevenly distributed in space: There are relatively abundant observations...
There are more data records on link AB than any path that goes through it (ABC or ABD) on some links but fewer on others. In the paths/intersections concerned, there can be insufficient observations for a proper characterization of the underlying structure. Therefore, travelers can make decisions based only on some information from a similar path or intersection. To take advantage of such similarity between different paths or interactions, a model that can explicitly describe such similarity needs to be developed.

The challenges above led to this research to find new models to study travel time, and copula methods have been chosen as they best fit the requirements. The details are discussed in the following sections.

3.2 **Copula based travel time distribution generation**

Copula is used to model the dependence between the travel times of related links.

3.2.1 **Copula and two step maximum likelihood estimation**

To estimate the link travel time distribution, copula models are selected to model the dependence. The corresponding definition is given for the bivariate case first:

Let $P$ be a conditional bivariate distribution function with continuous margins $F_X$ and
and let $F$ be some conditioning set. Then there exists a unique conditional copula $C : [0, 1] \times [0, 1]$ such that

$$P(x, y|F) = C(F_X(x|F), F_Y(y|F)|F), \forall x, y \in R$$ (3.1)

and the conditional distribution of $Y$ given $X$ can be calculated as follows:

$$P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)} = \frac{\partial C}{\partial u \partial v}(C(F_Y(y), F_X(x))) \frac{\partial F_Y(y)}{\partial y}(y)$$ (3.2)

By estimating the copula, the conditional distribution of travel time on the downstream link can be calculated as follows:

$$P(Y \leq y_0|X = x) = \int_0^{y_0} \frac{\partial C}{\partial u \partial v}(C(F_Y(y), F_X(x))) \frac{\partial F_Y(y)}{\partial y}(y)dy$$ (3.3)

To use the copula model, maximum likelihood estimators (MLEs) are used to estimate the model in two steps. Here $L_{XY}(\phi, \gamma, \kappa)$ is the log likelihood function of the joint distribution for $X$ and $Y$. $\phi$ is the parameter for the marginal distribution of $X$, $\gamma$ is the parameter for the marginal distribution of $Y$, $\kappa$ is the parameter for the marginal distribution of copula.

1. The total likelihood is expressed as $L_{XY}(\phi, \gamma, \kappa) = L_X(\phi) + L_{Y|X}(\phi, \gamma, \kappa)$
2. Maximize $L_X(\phi)$ to get an estimation $\hat{\phi}$ of $\phi$.
3. Maximize $L_{Y|X}(\phi_0, \gamma, \kappa)$ to get an estimation conditioned on $\hat{\phi}$.

Applying the two step MLE method to $L_{Y|X}$, the second step is equivalent to

1. The total likelihood is expressed as $L_{Y|X}(\phi_0, \gamma, \kappa) = L_Y(\gamma) + L_c(\phi_0, \gamma, \kappa)$
2. Maximize $L_Y(\gamma)$ first to get $\hat{\gamma}$. 

53
3. Maximize $L_c \phi_0, \gamma_0, \kappa$ conditioned on $\hat{\phi}$ and $\hat{\gamma}$.

The copula parameters are estimated in this two-step procedure. The first step is the identification of the marginal distribution, and the second step is the estimation of the conditional copula functions. The estimation is consistent and asymptotic efficient according to White (1994).

identification of marginal distribution using parametric and non-parametric estimators

In this section, both parametric and nonparametric estimators are used to identify the marginal distribution of link travel time. The parametric estimators compared include normal, lognormal, Gamma, Weibull and generalized Pareto distributions.

Non-parametric estimators include a kernel smoothing estimator with a Gaussian kernel. As reviewed in Chapter 2, the density at $x$ by the kernel smoothing method is given by:

$$\hat{f} = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$ \hspace{1cm} (3.4)

where $K$ is the kernel function, $h$ is the window size, and $X_i$ is the observed data. The Gaussian kernel is selected and the bandwidth chosen according to the optimality rules in Bowman & Azzalini (1997), which are implemented in Matlab.

Furthermore, different distance measures are calculated to show the quality of fit. They are defined in Table 3.1, where $F_{emp}$ is the cumulative distribution function of the empirical distribution, and $F_{est}$ is the cumulative distribution function of the estimated distribution.
Table 3.1: Definition of distance measures

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov distance</td>
<td>$KS = \max_{x \in \mathbb{R}}</td>
</tr>
<tr>
<td>L1 statistics</td>
<td>$L1 = \sum_{x \in \mathbb{X}}</td>
</tr>
<tr>
<td>L2 statistics</td>
<td>$L2 = \sum_{x \in \mathbb{X}} (F_{emp}(x) - F_{est}(x))^2$</td>
</tr>
<tr>
<td>Anderson-Darling statistic:</td>
<td>$AD = \max_{x \in \mathbb{R}} \frac{</td>
</tr>
</tbody>
</table>

Estimation of the conditional copula

After estimating the marginal distribution of travel time, the copula function between the travel times of different links is identified.

As travelers are interested in the co-occurrence of congestion on the focused links, the probability of experiencing high travel time in the unobserved links, given high travel time in the observed links, needs to be modeled. In copula theory, this is modeled by the upper tail dependence, as described in Definitions 2.3.5 and 2.3.6. Below, alternative copulas are proposed and compared according to their upper tail dependence.

The alternative copulas used in this chapter are bivariate Gaussian, T, Clayton, Gumbel, Frank, and BB1 copula. They are defined in Table 2.1 in Chapter 2. Their tail dependence are given below 3.2. According to theoretical results, there is no upper or lower dependence

Table 3.2: Tail dependence for typical copulas

<table>
<thead>
<tr>
<th>Name</th>
<th>Lower tail</th>
<th>Upper Tail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian Copula</td>
<td>$\lambda_L = 0$</td>
<td>$\lambda_U = 0$</td>
</tr>
<tr>
<td>T Copula</td>
<td>$\lambda_L = 2t_{\mu+1}(-\sqrt{\mu + 1} \sqrt{\frac{1-\epsilon}{1+\epsilon}})$</td>
<td>$\lambda_U = \lambda_L$</td>
</tr>
<tr>
<td>Clayton</td>
<td>$\lambda_L = 2^{-\frac{1}{\delta}}$</td>
<td>$\lambda_U = 0$</td>
</tr>
<tr>
<td>Gumbel</td>
<td>$\lambda_L = 0$</td>
<td>$\lambda_U = 2 - 2^{\frac{1}{\delta}}$</td>
</tr>
<tr>
<td>Frank</td>
<td>$\lambda_L = 0$</td>
<td>$\lambda_U = 0$</td>
</tr>
<tr>
<td>BB1</td>
<td>$\lambda_L = 2^{-\frac{1}{\delta}}$</td>
<td>$\lambda_U = 2 - 2^{\frac{1}{\delta}}$</td>
</tr>
</tbody>
</table>

in the Gaussian copula or Frank copula. There is only lower tail dependence for Clayton copula and only upper tail dependence for Gumbel copula. For T copulas both tails are the same and nonzero, while the two-tail dependence parameters of BB1 can be different.
Copulas with both upper and lower dependence are preferable to others if only one copula is used to describe the dependence for a link set. Based on this rationale, the empirical tail dependence described in Definition 2.3.7 is compared to the theoretical tail to select the best copula.

After the marginal distributions and the copulas are estimated, the travel time distribution of the dependent links can be generated.

### 3.3 Routing decision based on the estimated travel time distribution

After estimating the distribution, different risk measures are used for decision making. In finance, a risk measure is used to determine the amount of an asset or set of assets (traditionally currency) to be reserved for potential losses. In transportation studies, the risk measures should be defined in the transportation context as follows:

**Definition 3.3.1** Transportation risk measure

Transportation risk measure $\rho$ is the calculation to determine the amount of time to be reserved for potential transportation delay. It is a mapping from a set of random variables to real numbers, and it satisfies the following properties:

1. $\forall a \in R, \rho(a + X) = \rho(X) + f(a)$: when a real number $a$ is added to a stochastic travel time, the risk introduced by this change is a positive increasing function of $a$ as well. So if more travel time is needed to travel through a link, then there is more transportation risk associated with that link.

2. if $X_1 > X_2$ then $\rho(X_1) > \rho(X_2)$, if travel time on a link is constantly greater (that is, longer) than travel time on another link, then its risk measure is greater as well.
The following are some typical risk measures, most of which share the properties described above.

1. Rule Type 1: Mean-variance rule

**Definition 3.3.2** Rule 1 Mean-variance rule: The random variable $X$ and $Y$ characterized by the pair $(\mu_X, \sigma_X)$ dominate the other random variables characterized by the pair $(\mu_Y, \sigma_Y)$ if and only if $\mu_X < \mu_Y$ and $\sigma_X < \sigma_Y$. That is

$$X \succeq_{MV} Y \iff \mu_X < \mu_Y \text{ and } \sigma_X < \sigma_Y$$

If a candidate link dominates other candidate links in a mean-variance sense, then that link is preferred by the traveler when making his routing choices.

**Definition 3.3.3** Rule 1’ Transferred mean variance: The random variables $X$ and $Y$ characterized by the pair $(\mu_X, \sigma_X)$ dominate the other random variable characterized by the pair $(\mu_Y, \sigma_Y)$ if and only if $\mu_X + r\sigma_X^2 < \mu_Y + r\sigma_Y^2$ for some $r$. That is

$$X \succeq_{TMV} Y \iff \mu_X + r\sigma_X^2 < \mu_Y + r\sigma_Y^2$$

If a candidate link dominates other candidate links in a transferred mean-variance sense, then that link is preferred by the traveler when making routing choices.

2. Rule Type 2: Stochastic domination rules

**Definition 3.3.4** Rule 2 First-Order Stochastic Domination (FSD): The random variable $X$ first-order stochastically dominates the random variable $Y$, if $P(X > a) \geq P(Y > a), \forall a$. That is

$$X \succeq_{SSD} Y \iff P(X > a) \geq P(Y > a), \forall a$$
If a candidate link is dominated by other candidate links in a FSD sense, then that link is preferred by the traveler when making his routing choices.

**Definition 3.3.5** Rule 2’ Second-Order Stochastic Domination (SSD): Suppose the random variables \(X\) and \(Y\) have support on \([l; u]\). Then \(X\) second-order stochastically dominates \(Y\) if

\[
\int_l^a P(X > t)dt \geq \int_l^a P(Y > t)dt, \forall a \in [l, u]
\]

That is

\[
X \succeq_{\text{SSD}} Y \iff \int_l^a P(X > t)dt \geq \int_l^a P(Y > t)dt \forall a \in [l, u]
\]

If a candidate link is dominated by the others in SSD, then that link is preferred by the traveler when making his routing choices.

**Theorem 3.3.1** \(X\) second-order stochastically dominates \(Y\) if and only if

\[
E[h(X)] \geq E[h(Y)]
\]

for any increasing and concave function \(h\).

**Theorem 3.3.2** If \(X\) second-order stochastically dominates \(Y\), and if \(X\) and \(Y\) have the same mean, then \(X\) has a smaller variance than \(Y\)

3. Rule Type 3: Area ratio rule A new decision rule related to the integration of cumulative distribution function (CDF) is proposed in Rachel R. He (2005);

**Definition 3.3.6** Rule 3 Area ratio rule: If the area ratio of these CDF curves of two random variables \(X\) and \(Y\) before some critical point \(t\) is equal to or greater than a
threshold $\epsilon$, $X$ is dominating $Y$; otherwise, $Y$ is dominating $X$. That is

$$X \succeq Y \iff \frac{\int_0^t F_A(x)dx}{\int_0^t F_B(x)dx} \geq \epsilon$$

If a candidate link is dominating the others in the area ratio rule sense, then that link is preferred by the traveler.

As this rule was first constructed based on comparison of CDF curves, no direct interpretation in the transportation context was given to it. This thesis introduces its implications in transportation by modeling the probability characteristics of different link travel time later.

4. Rule Type 4: Expected exponential utility rule

**Definition 3.3.7** Rule 4: Expected exponential utility: Take the exponential utility function $U(x) = \exp(-ax)$ where $a$ stands for the risk aversion for the traveler. $1/a$ is the value of travel time that the traveler will give a utility value of $e^{-1}$. Then the random variable which yields a larger expected utility value dominates the other. That is

$$X \succeq Y \iff EU(E\exp(-aX)) \leq E\exp(-aY)$$

If a candidate link is dominating the others in the area ratio rule sense, then that link is preferred.

5. Rule Type 5: Value-at-risk rule

**Definition 3.3.8** Rule 5: Value-at-risk: Consider the top $\alpha$ percent quantile of the distributions. The distribution with a smaller $\alpha$ quantile value dominates the other, as there is less probability for extreme high values. That is

$$X \succeq Y \iff VAR Q^\alpha_X \leq Q^\alpha_Y$$
If a candidate link is dominating the others in a value-at-risk sense, then that link is preferred by the traveler.

Among the measures above, the first four are calculated based on the entire travel time distribution, and Rule 5 is calculated based on tails of distribution. After the conditional density of link travel time is estimated, these measures can be calculated to show the travel time risk associated with the estimated travel time distribution so that the travelers can generate routing decisions. Below, the area ratio rule is studied further, and its interpretation in transportation context is discussed.

**Theorem 3.3.3** Suppose the cars enter a path according to a Poisson process $\lambda$, and the distribution of travel time is $F$, then according to the properties of Poisson process: $P(M_t = k) \sim \text{Poisson}(\lambda')$, where $M_t$ is the number of cars traverse the path completely by time $t$. Here, the $\lambda' = \lambda \times \frac{1}{t} \int_{0}^{t} F(s)ds$

**Proof:**

Denote the number of probes which enters the path by time $t$ as $N_t$. It is assumed that the arrival event of a probe and its travel time on the link are independent.

First, we calculate the distribution of travel time conditioned on the number of probes that has entered the route at time $t$, denoted as $N_t$, then conditioned on the entering time $T_i$ is

$$P(T_{ps} <= s | T_i = t, N_i = n)$$

$$= P(T_{ps} <= s)$$

$$= F(s)$$
\[ P(T_{ps} + T_i \leq t | N_t = n) \]
\[ = \int_0^t P(T_{ps} \leq t - T_i | T_i = s, N_t = n) \cdot P(T_i = s | N_t = n) \, ds \]
\[ = \int_0^t P(T_{ps} \leq t - s | U_i = s) \cdot P(U_i = s) \, ds \]
\[ = \int_0^t P(T_{ps} \leq t - s) \cdot \frac{1}{t} \, ds \]
\[ = \int_0^t F(t - s) \cdot \frac{1}{t} \, ds \]
\[ = \int_0^t F(s) \cdot \frac{1}{t} \, ds \]

Please note the fact is used in the deduction above, that the distribution of the arrival time of the \( n \) probes is the same as \( n \) order statistics given the given \( N_t = n \), if the arrival of probes is subject to a Poisson process. In other words the arrival times of probes given the \( n \) probes completes the trip by time \( t \) are subject to a conditional uniform distribution between \([0, t]\).

Then since the behaviors of probes are independent from each other, the occurrence of event that a probe traverses the path completely before time \( t \) is subject to a Bernoulli distribution with parameter

\[ P_t = P(T_{ps} + T_i \leq t | N_t = n) \]

. Then the event that \( k \) out of the \( n \) probes traverse the path completely within time \( t \) is subject to a binomial distribution:

\[ P(M_t = k | N_t = n) \]
\[ = C_n^k P_t^k \cdot (1 - P_t)^{n-k} \]
Then

\[
P(M_t = k) = \sum_{n=k}^{\infty} P(M_t = k|N_t = n) * P(N_t = n)
\]

\[
= \sum_{n=k}^{\infty} C_n^k P_t^k * (1 - P_t)^{n-k} * \frac{(\lambda t)^n * e^{-\lambda t}}{n!}
\]

\[
= \frac{(\lambda t * P_t)^k * e^{-\lambda t * P_t}}{k!}
\]

Then the number of probes that traverse the path completely within time \( t \) is a Poisson process, and the parameter is

\[
\lambda * P_t = \lambda * \frac{1}{t} \int_0^t F(s)ds
\]

The expected number of probes that traverse the path completely within time \( t \), denoted as \( E(X_t) \), is then

\[
\lambda * P_t * t
\]

Q.E.D.
According to the theorem, the area-ratio between two paths can be calculated as:

\[
\rho = \frac{\int_{t_0}^{t_l} F_p(s) \, dt}{\int_{t_0}^{t_l} F_q(s) \, dt} = \frac{\int_{t_0}^{t_l} F_p(s) \ast \frac{1}{t_l} \, dt}{\int_{t_0}^{t_l} F_q(s) \ast \frac{1}{t_l} \, dt} = \frac{P_p(T_{ps} + T_i \leq t | N_t = n)}{P_q(T_{ps} + T_i \leq t | N_t = n)} = \frac{\lambda \ast P_{p-t_l} \ast t_l}{\lambda \ast P_{q-t_l} \ast t_l} = \frac{E(X_{p-t_l})}{E(X_{q-t_l})}
\]

Hence, the area ratio means the ratio of the expected number of cars that complete the route by time \(t_l\) when travel times on the paths are subject to the two different travel time distributions respectively. It measures the different capacity of the two paths to let traffic pass in the same time period. Hence, the area ratio rule means that if the expected road capacity of a candidate link is larger than the capacity of others, then that candidate link will be preferred.

### 3.4 Numerical analysis

In this section, a numerical example is given to illustrate the basic procedure and decision framework. First the data and experiment design are discussed as follows:

1. The testing network: The data in the experiment come from the GPS-travel time in the Princeton-Trenton highway system and are collected from the probe cars with installed Copilot systems (the GPS routing guidance and data collection system produced by [http://www.alk.com](http://www.alk.com) (2009)). The test area is shown in 3.2, and the targeting monu-
ment links are shown in Table 3.3 (the time unit used in this thesis is seconds). Please note monuments are defined as reference points (nodes) to and from which the travel times are measured. Usually, a monument is set at the middle point of a link, and the shortest path between two monuments is defined as a monument-to-monument (m2m) link.

Figure 3.2: Test network N95-US 195, the red frames are the targeted monuments

<table>
<thead>
<tr>
<th>Link name</th>
<th>Starting Monument</th>
<th>Ending Monument</th>
<th>Number of observations</th>
<th>The Latest Simultaneous Observation (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>101078</td>
<td>101094</td>
<td>909</td>
<td>251</td>
</tr>
<tr>
<td>BC</td>
<td>101094</td>
<td>101095</td>
<td>893</td>
<td>372</td>
</tr>
<tr>
<td>CD</td>
<td>101095</td>
<td>102825</td>
<td>980</td>
<td>392</td>
</tr>
<tr>
<td>BE</td>
<td>101094</td>
<td>101089</td>
<td>131</td>
<td>317</td>
</tr>
<tr>
<td>AF</td>
<td>101078</td>
<td>132460</td>
<td>24</td>
<td>-</td>
</tr>
<tr>
<td>GH</td>
<td>101069</td>
<td>101062</td>
<td>915</td>
<td>-</td>
</tr>
<tr>
<td>GK</td>
<td>101062</td>
<td>101061</td>
<td>76</td>
<td>-</td>
</tr>
</tbody>
</table>

2. Data sources and pre-processing
While there are many ways of collecting data, the processes described in Chapters 3 and 4 mainly use the GPS data points that are relayed during a trip by an onboard GPS device installed in each probe car. Each data record contains a sequence of positions and the positions’ corresponding time points collected along a given trip. Based on this raw data, link travel time observations in the research data set are generated by the following procedures:

(a) Map-matching the GPS points to appropriate links.
(b) Along a given recorded trip, find the monuments which the traveler traversed through. For each monument, record the two closest GPS locations to the monument: One is upstream of the monument, and the other is downstream of the monument. Denote the time that the traveler arrived at the upstream location as \((T_l, L_l)\), and denote that for the downstream location as \((T_n, L_n)\).
(c) Use the following formula to conduct a linear interpolation of the two times by distance to get the time that the vehicle arrives at the monument point.

\[
T_m = |L_m - L_l|T_l + |L_c - L_n|T_n
\]

(d) Generate the monument-monument link travel time by taking the difference of the \(T_m\)s for the two nearby monuments along the trip.

One more step is added to generate the path travel time: For a given trip data record, enumerate all the possible sub-trips in it. Record the travel time for each sub-trip by taking the difference of the \(T_m\) which corresponds to the starting monument and ending monuments of the sub-trip. Note that in the research, each trip starts and ends at a monument.

For practical application with the overall data set in the U.S., the data processing procedures are slightly different. The so-called "One-mon" data set is generated first by the following procedures:
(a) Map-matching the GPS points to appropriate links.

(b) For each physical link, for each trip that includes this link, record the speed $V_h$ and the position $L_h$ near the characteristic points of the link—the point $T_h$ closest to the geometrical center.

(c) Estimate the time in the center point of the link by the following formula

$$T_c = T_h + |L_c - L_h|/V_h$$

where $T_c$ means the time when the vehicle is at the center of the link. $L_c$ means exact halfway of the link, calculated from the link attributes.

(d) Store all $T_c$ for all the links where a monument is located. Therefore

$$T_m = T_c$$

for the specified links where monuments are located.

(e) Calculate the difference of $T_c$ between two given monuments to get the travel time for m2m links.

The above procedures build the data set that is used to validate the theory.

3. Experiment design.

The targeted link is a monument link AB, and the conditional travel time density on AB is estimated by considering the link sets AB-BE (on U.S. 195) and AB-BC-CD (on Interstate 95).

The estimated distribution for AB conditioned on the latest simultaneous observations on AF is compared with the latest observation on AB for validation. A leave-one-out cross validation is conducted as a validation. Routing decisions are generated by comparing the estimated travel time distribution for $AB$ with the smoothed empirical
travel time distribution on AF. GH and GK are used to illustrate how to conduct estimations based on the similarity between intersections.

4. Data and outliers.

In Pattanamekar et al. (2003), two methods to remove outliers are given. : 1) to reject a data record if the speed is greater than twice the speed limit or 100 mph, whichever is lower; 2) to reject a data record if the journey time falls outside 4 standard deviations from the mean. In this thesis, as travel time is calculated by taking the difference of the arrival time to two monuments, a large value of travel time can result from several reasons, including (a) congestion between the two monuments, (b) measurement errors, and (c) traveler’s non-traffic related stops between monuments. Usually, extremely large deviation is caused by Reasons B or C in this list. Thus, extreme values beyond \( \mu + 4\sigma \) are excluded from the analysis.

The detailed procedures and the results are as follows:

1. Fit marginal travel time distribution on each link.

For link AB, the fitted cumulative distribution function and probability density function are shown in Figure 3.3. To select a best parametrical distribution, the distance measure between the empirical distribution \( F_{emp} \) and the estimated distribution \( F_{est} \) are shown in Table 5.5

<table>
<thead>
<tr>
<th>Name</th>
<th>KS</th>
<th>L1</th>
<th>L2</th>
<th>AD</th>
<th>Pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lognormal</td>
<td>0.0055</td>
<td>134.4894</td>
<td>27.6103</td>
<td>549.4505</td>
<td>0</td>
</tr>
<tr>
<td>Normal (Gaussian)</td>
<td>0.0055</td>
<td>178.5836</td>
<td>48.2145</td>
<td>659.3407</td>
<td>7.001e-47</td>
</tr>
<tr>
<td>Gamma</td>
<td>0.0055</td>
<td>152.6858</td>
<td>35.8179</td>
<td>659.3407</td>
<td>( \approx 0 )</td>
</tr>
<tr>
<td>Weibull</td>
<td>0.0055</td>
<td>169.7458</td>
<td>43.1215</td>
<td>329.6703</td>
<td>( \approx 0 )</td>
</tr>
<tr>
<td>Reciprocal gamma</td>
<td>0.0055</td>
<td>120.9237</td>
<td>22.3283</td>
<td>254.7083</td>
<td>6.267e-39</td>
</tr>
<tr>
<td>Generalized Pareto</td>
<td>0.0049</td>
<td>290.0954</td>
<td>132.4392</td>
<td>109.8901</td>
<td>( \approx 0 )</td>
</tr>
<tr>
<td>Kernel Smoothing</td>
<td>0.0006</td>
<td>11.9310</td>
<td>0.2217</td>
<td>0.0653</td>
<td>0.7350</td>
</tr>
</tbody>
</table>

A Kolmogorov-Smirnov test is conducted for the estimated results at a 5
2. Select the appropriate copula function:

There are quite a few copula functions with different properties. To improve quality of fit, suitable copulas are selected by comparing the empirical observations to simulated values. The corresponding results are shown in Figure 3.4, and the corresponding AIC criteria values and the parametric bootstrap test for Archimedean copulas are shown in Table reftab:gfc.

<table>
<thead>
<tr>
<th>link</th>
<th>Normal</th>
<th>T</th>
<th>BB1</th>
<th>Clayton</th>
<th>Gumbel</th>
<th>Frank</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>-80.1444</td>
<td>-84.5242</td>
<td>-85.9260</td>
<td>-71.5360</td>
<td>-76.0585</td>
<td>-72.3326</td>
</tr>
<tr>
<td>Parametric bootstrap p-value for Extreme value copula</td>
<td>-</td>
<td>-</td>
<td>0.8110</td>
<td>0.1308</td>
<td>0.0058</td>
<td>0.0350</td>
</tr>
</tbody>
</table>

Combining the theoretical properties and the fitting statistics, the following conclusions can be drawn:

(a) A Gaussian copula can hardly capture the dependent structure between the travel times of different links. However, given that various estimators can be used to
Figure 3.4: Empirical distribution v.s. simulated data based on different copulas
estimate the parameter matrix for a Gaussian copula, and it is relatively easy to simulate a Gaussian copula, it is efficient in high dimensional cases. A mixture of Gaussian copulas may be both flexible in describing the dependence structure and efficient in estimation.

(b) A T copula fits the data better than the Gaussian copula. It is easier to calculate and simulate in high dimensional cases than the BB1 copula family. However, for T copula, both upper-tail and lower-tail dependence are the same. Nevertheless, the dependence between high travel time observations of two adjacent links is different from the dependence between the low travel time observations of them. Therefore, T copula may lack a certain flexibility when used to describe the travel time dependence.

(c) Frank copula does not reveal the upper or lower dependence relationships, and the fitting result also shows difference from empirical data.

(d) Clayton copula only reveals the lower tail dependence or the simultaneous occurrence of low travel time on both links. As travelers may be more concerned with the dependence between high travel times on different links, this copula is not particularly useful for modeling travel time.

(e) Gumbel copula only shows upper tail dependence or the occurrence of high travel time on different links. This copula can be a reasonably simple model, compared to the BB1 family.

(f) BB1 family can show both upper and lower tail dependence in the joint distribution of travel time. Therefore, it can describe the simultaneous occurrence of low and high travel time on both links. However, it cannot be efficiently estimated in high dimensional cases.

Next, the tail characteristics of travel time data are compared. Following the definitions of empirical lower/upper tail indices, given in Definition 2.2.4, the two tail indices for
this link set are calculated. Results are shown in Figure 3.5:

Figure 3.5: The comparison of tails of the joint structure: Red lower tail, Blue upper tail

The figure shows that the two tails of the copula between link travel times are asymmetric in nature and the upper tail is heavier than the lower tail. It is better to use a copula with asymmetric tail structures to describe this difference. Therefore, the BB1 copula is the best choice of the all candidates. In this chapter, BB1 copula is selected for the analysis of two dimensional cases. For higher dimensional cases, a Gaussian copula mixture model is used to approximate the empirical dependent structure.

3. Copula parameter estimation

With a two-step estimator, the parameters of the bb1 copula between links AB and BE are estimated. The parameters are $\theta = 0.7634$ and $\delta = 1.5154$. The corresponding tail dependence parameters are $\Lambda_U = 0.4201$ and $\Lambda_L = 0.5493$. The results show that there is a reasonable tail dependence in the travel time data, and it is generally more significant than that for the joint normal distribution (zero-tail dependence). The dependence structure is also asymmetric with respect to different levels of travel time. In other words, the probability of the simultaneous occurrence of low travel times on
both links is approximately 0.4201, while that for simultaneous high values is greater at 0.5493.

4. Routing decisions based on the estimated copula

Routing decisions are generated based on the estimated travel time distribution of alternative links.

Suppose a traveler starts at A and needs to go to G. The traveler will need to choose one path between links AB and AF. Then suppose the travel time on the link BG and FG is equal and deterministic. In this context, a comparison of travel time distribution between AB and AF is sufficient to yield a routing decision.

Following the assumptions above, there is no new travel time observation on the link AB, and the distribution of travel time on AB conditioned on the most recent observation on link BE should be estimated. This estimation should take into account the dependence between AB and BE. Following this methodology, the estimated copula between AB and BE are presented, and the estimated travel time distribution of AB and the empirical travel time distribution on AF are displayed in Figure 3.6.

The estimated mean of AB, conditioned on the most recent travel time observed on BE as 317, is \( \mu_{AB} = 258.9 \). The estimated standard deviation is \( \sigma_{AB} = 48.07 \). The observed data when the observation on BE is 317 is 251. The estimated mean is accurate compared to the observation. A further cross validation is conducted to show the difference between the estimated mean and the observation.

The conditional distribution is then used to generate routing decisions by calculating the following decision rules, as in Table 3.6.

According to the decision rules, the following decisions are made: AB is preferred to AF under the transferred mean variance rule, when \( r \) takes the value in \((0.0018, \infty)\); under the SSD rule as it is dominated in SSD the expected exponential utility with parameter

72
Table 3.6: Decision statistics for different rules

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$(\mu, \sigma)$</th>
<th>$\mu + r\sigma^2$</th>
<th>FSD</th>
<th>SSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>(258.9, 48.07)</td>
<td>305.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>AF</td>
<td>(245.4, 98.12)</td>
<td>438.1</td>
<td>259.7</td>
<td>$\infty$</td>
</tr>
<tr>
<td>Preference</td>
<td>None</td>
<td>AB</td>
<td>None</td>
<td>AB</td>
</tr>
<tr>
<td>Exponential utility</td>
<td>a=-1/250</td>
<td>AR $t_U=1500$</td>
<td>VAR(Quantile) 5%</td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>0.3850</td>
<td>1239</td>
<td>286.6</td>
<td></td>
</tr>
<tr>
<td>AF</td>
<td>0.3808</td>
<td>1255</td>
<td>321.6</td>
<td></td>
</tr>
<tr>
<td>Preference</td>
<td>AB</td>
<td>AF</td>
<td>AB</td>
<td></td>
</tr>
</tbody>
</table>

5. Cross validation for the mean of estimated travel time. A leave-one-out cross validation is conducted.

Total deviation measures between the estimation and the actual distributions are reported in Table 3.7. According to these measures, the BB1 copula yields a better $-1/250$; and under the Value-at-risk rule with 5% upper tail. AF is preferred to AB under the area-ratio rule (Passing capacity) if the threshold ratio is set as 1. Neither AB nor AF dominates the other under the mean-variance rule or the first order stochastic dominance rule. So travelers will choose different routes according to different rules, and this flexibility is a promising improvement for current routing guidance service.
estimate for the conditional mean.

| Copula | $\sum \frac{E(Y|X=x_i) - y_i}{\hat{f}_{x_i}}$ | $\sum |\frac{E(Y|X=x_i) - y_i}{\hat{f}_{x_i}}|$ |
|--------|---------------------------------|---------------------------------|
| BB1    | 0.0283                          | 0.1372                          |
| Normal | -0.0688                         | 0.1747                          |

6. Validate the conditional distribution using data set from loop detectors.

A validation of the distribution is conducted on a denser data set that measures the average speed for a given link in Los Angeles, California Chacon & Kornhauser (2009). In this experiment:

(a) Given a current speed (67 miles/hour) and current time on this link, the conditional distribution of speed after a fixed time interval (Lag=5/50/500 minutes) is estimated by the method in this section.

(b) To validate the estimator, the empirical conditional distribution is used for comparison. This empirical version is constructed by collecting the speed observations that appear with the same time lag, whenever the given velocity value (67 mile/hour) is observed again. The comparison is shown as in Figure 3.7. The calculated density yields a reasonably good estimate, as shown in Table 3.8.

<table>
<thead>
<tr>
<th>Lag</th>
<th>5 minutes</th>
<th>50 minutes</th>
<th>500 minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2</td>
<td>1.5060</td>
<td>0.7153</td>
<td>0.3836</td>
</tr>
</tbody>
</table>
Figure 3.7: Estimated conditional density (Red) v.s. Empirical conditional density (Blue: histogram at 1.2 miles/hour, Black: kernel smoothing) given current observation 67 mile/hour per hour at a loop detector

3.5 Further development: reliable estimates and similarity-based analysis

An additional challenge to the proposed method is how to obtain a reliable estimation given limitations in the data. This section discusses how model complexity and estimation
reliability can be controlled and how calibrated models can be used for intersections with similar physical conditions. They both help to address the problem of data insufficiency in real transportation networks.

3.5.1 Generation of reliable estimation

To obtain a reliable estimation given the limitations in the data, a weighted scheme is proposed in this section. Several estimations of the conditional distribution on AB are first generated based on different link sets, of which link AB is a part. The final estimation is a weighted sum of these several estimations. Mathematically, this weighted scheme is formulated as follows:

\[
P(t_b < x) = \sum_i f_{bi}(x)a_i \tag{3.5}
\]

where \( f_{bi}(x) \) is the conditional distribution of the travel time on the blind link \( b \) given the observation in \( i \) th link set.

To illustrate this weighing scheme, a numerical experiment is conducted. Following the settings of previous sections, the travel time of link AB is estimated. To generate the estimation, both the link set AB-BE and another link set AB-BC-CD are considered. In the link set AB-BC-CD, it is assumed that the latest travel time data are observed on BC and CD. As stated in Section 3, T copula is selected as the multidimensional copula for estimation of the three-link set AB-BC-CD. The parameter estimations for the T copula according to the two-step maximum likelihood procedure are conducted and displayed in Table 3.9.

Then based on linear combination rule, the modified estimation is given as follows:

\[
p(t_{AB} < x) = f_{AB|BE}(x)a_1 + f_{AB|BC,CD}(x)a_2
\]
Table 3.9: Parameter estimates for T copula

\[ \begin{array}{ccc}
\nu &=& 2.3409 \\
\rho &=& 1.0000 \\
& & 0.8662 \\
& & 0.8106 \\
& & 0.8662 \\
& & 1.0000 \\
& & 0.8671 \\
& & 0.8106 \\
& & 0.8671 \\
& & 1.0000 \\
\end{array} \]

The parameter \( \alpha \) is the weight given to individual estimation of each source. The weights can be set by normalizing the number of observations used for estimation in the two link sets. The parameters are \( \alpha_1 = \frac{546}{546+120} \) and \( \alpha_2 = \frac{120}{546+120} \). Based on the weighting scheme, the estimated conditional travel time distribution that is based on T copula and the three-link path AB-BC-CD is shown in Figure 3.9. Then a combined estimated conditional travel time distribution is obtained by calculating a weighted average between the T and BB1 estimator. Decision statistics for the two estimated distributions (Based on T and Combined) are calculated in Table 3.10.

Figure 3.8: Combined Conditional probability density function (Green) compare to two original estimation by T copula (Red) and by bb1 copula (Blue)

If travelers want to consider the traffic conditions on many links together, ideally they need to study the dependence among them all together. However, due to data limitations, it
Table 3.10: Decision statistics for mean-variance rule on AB by different estimation procedures

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Mean</th>
<th>Variance</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>266.4</td>
<td>239.04</td>
<td>15.4609</td>
</tr>
<tr>
<td>Combined</td>
<td>265.04</td>
<td>618.6</td>
<td>2.8716</td>
</tr>
</tbody>
</table>

is usually hard to estimate such high-dimensional copulas. In this case, the scheme proposed in this section can provide a practical approximation. The weighted average of estimates derived from different, relatively low-dimensional link sets considers the different dependent structures among the link and its neighbors. The more data and links are used to generate the estimation, the more reliable the estimated travel time distribution in the focused link is. In this way, the travelers can make trade-offs between accuracy and model complexity.

3.5.2 Similarity-based copula reconstruction

Another challenge when applying this method to the whole network is that there are not enough data to estimate the dependent structure for certain link sets for which travel time observations is too limited. Such link sets are called blind intersections/paths. For such blind intersections, the similarity between intersections can be considered. The copula type and copula parameters of similar intersections are used in them to generate an approximate estimation. The assumption is formally stated below:

**Assumption 3.5.1** Factor Determination (FD): There exists similarity in travel time dependence structure between link sets that satisfy similar physical conditions described by the following factors: 1 Each of the two link sets contains the same number of links; 2 The corresponding links in the two sets are similar in type. 3 The topologic relationship between the links in the link sets is similar (geometrical angle, etc.). 4 The link wise distances of the two link sets are similar.
This assumption implies a factor model in which the dependence structure parameters are responses and the physical attributes of a given link set (geometrical relationship, attributes of the links, distance etc.) are the independent variables. A generalized regression analysis of these factors can be conducted, and the significance of each factor can be studied. Due to limitations in the data, the numerical regression analysis will be conducted will be in future research. Below, a numerical example illustrates how to generate travel time estimation for a blind intersection assuming a similar intersection calibrated with a copula model has been identified.

Supposing that there are not enough data on link AB, and, hence, the copula parameters cannot be obtained correctly. A BB1 copula is estimated using the data on link set GH-GK, and it is then used in AB-BE. The copula parameters for GH-GK are $\theta = 4.10e-007$ and $\delta = 1.33$. The tail dependence parameters are $\lambda_L \approx 0$ and $\lambda_U = 0.32$. The copula has upper tail dependence similar to the one estimated from AB-BE but almost no lower tail dependence. As the simultaneous occurrence of high travel time is of more concern for travelers, this copula can be a reasonable substitute for the dependent structure on AB-BE. The difference between this estimated copula and the one associated with AB-BE pair is shown in Figure 3.9.

Figure 3.9: Difference in estimation between the borrowed copula(red) and the original copula(blue)) Empirical pdf(blue);Estimated pdf(red)
The estimated conditional distribution on AB when the observation on BE is 317 using the copula from the GH-GK pair is in Figure 3.9, (Green is the estimated conditional distribution), and the decision statistics in the mean-variance approach are obtained as $\mu_{AB} = 258.9$, $\sigma^2 = 5894$ and $\sigma = 76.77$. The estimation is slightly larger than the previous observation (251). Considering the availability of data and the stochastic nature of travel time, the estimation here is still reasonably accurate. Similarity-based dependent structure analysis is quite promising; as the example indicates it is an effective way to tackle insufficient data over large-scale transportation networks.

In all, the research in this chapter sets up a framework to forecast the travel time distribution for a given link. Copula methods are selected, as they can model sparse GPS travel time data effectively and model marginal distribution and dependent structure separately. Moreover, copula parameters can be used for other intersections based on similarity principles. In the next chapter, a generalized problem is studied: A traveler on a trip may be interested in estimating the travel time distribution in downstream paths that reach his final destination. To address this issue, multiple scenarios of multiple links in a given path should be studied, conditioned on different stress levels of the overall traffic conditions. The corresponding models will be studied in the following chapters.
Chapter 4

Path Travel Time Distribution
Estimation through Gaussian Copula Mixture Models

This chapter develops Gaussian copula mixture models (GCMM), which are used to model the interdependence of travel time distribution among neighboring links in a given path. Path travel time distributions are then generated through simulation using the estimated dependent structure. In Section 4.2, a scenario-based GCMM is studied. In this model, travel time is assumed to be a function of the systematic factor (path scenario). A set of path scenarios is considered first; in each path scenario, a Gaussian copula is used to model the dependence between the constructing links, and conditional path travel time distribution is estimated. The overall path travel time distribution is then constructed by integrating the conditional path travel time distributions over the set of path scenarios. In Section 4.3, a more general GCMM is studied, and two versions of expectation maximum methods are designed to fit the model to the empirical data. These estimators enable flexible estimation of GCMM based on empirical data without categorizing data into scenarios.
When estimating the GCMM models, two methods are introduced to address the problem of data insufficiency: first, pseudo path travel time observations are constructed using the travel time experienced by different travelers who enter and exit the path at the same location around the same time so that more path travel time observations can be obtained; second, Lasso method is used to estimate the Gaussian copula parameter matrix when data is limited (if the marginal distributions demonstrate normal distribution, it is the usual covariance matrix for joint normal distribution).

4.1 The Gaussian Copula Mixture Model (GCMM)

This section develops the universal model structure-Gaussian copula mixture model. The scenario-based model in Section 4.2 can be viewed as a special case of the model defined herein. The general format is studied extensively in Section 4.3. This section also assembles the procedures to generate a path travel time distribution from an estimated GCMM. This simulation method is shared by all the experiments in this chapter.

4.1.1 Definition of a Gaussian copula mixture model

A Gaussian copula mixture model (GCMM) consists of a weighted sum of a finite number of joint distributions whose dependent structures are modeled by Gaussian copulas. A GCMM is a generalization of the usual Gaussian mixture models (GMM). When the marginal distributions are restricted to be Gaussian as well, the model reduces to GMM.

To begin, the multivariate Gaussian copula is defined by the following probability function:

\[
F(u|P) = \int_{-\infty}^{\psi^{-1}(u_1)} \cdots \int_{-\infty}^{\psi^{-1}(u_d)} \frac{1}{(2\pi)^{n/2} \sqrt{\det P}} \exp\left(-\frac{1}{2} v^T P^{-1} v\right) dv
\]  (4.1)
whose density is given by

\[
f(u|P) = \frac{1}{(2\pi)^{n/2} P^{1/2}} e^{\exp(-\frac{1}{2} u^T P^{-1} u)} \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi}} e^{\exp(-\frac{1}{2} (\Psi^{-1}(u_d))^2)}
\]  

Please notice the parameter matrix \(P\) is the correlation matrix if the random variables are standard normal distributed.

Then with suitable variable transformation and considering the density of the marginal distributions, a GCMM model for the joint distribution of a random vector \(x\) can be defined as follows:

\[
F(X|\pi) = \sum_{k=1}^{K} \pi_k \int_{-\infty}^{Y_{k1}} \cdots \int_{-\infty}^{Y_{kd}} \frac{1}{(2\pi)^{n/2} P_k^{1/2}} e^{\exp(-\frac{1}{2} Y_k^T P_k Y_k)} dY \]  

where

\[
x = [x_1 \ldots x_d] \text{ is the marginal observation;}
\]

\[
Y_k = [Y_{1d} \ldots Y_{kd}] \text{ is the vector of the transferred data.}
\]

\[
Y_{kd} = \Psi^{-1}(F_{kd}(x_d)) \text{ is the d-th dimension of the transferred data.}
\]

\[
Z_{kd} = \frac{\partial F_{kd}}{\partial x}(x_d) \text{ is the density of the marginal distribution.}
\]

Its density is given by

\[
f(X|\pi) = \sum_{k=1}^{K} \pi_k \frac{1}{(2\pi)^{n/2} P_k^{1/2}} e^{\exp(-\frac{1}{2} Y_k^T P_k Y_k)} \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi}} e^{\exp(-\frac{1}{2} (Y_{kd})^2)}
\]  

In next few sections, a specific GCMM model is studied first, in which each Gaussian copula is estimated through the Lasso method after classifying observations into pre-defined path scenarios, and the weights of copulas are corresponding to the historical frequency of the
path scenarios. The general GCMM model is studied later, in which traffic scenarios on each link and the Gaussian copulas are all estimated using expectation maximum methods. This process optimizes marginal traffic scenario definitions, the weights and the parameters of the Gaussian copulas in a recursive fashion. Similar to GMM model, data are not classified to a specific scenario but instead they belong to different scenarios with different probabilities.

4.1.2 Estimation of the path travel time distribution based on a given GCMM model

Given the GCMM that describes the dependent structure in travel time between links in a given path, the path travel time distribution is generated according to the following procedures:

1. Given one of the copulas with parameter $P_k$ in the mixture, simulate a $N$-dimensional data record from Gaussian joint distribution.

2. Conduct monotone transformations $F_i^{-1}(\Psi_i(X_i))$ to convert the data into random variables with the marginal distribution.

3. Sum such marginal variables together to get a sample of the path travel time, and the path travel time distribution is then estimated by the histogram or kernel-smoothing estimator.

4. Conduct Steps 1 to 3 for each copula in the mixture, and sum the $k$ distributions according to the weights in the GCMM by the law of total probability.

In theory, multidimensional integration can be used to generate the distribution of the path, but the simulation procedure is more computationally efficient.
4.2 A Gaussian copula mixture model based on predefined link travel time distributions

The Gaussian copula mixture model in this section is defined based on different path scenarios. The fundamental assumption is that a probe/traveler may encounter different path scenarios in his trip and the probabilities of path scenarios that a probe may encounter in its trip are used as the weights to aggregate the scenario-specific path travel time distributions for the overall path travel time distributions: First, different path scenarios are defined to represent the different traffic conditions along it. For each scenario, there is a stationary dependent structure between the travel times of the links along the path, and a Gaussian copula is used to model this dependent structure. A conditional path travel time distribution is estimated using such a structure for each path scenario. Then, the weights $\pi_k$ in the GCMM are estimated through the frequency of these scenarios. In other word, the overall path travel time distribution is estimated by integrating the conditional path travel time distributions using the distribution of the systematic factor. Usually historical measures can be used to describe the observed path performance. More details are discussed in the following paragraphs:

1. For each link, there are several levels of congestion, each of which yields a different conditional link travel time distribution. A path scenario is defined as a stationary joint distribution between the congestion levels of the constructing links of the path. This concept is illustrated in Figure 4.1: in this figure, a three-link path is displayed with three path travel time observations. Each path observation is a vector of the travel time of the three links experienced by a traveler. The side length of a triangle means the travel time which a particular traveler takes to traverse the link completely. The link travel time observations are classified into a normal traffic scenario(1) and a congestion scenario(2) There are three path observations: two path observations belong to the path scenario (1,1,2) and the third path observation belongs to scenario (1,1,1).
These data vectors are used to estimate the joint distribution between link travel times in the corresponding path scenarios.

2. A Gaussian copula structure exists between the link distributions for a given path scenario. The scenario-specific path travel time distribution can be estimated using this structure. This step is illustrated in Figure 4.2: in this figure, there are three link travel time distributions: that of the normal scenario of link 1, that of the normal scenario of link 2 and that of the congestion scenario of link 3, are aggregated using a Gaussian copula. The outcome is the path travel time distribution of that path scenario.

3. The overall path travel time distribution can be estimated by integrating scenario-specific path travel time distributions using suitable weights. As stated at the beginning of this section, the fundamental assumptions of this treatment is that a probe may encounter different path scenarios in his trip; the probabilities of these path scenarios...
are then used as the weights to aggregate the scenario-specific path travel time distributions for the overall path travel time distributions. If there are finite path scenarios, the overall path travel time distribution can be estimated using the weighted sum of the corresponding scenario-specific travel time distributions, where the weights are the historical frequencies of the path scenarios, which is illustrated in Figure 4.3.

Each step is addressed in more detail in the following subsections:
4.2.1 Scenario decomposition and summation

First, a path scenario is defined. A path scenario is defined as a specific state in which the travel time of the links in the path are subject to a specific joint distribution. In a path scenario, the marginal link travel time distributions are all stationary with respect to the traffic condition implied by the path scenario and the copula structure is stationary. In mathematical terms, a path scenario specifies the following aspects:

• The travel time distribution in the links of the path: \(\{T_{s1}, \ldots, T_{sI}\}\);

• The dependence structure between these link travel times. This dependence structure is described by a copula \(C^s\).

Internally, path scenarios are specified by the partition of arrival times to the links in the path. During a trip, the arrival times to the consecutive links in the path form a high dimensional space. It is assumed the joint distribution of these link travel time is stationary on the partition of this space. To clarify, a partition of a high dimensional space means a group of subsets of this space, it does not contain the empty set and: First, the union of these subsets is equal to the space; Second, the intersection of any two distinct subsets in this group is empty. Applying this concept to the space of the arrival times to the consecutive links in a path, we get a partition of the space of the arrival times and joint travel time distribution is assumed to be stationary on a given subsets in this partition.

Externally, path scenarios are determined by a set of factors, including time of day at which the trip starts, the weather conditions and whether there is any special event happening on the path. The changes of these external factors determines the total number of path scenarios and the parameters of path scenarios.

Consider an example of three consecutive links in a path, if measured from the time that the trip starts, the arrival times to the three links form the three dimensional space \([0]x[0, \text{max } T_1]x[T_1, \text{max } T_2]\), where \(T_1, T_2\) are the travel time to link 1 and link 2 in a trip.
following this path. Suppose under certain external conditions, the arrival time to the second
link is subject to a stationary distribution with parameters $P_1$; the arrival time to the third
link is a stationary distribution with parameters $P_2$; the joint distribution of the travel times
on the three links is stationary: a path scenario can be defined under such situation.

Finally, the overall path travel time distribution is a weighted average of the path travel
time distributions in all path scenarios defined under different external conditions. This esti-
mator assembles the unknown high dimensional joint distribution by parts. Comparatively,
traditional recursive convolution of link travel time distribution failed numerically as the
joint distribution of link travel times in a path is unknown.
According to the discussion above, the estimator for path travel time distribution is given as follows:

\[
P(\sum_{i=1}^{l} T_{t_i} \leq t) = \int_{t_1 \leq t_2 \leq ... \leq t_I} \sum_{(s_1,\ldots,s_I)} P(\sum_{i=1}^{l} T_{t_i} \leq t | (S^1_{t_1} = s_1, \ldots, S^I_{t_I} = s_I)) P(S^1_{t_1} = s_1, \ldots, S^I_{t_I} = s_I) dt_1 \ldots dt_I
\]

Conditioned on the arrival times to the links first, then conditioned on the scenarios that the traveler will meet at their arrival times. Denote \( t_i \) as arrival time, \( s_i \) as travel time.

\[
P(\sum_{i=1}^{l} T_{t_i} \leq t) = \sum_{(s_1,\ldots,s_I)} \sum_{j=1}^{J} \int_{(t_1,\ldots,t_I) \in Q_j} P(\sum_{i=1}^{l} T_{t_i} \leq t | (s_1, t_1, \ldots, s_I, t_I)) \Pi(s_1, t_1, \ldots, s_I, t_I) dt_1 \ldots dt_I
\]

As the probability is positive, integral and summation can exchange.

Assuming the dependent structure is stationary on the in each path scenario.

A series number \( j \) is given to the path scenarios.

\[
P(\sum_{i=1}^{l} T_{t_i} \leq t) = \sum_{(s_1,\ldots,s_I)} \sum_{j=1}^{J} \int_{(t_1,\ldots,t_I) \in Q_j} P(\sum_{i=1}^{l} T_{t_i} \leq t | (s_1, \ldots, s_I, Q_j)) \Pi(s_1, \ldots, s_I, Q_j) dt_1 \ldots dt_I
\]

Integrating over all possible path scenarios \( Q_j \).

\[
\sum_{k=1}^{K} P(\sum_{i=1}^{l} T_{t_i} \leq t | S_k, Q_k) \pi(S_k, Q_k)
\]

The overall path travel time distribution is a weighted sum of the path travel time distributions in different path scenarios.

Consistent with the discussion, the last formula above means the overall path travel time distribution can be calculated as the weighted average of path scenarios where weights are calculated based on the frequency of these path scenarios (systematic factor).
4.2.2 Scenario-specific estimation of conditional path travel time

Within each path scenario, the distribution of each link travel time is estimated via one-dimensional kernel estimators and the dependent structure between the travel times of different links in the path is estimated through a Gaussian copula. In this section, related procedures are addressed: The construction of path travel time observations are addressed first, followed by the methods and assumptions that are used to estimate the copula parameters based on such path travel time observations. In this process, two special treatments are used to tackle the data insufficiency problem:

1. Construction of pseudo path travel time observations to maximize number of path travel time observations.

2. Estimation of the Gaussian copula parameters using Lasso method to avoid singularity in parameter estimation due to data insufficiency.

4.2.2.1 Construction of the data vector for one path scenario

To estimate the dependent structure between the travel time of the links in a path, travel time observations on different links in the path within the given path scenario should be studied together. Consider a traveler traversing through a three-link path within a given path scenario, if travel time on the first link is 80 seconds, travel time on the second link after the traveler traverses the first link is 60 seconds, and travel time on the third link after the traveler traverses the second link is 75 seconds, then the path travel time observation which can be used to estimate the copula for this path scenario is (80, 60, 74).

The challenge when constructing such observations is that the number of observations for a long path is very limited. On the other hand, for a longer path, more observations should be used to estimate the high dimensional copula than that for a shorter path. Such data insufficiency is related to how the data set was constructed:
1. In the original data set, the path travel time are measured by short trips of pilot cars within a certain time limit. Due to the time limit, the number of links that the pilot cars can traverse is limited. As a result, if Path A contains more links and is longer than Path B, it is very likely that there are less path travel time observations which transverse path A than those which traverse path B.

2. As a path observation is constructed using the travel time observations for the links in the path, the number of available path travel time observations is less than the number of available link travel time observations on any link in the path. If the number of link travel time observations on one link is given, the number of path travel time observations for any path that contains that link is smaller or equal than this number.

To address this data insufficiency problem, pseudo path observations are constructed in this study. By constructing additional pseudo path travel time observations, more path data can be used to estimate the copula parameters. Related definitions are given below:

**Definition 4.2.1** An original observation and the original frequency $\pi$ for a given path scenario

1. The original observations for the given path are the travel time observations generated by travelers who traversed the given path p themselves in a trip. These observations are added to the observation data set and the count $N_i$ for the corresponding path scenario increases by 1.

2. The original frequency of a given path scenario, $\pi_i$, is calculated based on the counts $N_i$ of all path scenarios.

$$\pi_i = \frac{N_i}{\sum_j N_j}$$  \hspace{1cm} (4.5)
An original path travel time observation is the path travel time experienced by a traveler. As the data set is sparse, the number of such original path travel time observations are limited. Pseudo path travel time observations are the path travel time observations which are constructed using the experienced travel time of different travelers under the condition that the exiting of the former traveler is close adequately to the entering of the successor both in time and location. That is, pseudo path travel time observations are constructed by connecting the experienced link travel time of different travelers along a given path. To construct such pseudo observations produces more data to estimate path travel time distributions. Related definitions and properties are discussed below.

**Definition 4.2.2** A pseudo observation and the pseudo frequency $\hat{\pi}$ for a given path scenario

1. When traveler $A$ exits the given path $p$ before he traverses the path completely, the second traveler $B$ enters the immediate downstream link of the place that $A$ exits, and $B$ traverses until the end of the path $p$.

2. The time that traveler $A$ leaves the $n$-th link in the path and the time that traveler $B$ enters the immediate downstream link($(n+1)$-th) is with mean of zero and bounded by $\theta_0$, i.e., $E\theta = 0$ and $|\theta| < \theta_0$.

3. The pseudo path travel time vector is the vector of the link travel time of traveler $A$ and traveler $B$, i.e. $[T_A^1, \ldots, T_A^n, T_B^{n+1}, \ldots, T_B^N]$, then this data vector is added into the observation data set, and all the original observations of $p$ are added into the data matrix as well. The frequency count $\hat{N}_i$ is obtained.

4. If traveler $B$ does not traverse the path completely either, and traveler $C$ enters the immediate downstream link of the link where traveler $B$ left the path, traveler $C$’s travel time is recorded into the data vector. This process continues until the full specified trip $p$ is traversed by a sequence of travelers $A, B, C, \ldots$. The pseudo path travel time observation is the vector of their link travel times respectively, and the pseudo count of
the path scenario $\hat{N}_i$ increases by 1.

5. Finally the pseudo frequency of this path scenario $\hat{\pi}_i$ is calculated based on the pseudo count $\hat{N}_i$ of all path scenarios. And such pseudo count include both the original path travel time observations and the pseudo path travel time observations.

$$\hat{\pi}_i = \frac{\hat{N}_i}{\sum_j \hat{N}_j}$$  \hfill (4.6)

This procedure is illustrated in Figure 4.4

![Figure 4.4: Generate pseudo observations](image)

Some related theoretical properties for such construction are studied in Section 5.3.4. It is shown that under suitable assumptions, the estimation based on pseudo path observations converges to the underlying true value as the number of observations increases to infinity.

4.2.2.2 Lasso estimation of the Gaussian copula

Although pseudo path travel time observations are constructed to obtain more observations, data vectors may be still very limited for a long path. As a result, beyond constructing the pseudo path travel time observations, special estimation method should also be used to obtain reliable copula parameters, which contributes to tackle the potential data insufficiency.
problem. Under this context, a Lasso estimator is used to estimate the Gaussian copula parameters within each path scenario. Related background and mathematical details are discussed below:

1. First, the Gaussian copula parameter can be estimated through the following monotonic transformation:

\[ \Sigma = \frac{1}{n} Z^T Z - \frac{1}{n} Z^T 1 1^T Z \]

where \( Z = \Psi^{-1}(F(X)) \)

\( F = (F_1, F_2, \ldots, F_d) \) are CDF of the link travel time in that path scenario.

Intuitively, this monotone transformation transforms any non-Gaussian variable into standard Gaussian variables. The transformed Gaussian variables share a Gaussian copula so that they are joint Gaussian distributed. Therefore estimating the covariance matrix using Lasso method would yield an estimator of the Gaussian copula parameter matrix for the original random variables. The following theorem validates the intuitions above more rigorously:

**Theorem 4.2.1** Any covariance matrix estimation method can be applied to the estimation of Gaussian copula for non-Gaussian marginal data whose dependent structure is subject to a Gaussian copula, under the condition that monotone transforms are conducted on the marginal observations using the corresponding marginal cumulative distribution functions.

Proof to Theorem 4.2.1:

An n-dimensional joint Gaussian distribution is n Gaussian marginal distributions composed by a Gaussian copula. If any data \( X_i \) with Gaussian copula as their dependent structure but with non-normal marginal distributions \( F_i \), then \( F_i(X_i) \) is a uniform distributed random variable, and \( \Psi^{-1}(F_i(X_i)) \) will be subject to a Gaussian distribution, this holds for every \( i \). So by definition the n-dimensional data \( \Psi^{-1}(F_i(X_i)) \)
are of $n$-dimensional joint Gaussian distribution, and, therefore, any covariance matrix estimation method can be applied. The parameter matrix for Gaussian copula can be obtained by converting the estimated covariance matrix into the corresponding correlation matrix.

Q.E.D.

2. After the monotone transformation, the parameter matrix of the Gaussian copula is obtained by estimating the correlation matrix and covariance matrix of the transformed standard normal variables.

The issue in this estimation process is when the number of data is equal to the number of dimensions $n \sim p$, the estimated covariance matrix $\Sigma$ is ill-conditioned and hence is not considered as a good estimator for the covariance matrix. When $n \preceq p$, the empirical covariance $S$ is singular. To solve this problem, Lasso is introduced: instead of estimating the covariance matrix, Lasso method estimates the inverse of covariance matrix directly using optimization techniques so that the estimated covariance matrix is always invertible.

In more detail, Lasso method is used to yield a sparse, invertible estimate covariance matrix by maximizing an objective function penalized by the $L^1$ norm of the inverse of the covariance matrix $\Sigma^{-1}$. The optimization problem is given below, Friedman et al. (2008), Banerjee & Natsoulis (2006) and Banerjee & El Ghaoui (2008):

$$\max_X \log(det X) - tr(\Sigma^TX) - v|X|_1$$

(4.7)

where $|X|_1$ is the sum of the terms in $X$.

Equation (4.7) can be solved in a block coordinate descent algorithm as indicated in Banerjee & El Ghaoui (2008). In this process, the problem in focus is a series of $L^1$ constrained problems called lasso and Least Angle Regression (LAR) or other Lasso solvers can be used to obtain the solutions to them Efron et al. (2004). In this research,
Matlab version of Lasso solver is used for experiments.

This procedure is tractable because:

1. $X$ is the inverse covariance matrix of the transferred data. By estimating $X$ and keeping it invertible, $X^{-1}$, which is the covariance matrix of the data is always invertible. On the other side, the initial sample covariance matrix $\Sigma$ is not necessarily invertible when the number of common observations is smaller than the number of dimensions.

2. The change of the parameter $v$ changes the scarcity of the estimated covariance matrix. Small correlation pairs between faraway links can be adjusted to zero while keeping the covariance matrix positive definite. This property leads to parsimonious models for travel time research.

### 4.2.3 Properties of the scenario-based GCMM model

In this section, the properties of the path travel time estimator introduced in previous sections are studied. First, the properties when there is only one scenario are discussed; Second the properties when there are infinite scenarios are addressed.

#### 4.2.3.1 Main scenario analysis

By definition, the scenario based GCMM model is composed of several Gaussian copulas for different path scenarios. To study the property of the estimator thoroughly, an error analysis is first conducted for the case in which there is only one path scenario. In this case, the scenario-specific mean for each link’s travel time is set as its unconditional mean of the corresponding link, and only one Gaussian copula is used to model the dependence between the travel time of links. Therefore, the estimated path travel time can be expressed as $\sum_{i=1}^{N} T_{ti}$ and $t_i = \sum_{j=1}^{l_t} ET_{t_j}$. On the other hand, the true path travel time is calculated
by summing the travel time on each link $i$, and the entering time to link $i$ depends on the travel time on previous links $\sum_{i=1}^{N} T_{s_i}$ and $s_i = \sum_{j=1}^{i} T_{s_j}$. Based on these formulas, the bias of the estimator is calculated, and a bound to the estimation error is given.

**Theorem 4.2.2** Define $K_i(x)$ such that

$$
\int P(T_{s_i} = x|s_i = a)P(s_i = a)da = K_i(x)P(T_{s_i} = x|s_i = t_i)P(s_i = t_i)(t_i - mins_i)
$$

Define $k_i = \inf_x K_i(x)$, $K_i = \sup_x K_i(x)$ and

$$
\Delta^L_i \doteq E(T_{s_i})(k_i P(s_i = t_i)(t_i - mins_i) - 1)
$$

$$
\Delta^U_i \doteq E(T_{s_i})(K_i P(s_i = t_i)(t_i - mins_i) - 1)
$$

$$
\Gamma^L_i \doteq E(T^2_{s_i})(k_i P(s_i = t_i)(t_i - mins_i) - 1)
$$

$$
\Gamma^U_i \doteq E(T^2_{s_i})(K_i P(s_i = t_i)(t_i - mins_i) - 1)
$$

then the total mean error of the estimation is bounded by

$$
\sum_{i=1}^{N} \Delta^L_i \leq \sum_{i=1}^{n}(ET_{s_i} - ET_{t_i}) \leq \sum_{i=1}^{N} \Delta^U_i
$$

The error of each marginal variance is bounded by

$$
\Gamma^l_i - \Delta^U_i (ET_{s_i} + ET_{t_i}) \leq \text{Var}T_{s_i} - \text{Var}T_{t_i} \leq \Gamma^U_i - \Delta^L_i (ET_{s_i} + ET_{t_i})
$$

for each $i$
proof:

\[
\int P(T_{s_2} = x|T_{s_1} = a)P(T_{s_1} = a)da = K_2(x)P(T_{s_2} = x|T_{s_1} = ET_{t_1})P(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{t_1}) = K_2(x)P(T_{s_2} = x|T_{s_1} = ET_{t_1})P(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{t_1})
\]

\(K_2(x)\) is a modifying constant determined by the distribution \(s_2 = s_1 + T_{s_1} = T_{s_1}\) such that

\[
\int P(T_{s_2} = x|T_{s_1} = a)P(T_{s_1} = a)da = K_2(x)P(T_{s_2} = x|T_{s_1} = ET_{s_1})P(T_{s_1} = ET_{t_1})(ET_{s_1} - minT_{s_1})
\]

Since \(s_1 = t_1\), it is therefore the \(K_2(x)\) such that

\[
\int P(T_{s_2} = x|T_{s_1} = a)P(s_2 = a)da = K_2(x)P(T_{s_2} = x|s_2 = t_2)P(s_2 = t_2)(t_2 - mins_2)
\]

Then

\[
ET_{s_2} = \int xK_2(x)P(T_{s_2} = x|T_{s_1} = ET_{t_1})P(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{t_1})dx = \int xP(T_{s_2} = x|X_{s_1} = ET_{t_1})K_2(x)dxP(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{t_1})
\]

As when \(s_1 = t_1\) and \(T_{s_1} = ET_{s_1} = ET_{t_1}\) we have

\[s_2 = s_1 + T_{s_1} = s_1 + ET_{s_1} = t_1 + ET_{s_1} = t_2\]

we have

\[P(T_{s_2} = x|T_{s_1} = ET_{t_1}) = P(T_{t_2} = x)\]
and

\[ ET_{s_2} = \int x K_2(x) P(T_{t_2} = x) dx \mathbb{P}(T_{s_1} = ET_{t_1})(ET_{t_1} - \min T_{t_1}) \]

Take \( K_2 = \sup_x (K_2(x)) \)

\[ \Delta_2 = ET_{s_2} - ET_{t_2} \]
\[ = \int x K_2(x) P(T_{t_2} = x) dx \mathbb{P}(T_{s_1} = ET_{t_1})(ET_{t_1} - \min T_{s_1}) - \int x P(T_{t_2} = x) dx \]
\[ \leq \int x P(T_{t_2} = x) dx (K_2 P(T_{s_1} = ET_{s_1})(ET_{t_1} - \min T_{s_1}) - 1) \]
\[ = E(T_{t_2}) (K_2 P(T_{s_1} = ET_{s_1})(ET_{t_1} - \min T_{s_1}) - 1) \]
\[ = E(T_{t_2}) (K_2 P(s_2 = t_2)(t_2 - \min s_2) - 1) \]

Similarly, for lower bound

\[ \Delta_1 \geq E(T_{t_2})(k_2 P(s_2 = t_2)(t_2 - \min s_2) - 1) \]

where \( k_i = \sup_x K_i(x) \)

Similarly when \( n = i \),

\[ \int P(T_{s_i} = x|s_i = a) P(s_i = a) da = K_i(x) P(T_{s_i} = x|s_i = t_i) P(s_i = t_i)(t_i - \min s_i) \]

Since

\[ P(T_{s_i} = x|s_i = t_1 + \sum_{j=1}^{i-1} ET_t_j) = P(T_{t_i} = x) \]
as we have $s_i = s_1 + \sum_{j=1}^{i-1} T_{s_j} = t_1 + \sum_{j=1}^{i-1} ET_{t_j} = t_i$

$$ET_{s_i} = \int K_i(x) P(T_{s_i} = x | s_i = t_i) P(s_i = t_i)(t_i - mins_i) \, dx$$

$$= \int x K_i(x) P(T_{t_i} = x) dx P(s_i = t_i)(t_i - mins_i)$$

And then take $K_i = \sup_x K_i(x)$

$$\Delta_i = ET_{s_i} - ET_{t_i}$$

$$= \int x K_i(x) P(T_{t_i} = x) dx P(s_i = t_i)(t_i - mins_i) - \int x P(T_{t_i} = x) dx$$

$$\leq E(T_{s_i} | s_i = t_i)(K_i P(s_i = t_i)(t_i - mins_i) - 1)$$

$$= E(T_{t_i})(k_i P(s_i = t_i)(t_i - mins_i) - 1)$$

$$= \Delta_i^L$$

Similarly, for lower bound

$$\Delta_i = ET_{s_i} - ET_{t_i}$$

$$\geq E(T_{t_i})(k_i P(s_i = t_i)(t_i - mins_i) - 1)$$

$$= \Delta_i^U$$

where $k_i = \sup_x K_i(x)$

The sum of $\Delta_i$ can be used as modification to the initial estimate.

$$\sum_{i=1}^{n} ET_{s_i} = \sum_{i=1}^{n} ET_{t_i} + \sum_{i=1}^{n} \Delta_i$$
\[\sum_{i=1}^{n} (E_T s_i - E_T t_i) \leq \sum_{i=1}^{N} \Delta_i \leq \sum_{i=1}^{N} \Delta_i^U\]

Similarly for the variance, there is a bound of error in each of the \(T_{s_i}\). Consider

\[\Gamma_i = E_T^2 s_i - E_T^2 t_i\]
\[= \int x^2 K_i(x) P(T_{t_i} = x) dx P(s_i = t_i)(t_i - \text{mins}_i) \leq \int x^2 P(T_{t_i} = x) dx\]

So

\[\Gamma_i^L \leq \Gamma_i \leq \Gamma_i^U\]

where

\[\Gamma_i^L = E(T_{t_i}^2)(k_i P(s_i = t_i)(t_i - \text{mins}_i) - 1) \quad \Gamma_i^U = E(T_{t_i}^2)(K_i P(s_i = t_i)(t_i - \text{mins}_i) - 1)\]

And then

\[\text{Var}_{T_{s_i}} - \text{Var}_{T_{t_i}} = E_T^2 s_i - (E_T s_i)^2 - (E_T^2 t_i - (E_T t_i)^2)\]
\[= \Gamma_i - \Delta_i (E_T s_i + E_T t_i)\]

So

\[\Gamma_i^L - \Delta_i^U (E_T s_i + E_T t_i) \leq \text{Var}_{T_{s_i}} - \text{Var}_{T_{t_i}} \leq \Gamma_i^U - \Delta_i^L (E_T s_i + E_T t_i)\]

for each \(i\) Q.E.D. These bounds provide estimation for the error, and they help to control estimation error when a limited number of scenarios is used for estimation.
4.2.3.2 Over-all error analysis

In this section, the properties of the estimator when the number of scenarios increases to infinity are studied. The impact of constructing pseudo observations is studied in detail. Under suitable assumptions, the estimator is proven to converge to the true underlying path travel time distribution if the number of path scenarios increases to infinity.

**Theorem 4.2.3** If the overall path travel time distribution is stationary, the frequency estimates of path scenario \(i\) converges to the true occurring probability of the scenario \(i\).

\[ \pi_i = \frac{N_i}{\sum_j N_j} \rightarrow p_i \]

as \(N \rightarrow \infty\)

Proof to Theorem 4.2.3:

By the law of large numbers, the empirical frequency converges to the distribution of the path scenarios. Q.E.D.

Recall that pseudo path observations are constructed by combining the segmented observation of different travelers together to constitute a path observation vector. After considering such pseudo path observations, the following conclusion holds:

**Lemma 4.2.1** Assume the travelers are homogeneous in their driving pattern, and the time mismatch when constructing pseudo path observations is bounded by \(\theta_0\), then the frequency of a path scenario estimated using both original and pseudo path observations converges to the true frequency as number of data points increases to infinity, \(\hat{\pi}_i \rightarrow \pi_i\), as \(N \rightarrow \infty\) if one of the following three conditions are satisfied:

1. if \(\theta_0 \rightarrow 0\), as \(N \rightarrow \infty\);
2. if the weights of pseudo observations exponentially decay, as \( N \to \infty \):

\[
\hat{\pi}_i = \frac{N_i + \exp(-n_i/n_0)\Delta_i}{\sum_j(N_j + \exp(-n_j/n_0)\Delta_j)}
\]

3. if the following set of assumptions hold

(a) the arrival events for different scenarios form an independency.
(b) the entry of a traveler to a given link in a certain scenario is a Poisson process with parameter \( \lambda \)
(c) the behavior of different travelers are independent, including the start and the route choice to different links.

Proof to Lemma 4.2.1:

If \( \theta = 0 \) a.s., then these link observations constitute a perfect path observation, and there is no time lag between the travelers whose sectional trips constitute the pseudo path observations together. The constructed sample is a real sample of the travel time on the path. Since the system is stationary, the occurring frequency of such a path observation is in proportion to the stated probability by the law of large numbers.

If there is an error \( \theta \) between the time when the previous vehicle left the path and the time when the next vehicle entered for the immediate downstream link, and it is further assumed that the time lag satisfies these conditions: \( E\theta = 0 \) and \( |\theta| < \theta_0 \), then the following analysis follows:

By Assumption (b), the arrival of one vehicle at a given link is a Poisson process with parameter \( \lambda \), then when Car 1 traverses Link 1, the probability that more than zero vehicles arrive to Link 2 within the next \( \theta_0 \) seconds is

\[1 - \exp(-\lambda \theta_0)\]
By Assumption (c), the entrance of different vehicles to different links is independent. Consider the path with \( n \) links under a specific path scenario:

If two different sub-trips are needed to cover the whole path, then the probability to get such a pseudo path observation is \( (1 - \exp(-\lambda \theta_0)) \). If three different sub-trips are needed, the probability of getting such a pseudo path observation is \( (1 - \exp(-\lambda \theta_0))^2 \). If \( n \) different sub-trips are needed, the probability is \( (1 - \exp(-\lambda \theta_0))^{n-1} \).

Furthermore, it is assumed that at a given intersection, the probability of keeping to this given path is of probability \( q \), so the probability that it takes \( i \) sub-trips to generate a full path observation is \( q^{I-i} \). And by the law of total probability, the probability to get a pseudo observation is:

\[
p = \sum_{i=1}^{I} q^{I-i}(1 - \exp(-\lambda \theta_0))^{i-1}
\]

By assuming the different vehicles start their trips independently, as in Assumption (c), the probability of getting \( N \) regular observations in the \( T \) vehicles that enter the first link in the given path is a binomial distribution with parameter \( p_R = q^{I-1} \).

The probability of getting \( M \) pseudo observations of the \( T \) vehicles that enter the first link in the given path is a binomial distribution with parameter \( p \).

\[
p_S = q^{I-1} + \sum_{i=2}^{I} q^{I-i}(1 - \exp(-\lambda \theta_0))^{i-1}
\]

By Assumption (b), the arrival of a vehicle to the first link is subject to a Poisson process with parameter \( \lambda \). The overall distribution of regular observations is as follows:

\[
P(n_R = k) = \sum_{n=0}^{\infty} \frac{e^{-\lambda t}}{n!} k! \sum_{k=0}^{n} \frac{e^{\lambda t}}{n} \left( p_R \right)^k (1 - p_R)^{n-k} 1_{k \leq n}
\]
while the overall distribution with the pseudo observations is

\[
P(n_S = k) = \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n} \frac{1}{n} C_n^k p_S^k (1 - p_S)^{n-k} 1_{k \leq n}
\]

Then three alternative sets of conditions are studied, and if one of the three condition sets is satisfied, the convergence holds:

1. First, \( \delta \) can be controlled. When \( \delta_0 \to 0 \), the two binomial distributions above tend to be identical, which implies that the difference between pseudo path observations and original path observations reduces to zero, therefore the frequency of the former converges to the true frequency of the path scenarios: \( \hat{\pi}_i \to \pi_i \).

2. Second, the pseudo path observations can be gradually excluded when the original observations are abundant. In this case, the frequency estimated using both original and pseudo path observations also converges to the true frequency. The formulas are as follows:

\[
\hat{\pi}_i = \frac{\hat{N}_i}{\sum_j \hat{N}_j} = \frac{N_i + \exp(-n_i/n_0) \Delta_i}{\sum_j (N_j + \exp(-n_j/n_0) \Delta_j)}
\]

where \( n_i \) is the number of total number of true path travel time observation and \( n_0 \) is the expected number of observation, beyond which the original observations are used. As \( n_i \) goes to \( \infty \), this frequency converges to the true probability for each scenario. Then the conclusion is \( \hat{\pi}_i \to \pi_i \).

3. Third, the different path scenarios are considered. If in any two scenarios \( S_1 \) and \( S_2 \), Poisson arrival processes take two different parameters \( \lambda_1 \) and \( \lambda_2 \). Suppose the number
of vehicles which enter the first link are $M_1$ and $M_2$ respectively, then the following result can be obtained:

Claim

$$E \left\{ \frac{N_1}{N_2} | M_1, M_2 \right\} \to \frac{M_1}{M_2}$$

as $n \to \infty$

Since $M_1$ and $M_2$ are independently drawn, as $n - > \infty$:

$$E \left\{ \frac{N_1}{N_2} | M_1, M_2 \right\} = E \left\{ \frac{1}{N_2} | M_1 \right\} E \left\{ \frac{1}{N_1} | M_2 \right\}$$

$$E \left\{ \frac{1}{N_1} | M_1 \right\} = M_1 p_R$$

and by the approximation formula in Rempa (2004):

$$E(\frac{1}{N_2} | M_2) = \frac{1}{M_2 p_R} + \frac{1 - p_R}{M_2^2 p_R^2} + \frac{2(1 - p_R)^2}{M_2^3 p_R^3} + \frac{6(1 - p_R)^3}{M_2^4 p_R^4} + \frac{24(1 - p_R)^4}{M_2^5 p_R^5}$$

And therefore:

$$E(\frac{N_1}{N_2} | M_1, M_2) = M_1 p_R \left\{ \frac{1}{M_2 p_R} + \frac{1 - p_R}{M_2^2 p_R^2} + \frac{2(1 - p_R)^2}{M_2^3 p_R^3} + \frac{6(1 - p_R)^3}{M_2^4 p_R^4} + \frac{24(1 - p_R)^4}{M_2^5 p_R^5} \right\}$$

$$= \left\{ \frac{M_1}{M_2} + \frac{M_1(1 - p_R)}{M_2^2 p_R^1} + \frac{M_1(1 - p_R)^2}{M_2^3 p_R^2} + \frac{6M_1(1 - p_R)^3}{M_2^4 p_R^3} + \frac{24M_1(1 - p_R)^4}{M_2^5 p_R^4} \right\}$$

$$\to \frac{M_1}{M_2} \text{ when } M_2 \to \infty$$

Similarly the following formula holds:
$$E\left(\frac{\hat{N}_i}{\hat{N}_2}|M_1, M_2\right) \rightarrow \frac{M_1}{M_2} \text{ when } M_2 \rightarrow \infty$$

So

$$E\left(\frac{\hat{N}_i}{\hat{N}_2}|M_1, M_2\right) \rightarrow E\left(\frac{N_i}{N_2}|M_1, M_2\right) \text{ when } M_2 \rightarrow \infty$$

The above holds for any pair of scenarios. Since $\sum_j \hat{N}_j$ is the sum of independent binomial random variables with parameters $(M_i, P_S)$, it is another binomial random variable with parameter $(\sum M_i, P_S)$. Similarly $\sum_j N_j$ is a binomial random variable with parameter $(\sum M_i, P_R)$. Then,

$$E\left(\frac{\hat{N}_i}{\sum_j \hat{N}_j}|M_i\right) \rightarrow E\left(\frac{N_i}{\sum_j N_j}|M_i\right) \text{ when } M_i \rightarrow \infty$$

Then by definition:

$$\hat{\pi}_i = \frac{\hat{N}_i}{\sum_j \hat{N}_j} \rightarrow \frac{N_i}{\sum_j N_j} \text{ when } M_i \rightarrow \infty$$

This means the time lag $\theta$ when constructing the pseudo path observations does not introduce error in the asymptotic sense to the estimator.

Q.E.D.

Based on the lemma above, the convergence can be proven below:

**Theorem 4.2.4** If the arrival events of different path scenarios form an independency, then
as $N \to \infty$, the estimated path travel time distribution weakly converges to the true distribution.

Proof to Theorem 4.2.4:

By the law of total probability, the estimated cumulative distribution is the weighted sum of the scenario-specific cumulative probability functions for the path travel time:

$$F(x) = \sum_n \pi_n F_n(x)$$

Then the characteristic function of the estimated distribution is calculated as follows:

$$C_n(t) = \int \exp(itx)dF(x)$$

$$= \int \exp(itx) \sum_n \hat{\pi}_n dF_n(x)$$

$$= \sum_n \hat{\pi}_n \int \exp(itx)dF_n(x)$$

$$\to \sum_n \pi_n \int \exp(itx)dF_n(x)$$

$$\to \int \int \exp(itx)1_{x \in V} \mu(dx, dV)$$

$$= \int E(\exp(tx)|x \in V) \mu_V(dV)$$

$$= \mathbb{E} \exp(itx)$$

Here, the fact is used that $\hat{\pi} \to \pi_n$, under Assumption (a). The result is the characteristic function for any distribution. Therefore, the infinite partition will converge to the true distribution of the path scenarios.

Q.E.D.
The model in this section is a special Gaussian copula mixed model where the weights \( \pi_k \) are the frequency of historical occurrence of the path scenarios: pseudo path observations are constructed for more reliable estimation and the dependent parameter \( P_k \) of each copula is estimated through the Lasso method using the data in the corresponding states. It is shown that the error of such an estimator is bounded when number of path scenarios is one, and the estimator converges to the true distribution as number of states increases to infinite.

An important feature of the scenario-based GCMM model in this section is that each data vector is classified precisely into a path scenario for further estimation of copula parameters. If the classification of each data vector is conducted in a probability sense, i.e. a data vector can be classified into different path scenarios with different probabilities, the scenario GCMM model in this section can be generalized to more general GCMM models which are introduced in next section. Such flexibility in classification helps to improve the properties of these estimators when the number of path scenarios is finite.

4.3 A general GCMM and the extended expectation maximum algorithms

In this section, the general GCMM model is introduced. Instead of classifying path observations strictly into distribution scenarios, the general GCMM model automatically fits the empirical data in an integrated fashion and each data point can belong to different scenarios with different probabilities. Essentially, to estimate a general GCMM is to optimize the parameters of the GCMM model (all weights \( \pi_k \) and dependency parameter \( P_k \) of the Gaussian copulas) by the maximum likelihood method using the whole data set. Compared to scenario-based GCMM, there are fewer constraints because of pre-classification of observations into scenarios. Hence, when the number of path scenarios is finite, this optimization with fewer constraints may lead to higher values of the likelihood function and better quality
of fit.

To conduct the estimation of a general GCMM, two estimation algorithms are designed. The difference between the two algorithms concerns whether the link travel time distribution is fixed or not during the iteration. If the link travel time distributions are defined according to transportation context first and hence fixed in the iteration, a fixed-marginal-distribution GCMM can be used to estimate the parameters; if the path scenarios are initially defined by abstract cluster analysis and link travel time distributions are not fixed during the iteration, a varying-marginal-distribution GCMM can be used to estimate the parameters.

4.3.1 Fixed-marginal-distribution GCMM

By fixing the marginal travel time distribution of each link in the path scenario, an expectation-maximum algorithm can be designed to search the optimal parameters of Gaussian copulas, which explains the pattern of historical path travel time observations. It is the specific expectation-maximum algorithm for the likelihood function of GCMM. The steps of the algorithm are designed as follows:

**Fixed-marginal-distribution GCMM**

1. Generate initial clusters for $X_i$ on each link and classify the link travel time data to each clusters

2. Expectation step: update $r_{nk}$

3. Maximum step:
   
   (a) Conduct maximum likelihood estimation for marginal link travel time distribution $F_{X_i}$.

   (b) Conduct maximum likelihood estimation for the Gaussian copula mixture model.

4. Check convergence, if the convergence rule is not satisfied, goto Step 2.
5. Eliminate redundant clusters and copulas.

Below, each step is discussed in detail, and necessary mathematic formulas are introduced:

- **Step 1 Initialization**
  - Get \( \{x_n\} \) and \( n = 1...N \) are the N data points; each record is of dimension \( D \) for the joint structure and some extra observations on each dimension \( y_{ds} \) where \( d = 1...D \) and \( s = 1...S_d \) and \( S_d \) is an arbitrary finite number.
  - Initial clusters are generated on the travel time distribution for each link. Denote \( S_{ds}^m \) is the s-th cluster for the d-th dimension in the m-th iteration. One dimensional Gaussian Mixture Models can be used to generate the clusters for each link travel time distribution.
  - Consider the n links together. For every possible combination of such link-clusters, the dependent structure is modeled using a Gaussian copula with parameter \( P_k \). Meanwhile, a mapping \( M : C_{kd} -> S_dj \) is generated to tell that the d-th dimension of the k-th copula is corresponding to the j-th cluster of the d-th link. A Gaussian copula mixture model is then generated by doing linear combination of these \( K \) Gaussian copulas.
  - The overall likelihood function is denoted as

\[
L = \sum_{n=1}^{N} \ln(\sum_{k=1}^{K} \pi_k \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} (Y_{n,k}^m)^T P Y_{n,k}^m\right) \prod_{i=1}^{D} \frac{Z_{n,ki}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (Y_{n,ki}^m)^2\right))
\]

where \( Y_{n,k}^m = [Y_{n,ki}] \), \( Y_{n,ki} = \Psi^{-1}(F_{ki}^m(x_{ni})) \) and \( Z_{n,ki} = \frac{\partial F_{ki}^m}{\partial x}(x_{ni}) \).

- **Step 2: Expectation.** By taking partial derivatives to the Likelihood function with respect to \( \pi_k \), a ratio \( r_{nk} \) can be defined

\[
D_{nk}^m = \prod_{i=1}^{D} \frac{Z_{n,ki}}{\sqrt{2\pi}} \exp\left(\frac{1}{2} (Y_{n,ki}^m)^2\right)
\]
\[
\frac{\partial L}{\partial \pi_k} = \sum_{n=1}^{N} \frac{r_{nk}}{\pi_k}
\]

where

\[
r_{nk}^m = \frac{D_{n,k}^m \pi_k}{\sum_{j=1}^{K} D_{n,j}^m} \exp\left(-\frac{1}{2} (Y_{nk}^m)^T P_k^{-1} Y_{nk}^m\right)
\]

(4.9)

- **Step 3: Maximum**
  
  - Conduct maximum likelihood estimation for the \(F_{ki}^m\) according to the latest cluster of marginal data. Update \(Y_{n,k} Z_{n,k}\) and \(D_{n,k}\).
  
  - Conduct the maximum for Gaussian copula mixture model and define Lagrange objective as follows:

\[
\bar{L} = L + \lambda \left(\sum_{k=1}^{K} \pi_k - 1\right)
\]

Then

\[
\frac{\partial L}{\partial \pi_k} = 0
\]

\[
\sum_{n=1}^{N} \frac{r_{nk}}{\pi_k} - \lambda = 0
\]

\[
\sum_{k=1}^{K} \sum_{n=1}^{N} r_{nk}^m = \lambda
\]

\[
\sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk}^m = \lambda
\]

\[
\lambda = N
\]
and

$$\pi_k^m = \frac{\sum_{n=1}^{N} r_{nk}^m}{N}$$ (4.10)

Next, consider the partial for $P_k^{-1}$

$$\frac{\partial L}{\partial P_k^{-1}} = \frac{\partial \pi_k^m \sum_{n=1}^{N} r_{nk}^m \exp\left(\frac{1}{2}(Y_{nk} Y_{nk})^T P_k P_k^T\right)}{\partial P_k^{-1}}$$

$$= \frac{\sum_{j=1}^{K} D_{nj}^m \pi_k^m \exp\left(-\frac{1}{2}(Y_{nk} Y_{nk})^T P_j P_j^{-1} P_k P_k^{-1}(Y_{nk} Y_{nk})^T\right)}{\sum_{j=1}^{K} \pi_j^m \sum_{n=1}^{N} r_{nj}^m \exp\left(-\frac{1}{2}(Y_{nj} Y_{nj})^T P_j P_j^{-1} P_k P_k^{-1}(Y_{nj} Y_{nj})^T\right)}$$

$$= \frac{1}{2} \sum_{n=1}^{N} r_{nk}^m (P_k - Y_{nk} Y_{nk}^T)$$

By making the derivative to equal to zero:

$$P_k^m = \frac{\sum_{n=1}^{N} r_{nk}^m Y_{nk} Y_{nk}^T}{\sum_{n=1}^{N} r_{nk}^m}$$ (4.11)

- Step 4: Check convergence by comparing $|L^m - L^{m-1}| < \varepsilon$ or comparing the parameter values. If the convergence criterion is not satisfied, return to Step 2 Expectation.
- Step 5: Eliminate the copulas whose weights are small or whose parameter values do not make sense.

By Theorem 2.14 of White (1994), the estimator in this section is a two-stage quasi-maximum-likelihood estimator for the joint distribution. Given regular conditions and the condition that the second stage maximum(the EM part) is successfully conducted, there exists a function in the function space that the likelihood functions converge to. Furthermore, consider the second stage of this procedure; by the theorem in Boyles (1983) the two-step expectation-maximum estimator converges to the local minimum of the likelihood function.
under suitable technical conditions. Hence, overall, the estimator converges.

### 4.3.2 Varying-marginal-distribution GCMM

In the algorithm in the previous section, the link travel time distributions in each path scenario are fixed before the iteration starts. A further consideration over the algorithm structure indicates that it may be more flexible to update the link travel time distributions in each path scenarios during each iteration of the expectation-maximum algorithm so that a better mixture model can be identified based on the overall data set. This algorithm is useful when the initial path scenario is generated by some statistical cluster analysis and the travel time distributions for each link are not fixed. During the iteration of this algorithm, both link travel time distributions and the parameters of the Gaussian copulas are updated, which leads to potentially higher likelihood values. The steps of the algorithm are designed as follows:

**Varying-marginal-distribution GCMM**

1. Generate an initial cluster of the link travel time distribution for $X_i$ and classify the marginal data to each such clusters.

2. Expectation step: update $r_{nk}$.

3. Maximum step:
   
   (a) Convert the marginal observations into probability values using the marginal cumulative distribution functions.
   
   (b) Conduct maximum likelihood estimation for the Gaussian copula mixture model.

4. Update the marginal link travel time probability distribution functions $F_n$ according to the posterior classification probability $r_{nk}$.

5. Check convergence. If not, goto Step 2.

This algorithm combine the two-step estimator maximum likelihood estimator for copula and the expectation-maximum algorithm for GMM together in each iteration. The following theorem shows the properties of the likelihood function and the convergence property of the new algorithm can be studied based on the results below.

Although the log-likelihood is not convex in $Z_{n,ki}$ or $Y_{n,ki}$, the bound on the log-likelihood function can be given as follows:

**Theorem 4.3.1** Given the $P$ and $x_n$, assume there are bounds to the probability values on $x_n$ for each of the marginal distribution $F_{kd}$

Then the log-likelihood function is bounded above.

Proof:

Consider maximizing the following function

$$L = \sum_{n=1}^{N} \ln \left( \sum_{k=1}^{K} \pi_k \left( \frac{1}{2\pi} \right)^{n/2} P^{1/2} \exp \left( -\frac{1}{2} (Y_{n,k})^T P Y_{n,k} \right) \prod_{i=1}^{D} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} (Y_{n,ki})^2 \right) \right)$$

with the constrains:

$$Z_{n,k} \geq 0, \ Z_{n,k} \leq C, \ Y_{n,k} \geq 0 \text{ and } Y_{n,k} < 1$$

If it is changed into a minimization problem by multiplying the objective by -1, the full Lagrange objective function will be:

$$\hat{L} = -\sum_{n=1}^{N} \ln \left( \sum_{k=1}^{K} \pi_k \left( \frac{1}{2\pi} \right)^{n/2} P^{1/2} \exp \left( -\frac{1}{2} (Y_{n,k})^T P Y_{n,k} \right) \prod_{i=1}^{D} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} (Y_{n,ki})^2 \right) \right)$$

$$+ \sum_{n} \sum_{k} \alpha_{n,k}^T (-Y_{n,k}) + \sum_{n} \sum_{k} \beta_{n,k}^T (Y_{n,k} - 1) + \sum_{n} \sum_{k} \gamma_{n,k}^T (-Z_{n,k})$$

$$+ \sum_{n} \sum_{k} \theta_{n,k}^T (Z_{n,k} - C)$$
with

\[ \alpha_{n,k} \geq 0, \quad \gamma_{n,k} \geq 0, \quad \beta_{n,k} \geq 0, \quad \theta_{n,k} \geq 0 \]

and

\[-Z_{n,k} \leq 0, \quad Z_{n,k} - C \leq 0, \quad -Y_{n,k} \leq 0 \quad \text{and} \quad Y_{n,k} - 1 < 0\]

Then

\[ \frac{\partial L}{\partial Z_{n,kj}} = \frac{1}{S_n} \pi_k \frac{1}{(2\pi)^{n/2} P^{1/2}} \exp\left(-\frac{1}{2} Y_{n,k}^T P Y_{n,k}\right) \prod_{i \neq j} \frac{Z_{n,ki}}{\sqrt{2\pi} \exp\left(-\frac{1}{2} Y_{n,ki}^2\right)} \]

\[ \geq 0 \]

Where

\[ S_n = \sum_{k=1}^K \pi_k \frac{1}{(2\pi)^{n/2} P^{1/2}} \exp\left(-\frac{1}{2} Y_{n,k}^T P Y_{n,k}\right) \prod_{i \neq j} \frac{Z_{n,ki}}{\sqrt{2\pi} \exp\left(-\frac{1}{2} Y_{n,ki}^2\right)} \]

\[ \frac{\partial L^2}{\partial Z_{n,kj}^2} = \frac{1}{S_n^2} (0 - P_{n,kj}^2) \]

\[ \leq 0 \]

where

\[ P_{n,kj} = \pi_k \frac{1}{(2\pi)^{n/2} P^{1/2}} \exp\left(-\frac{1}{2} Y_{n,k}^T P Y_{n,k}\right) \prod_{i \neq j} \frac{Z_{n,ki}}{\sqrt{2\pi} \exp\left(-\frac{1}{2} Y_{n,ki}^2\right)} \]

And

\[ \frac{\partial \hat{L}}{\partial Z_{n,kj}} = -\frac{1}{S_n} \pi_k \frac{1}{(2\pi)^{n/2} P^{1/2}} \exp\left(-\frac{1}{2} Y_{n,k}^T P Y_{n,k}\right) \prod_{i \neq j} \frac{Z_{n,ki}}{\sqrt{2\pi} \exp\left(-\frac{1}{2} Y_{n,ki}^2\right)} - \gamma_{n,kj} + \theta_{n,kj} \]

\[ \gamma_{n,k}^T Z_{n,k} = 0 \]

\[ \theta_{n,k}^T (Z_{n,k} - C) = 0 \]

117
By taking \( \frac{\partial \hat{L}^2}{\partial Z_{n,kj}} = 0 \), there should be the following relationship:

\[
Z_{n,k} = \frac{S_n(-\gamma_{n,kj} + \theta_{n,kj})}{D_{n,k} \frac{1}{2\pi} \exp\left(-\frac{1}{2}Y_{n,ki}^2\right)}
\]

(4.13)

\[
\gamma_{n,k} \frac{S_n(-\gamma_{n,kj} + \theta_{n,kj})}{D_{n,k} \frac{1}{2\pi} \exp\left(-\frac{1}{2}Y_{n,ki}^2\right)} = 0
\]

(4.14)

\[
\theta_{n,k} \left( \frac{S_n(-\gamma_{n,kj} + \theta_{n,kj})}{D_{n,k} \frac{1}{2\pi} \exp\left(-\frac{1}{2}Y_{n,ki}^2\right)} - C_i \right) = 0
\]

(4.15)

The objective \( \hat{L} \) may be minimized in the inner area, that is: (1) \( \gamma_{n,kj} = 0 \) and \( \theta_{n,kj} \neq 0 \) the solution is denoted as \( Z_{n,k}^{b1} \); (2) \( \gamma_{n,kj} \neq 0 \) and \( \theta_{n,kj} = 0 \), the solution is denoted as \( Z_{n,k}^{b2} \); (3) \( \gamma_{n,kj} = 0 \) and \( \theta_{n,kj} = 0 \), the solution is denoted as \( Z_{n,k}^c \).

For \( Y_{n,k} \), the following analysis is conducted:

\[
\frac{\partial L}{\partial Y_{n,k}} = \frac{1}{S_n} B_{n,k} D_{n,k} (-P + I) Y_{n,k}
\]

where \( B_{n,k} = \pi_k \left( \frac{1}{2\pi} \right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}Y_{n,kj}^T P Y_{n,kj}\right) \) and \( D_{n,k} = \prod_{i=1}^{D} \sqrt{2\pi} \exp\left(-\frac{1}{2}Y_{n,ki}^2\right) \) as defined in the previous section.
\[
\frac{\partial L^2}{\partial Y_{n,k}} = \frac{1}{S_n} (B_{n,k}D_{n,k}(-P + I)Y_{n,k}Y_{n,k}^T(-P + I) + B_{n,k}D_{n,k}(-P + I))S_n
\]
\[
- B_{n,k}D_{n,k}(-P + I)Y_{n,k}Y_{n,k}^T(-P + I)B_{n,k}D_{n,k}
\]
\[
= \frac{1}{S_n^2} (B_{n,k}D_{n,k}(-P + I)(Y_{n,k}Y_{n,k}^T)S_n + S_n - Y_{n,k}Y_{n,k}^T)B_{n,k}D_{n,k}(-P + I)
\]
\[
= \frac{1}{S_n} (B_{n,k}D_{n,k}(Y_{n,k}Y_{n,k}^T)S_n + S_n - B_{n,k}D_{n,k}Y_{n,k}Y_{n,k}^T)(-P + I)(-P + I)
\]

Notice here \((-P + I)\) is diagonalizable since \(P\) is the covariance matrix of two normally distributed random vectors. \((-P + I)(-P + I)\) is then positive semi-definite. Define

\[
\Lambda_{n,k} = Y_{n,k}Y_{n,k}^T S_n + S_n - B_{n,k}D_{n,k}Y_{n,k}Y_{n,k}^T
\]

Since \(\frac{1}{S_n} B_{n,k}D_{n,k}\) is positive, \(\Lambda_{n,k}\) will determine the properties of the function with respect to \(Y_{n,k}\).

\[
\frac{1}{S_n} B_{n,k}D_{n,k}(-P + I)Y_{n,k} - \alpha_{n,k} + \beta_{n,k} = 0
\]
\[
\alpha_{n,k}^T Y_{n,k} = 0
\]
\[
\beta_{n,k}^T (Y_{n,k} - 1) = 0
\]

if \(\Lambda_{n,k} \geq 0\), \(\frac{\partial L^2}{\partial Y_{n,k}} \geq 0\) and \(\frac{\partial \hat{L}^2}{\partial Y_{n,k}} \leq 0\)

if \(\Lambda_{n,k} < 0\), \(\frac{\partial L^2}{\partial Y_{n,k}} < 0\) and \(\frac{\partial \hat{L}^2}{\partial Y_{n,k}} > 0\)

Then

\[
Y_{n,k} = \frac{S_n}{B_{n,k}D_{n,k}} (-P + I)^{-1} (\alpha_{n,k} - \beta_{n,k})
\]

(4.16)
\[
\frac{S_n}{B_{n,k}D_{n,k}} \alpha_{n,k}^T(-P + I)^{-1}(\alpha_{n,k} - \beta_{n,k}) = 0 \tag{4.17}
\]

\[
\beta_{n,k}^T\left(\frac{S_n}{B_{n,k}D_{n,k}}(-P + I)^{-1}(\alpha_{n,k} - \beta_{n,k}) - 1\right) = 0 \tag{4.18}
\]

Then \(Y_{n,k}\) can be solved through the above equations. Furthermore, the extreme values of the \(Y_{n,k}\) are considered as follows:

If \(\Lambda_{n,k} \geq 0\) then the data point \(x_n\) is classified as Type 1 for \(k\)-th copula. The objective \(\hat{L}\) is always minimized on the boundary. That is: (1) \(\alpha_{n,k} = 0\) and \(\beta_{n,k} \neq 0\), the solution is denoted as \(Y_{n,k}^{1}\); (2) \(\alpha_{n,k} \neq 0\) and \(\beta_{n,k} = 0\) the solution is denoted as \(Y_{n,k}^{1b}\).

If \(\Lambda_{n,k} < 0\), then the data point \(x_n\) is classified as Type 2 for \(k\)-th copula. The objective \(\hat{L}\) may be minimized in the inner area. That is: (1) \(\alpha_{n,k} = 0\) and \(\beta_{n,k} \neq 0\), the solution is denoted as \(Y_{n,k}^{2}\); (2) \(\alpha_{n,k} \neq 0\) and \(\beta_{n,k} = 0\), the solution is denoted as \(Y_{n,k}^{2b}\); (3) \(\alpha_{n,k} = 0\) and \(\beta_{n,k} = 0\), the solution is denoted as \(Y_{n,k}^{2c}\).

In all cases, the value of the likelihood function is bounded above by a value determined by these finite extreme values in \(Y_{n,k}\) and \(Z_{n,k}\). Q.E.D.

Hence by solving the maximum likelihood problem in each step, historical observations are reclassified in a probabilistic sense and a new set of marginal link travel time distributions is generated for each dimension of the copulas. The log likelihood function is not convex while local optimality can be obtained using the proposed algorithm. Stochastic optimization or heuristic optimization can be further introduced to get better solutions.
4.4 Numerical analysis

4.4.1 Experiments for scenario-based GCMM models

For the experiment on path travel time estimation, a twelve-link path between Allentown and Clinton along Highway 78 is selected to demonstrate the method to estimate the path travel time based on the available GPS data set; two competing paths near Philadelphia are selected to illustrate the decision making process. The networks are shown in Figure 5.6.

![Figure 4.5: The experimental network in New Jersey](image)

(a) A twelve-link path  
(b) The comparison of two paths

First, the sensitivity of the estimation with respect to $\Delta$ is studied in4.6, which is the bound of the time difference when different travelers assemble pseudo path observations together. As $\Delta$ changes, the estimated result does not change significantly. One reason is that as the $\Delta$ changes, the path scenario does not change, and the estimated result is similar. The other reason is that the data set is sparse, and when $\Delta$ changes, not so many new observations are included. $\Delta$ is then set as small portion of the standard deviation in later experiments.

Second, the sensitivity of the estimation with respect to Lasso penalty $\nu$ is studied while fixing $\Delta$. The estimates based on different lasso penalties are shown in Figure 4.7. The
more penalty is used, the more independent the estimated covariance structure tends to be, and the worse the approximation is. This result shows that the dependence between links is heavy, and independent assumption does not work. Generally speaking, small penalty yields better estimation, and \( v \) is set as the smallest value \( v = 0.001 \) in later experiments.

Third, the sensitivity of the estimation with respect to the number of scenarios is studied. The number of congestion levels in a link is one (the main scenario), two, four, eight, and sixteen are considered in each link. The estimated cumulative distribution functions are given in Figure 4.8. The experiment shows that (1) as the number of states in each link increases, the difference between estimated distribution and the empirical distribution tends to be smaller; (2) if assuming the links are independent from each other, the estimated path travel time (the black dots) tends to ignore the upper tails (i.e., it underestimates the probability of congestion); and (3) there is bias when the number of states is limited. This bias can be further reduced using the expectation-maximum algorithms introduced in Section 4.3. Please note, due to constrains on dimensionality, the experiment with respect
Figure 4.7: Change of estimation as the Lasso penalty changes. (Empirical: red dot; Independent: yellow; $\nu=0.001$: Cyan; $\nu=0.01$: red; $\nu=0.05$: Blue; $\nu=0.1$: Black; $\nu=0.5$: Green)

to number of scenarios is conducted on two adjacent links (part of the original 12-link path).

Figure 4.8: Estimated cdf when number of scenarios change: 1: yellow; 2: red; 4: cyan; 8: green; 16: blue; empirical: black; independent: black dots; Left figure displays the lower tail, Right figure displays the upper tail

The probability density function is shown in Figure 4.9. When the number of scenarios
is two, the upper tail of the empirical distribution is not well described. When the number of scenarios is eight, the upper tail is better described by the model.

![Graphs showing travel time distributions](image)

**Figure 4.9:** Left: 2 scenarios per link; Right: 8 scenarios per link; Estimated pdf: red; empirical: green; scenario-specific cdfs: blue

Fourth, to show the comparison of the two competing paths using different decision statistics, the path travel time distribution on two paths are compared based on their main scenarios (the lasso penalty is $v = 0.001$). The estimated path travel time distributions are displayed in Fig 4.10.

Then the decision statistics are calculated to generate routing decision, as in Table 4.1. Although the two paths are not comparable under mean-variance decision rules, under most other decision rules (stochastic dominance, value-at-risk, exponential utility, area-ratio rule, etc.), Path 2 should be selected. The traveler can then select one rule to make his routing decision. Different decision rules may lead to different routing decisions.
Figure 4.10: 12 Path travel time based on the approximation; red: Path 1; blue: Path 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$(\mu, \sigma)$</th>
<th>$\mu + r\sigma^2$</th>
<th>FSD</th>
<th>SSD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$r=0.05$</td>
<td>first violation</td>
<td>first violation</td>
</tr>
<tr>
<td>Path 1</td>
<td>$(1199,130.5)$</td>
<td>2051</td>
<td>$\infty$</td>
<td>$\infty$</td>
</tr>
<tr>
<td>Path 2</td>
<td>$(993.8,145.9)$</td>
<td>2058</td>
<td>739.4</td>
<td>0</td>
</tr>
<tr>
<td>Preference</td>
<td>None</td>
<td>Path 1</td>
<td>Path 2</td>
<td>Path 2</td>
</tr>
<tr>
<td>Exponential utility</td>
<td>$a=-1/1000$</td>
<td>$t_U=1600$</td>
<td>VAR(Quantile)</td>
<td>5%</td>
</tr>
<tr>
<td>Path 1</td>
<td>0.2476</td>
<td>251.5</td>
<td>1832</td>
<td></td>
</tr>
<tr>
<td>Path 2</td>
<td>0.3509</td>
<td>559.5</td>
<td>1394</td>
<td></td>
</tr>
<tr>
<td>Preference</td>
<td>Path 2</td>
<td>Path 2</td>
<td>Path 2</td>
<td></td>
</tr>
</tbody>
</table>

4.4.2 Experiment for general GCMM models

To demonstrate the difference between the Fixed-marginal-distribution GCMM and the Varying-marginal-distribution GCMM, an experiment is conducted: based on two simulated link travel time samples, the model was used to estimate the path travel time distribution. The estimated distribution are then compared to the empirical path travel time distribution.

In this experiment, five methods are compared: Model(1): a Gaussian mixture models; Model(2): scenario based GCMM model; Model (3): a fixed-marginal-distribution Gaussian copula mixture models; Model (4): a varying-marginal-distribution Gaussian copula mixture model using only the path travel time observations; Model (5): a varying-marginal-
distribution Gaussian copula mixture model with additional travel time observations on each link. To clarify, in the data set, path travel time observations are the vectors of link travel time which are constructed to estimate the dependent structure between the two links. There are additional travel time observations on each link which are not matched (some travelers just traverse one of the two links at the observation time).

First, the link travel time distribution in each path scenario is compared and Figure 4.11 displays the results for the second link: For GMM (Model 1), the estimated link travel time distribution in every path scenario is a Gaussian distribution. For scenario based GCMM model (Model 2), the estimated travel time distribution in every path scenario is the smoothed empirical travel time distribution in that scenario. For this experiment, two scenarios are defined in each link, normal and congestion, and the travel time observations are divided into these two scenarios using a threshold. For general GMM models (Model 3,4,5), the estimated travel time distribution in every path scenario is estimated using weighted kernel smoothing based on all path travel time observations and the weights are the probability that a travel time observation vector belongs to different path scenarios (Similar to GMM, the classification of data to different scenarios are conducted in probability sense instead of strict classification).

Second, the dependent structure is estimated using different models and compared in Figure 4.12: For GMM, the covariance structure of the joint Gaussian distributions in GMM are displayed; For the scenario-based GCMM model, the covariance structure of each path scenario is displayed; For general GCMM methods (Model 3,4,5), the link travel time observations are first transformed into standard normal random variables to estimate copula parameters. Then the covariance structures between the transformed standard normal random variables in each path scenario is displayed.

Third, the travel time on the two links are simulated using the estimated model; their sum are calculated and its distribution is generated; then the distribution of their sum is
compared with the empirical path travel time distribution, which is the distribution of the sum of the travel time on the two links in the empirical dataset. The comparisons are displayed in Figure 4.13.

This figure shows that:

- The estimated path travel time distribution based on GMM shows a smaller mean value and a lighter tail than the empirical distribution;

- The estimated path travel time distribution based on the scenario-based GCMM model shows a larger mean value and a lighter tail.

- The estimated path travel time distribution based on the general GCMM models all show larger mean value and heavier tails.

  - The tail of the estimated path travel time distribution based on Fixed-marginal-distribution GCMM deviates from the true tail probably because the chosen marginal tail may not quite match the true scenarios on the link.

  - The estimated path travel time distribution based on Varying-marginal-distribution GCMM shows good risk properties by estimating the tail more reliably. The tails are heavier than Gaussian models but contain less bias or overestimation.

In all, the research in this chapter generalizes the framework in last chapter to forecast the path travel time distributions. Multiple scenarios of multiple links in a given path are studied, conditioned on different stress levels of the overall traffic conditions. In this process, Gaussian copula mixture models (GCMM) are defined: a scenario based GCMM is introduced first, in which, pseudo path travel time observations are constructed and Lasso methods are used to obtain reliable estimation; more general GCMM are introduced second, in which expectation maximum algorithms are used for parameter estimation to achieve better fitting given limited data and dimensionality. These models help to conduct the path travel time
distribution estimation in real transportation networks; the proposed methodology is also related to the discipline of machine learning and can be applied to other areas.
Figure 4.11: Comparison of the Marginal scenario decomposition for the two links
Figure 4.12: Comparison of the copula structure for the two links
Figure 4.13: Comparison of path travel time distributions estimated through different models: Black: empirical; Red: Model 1; Green: Model 2; Magenta: Model 3; Blue Model 4; Cyan: Model 5
Chapter 5

Travel Time Derivatives: Market Analysis and Pricing

This chapter introduces the concept of a travel time derivative as an alternative approach to congestion pricing, sometimes referred to as value pricing. Travel time derivatives are useful tools for (a) hedging against transportation-related risk, (b) changing travelers’ behavior as a dynamic road toll method, (c) promoting portfolio risk diversification, and (d) providing funding for transportation related projects. In this chapter, potential market participants are analyzed first, and then major products are designed. Alternative models for describing underlying travel time changes are discussed, together with corresponding pricing methods.

5.1 Initiation and necessity analysis

5.1.1 Derivative and weather derivative as a hedging tool

Derivatives are financial instruments whose prices are derived from the value of something else (known as the underlying asset). The major types of derivatives are forwards, futures,
options, and swaps John (2000). Any stochastic changing element that generates changes in cash flow can serve as the underlying asset. Therefore, the underlying element on which a derivative is based can be the price of an asset (e.g., commodities, equities [stock], residential mortgages, commercial real estate, loans, bonds), the value of an index (e.g., interest rates, exchange rates, stock market indices, consumer price index [CPI]), or other items (e.g., temperature, precipitation).

The underlying of derivatives can be further classified as tradable and non-tradable. Most items above can be traded in a market and can be called tradable underlying assets, while a few, such as temperature, precipitation, and travel time, are non-tradable. This difference in the underlying asset will motivate different pricing methods.

In 1999, the Chicago Mercantile Exchange introduced weather futures contracts, the payoffs for which are based on average temperatures at specified locations. According to Stewart (2002), weather derivatives offer an innovative hedging instrument to firms facing the possibility of significant earnings declines or advances because of unpredictable weather patterns. Banks (2002) analyzes various participants, and roles in the weather derivative markets are analyzed. The weather derivative market is formed in need of hedging against weather related risk. Weather derivatives act as an alternative and more flexible way of insuring against weather related risk. As a result, industries which are subject to weather risk participate in the buy/sell side of the market. Speculators who trade purely for profit then come in as an important source of liquidity. The trading and capital activities enable a beneficial mechanism for all related parties.

Weather derivatives provide insurance to farmers/agriculture companies against bad weather and low crop output. The payoff for one party who grows corn and buys a weather derivative contract is as follows: When the weather is good, the insured benefits from abundant corn output; when the weather is bad, the insured receives extra compensation from the derivative to cover losses in corn sales. In this way, the insured hedges risk, and this
risk-protection mechanism is shown in Table 5.1. $G$ means the gain on the derivative, and $P$ means the premium that the farmer pays for the contract.

<table>
<thead>
<tr>
<th>Weather Condition</th>
<th>Corn production payoff</th>
<th>Derivative payoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>$G_{\text{good}}$</td>
<td>$-P$</td>
</tr>
<tr>
<td>Bad</td>
<td>$G_{\text{bad}}$</td>
<td>$G - P$</td>
</tr>
</tbody>
</table>

### 5.1.2 Travel time derivatives is a flexible congestion pricing scheme

There exists similar stochastic nature between temperature changes at a given location and the travel time along a given path. Similar to farmers, travelers should be insured against the economic cost caused by low-quality service. This insurance can be generated by using financial derivatives derived on travel time. Furthermore, because the price of travel time derivatives changes as the predicted traffic conditions change, travel time derivatives can be used to predict future travel time and change travelers’ routing, through which their true time cost caused by traffic delays can be reduced.

First, the payoff of a typical travel time derivative contract is illustrated. When a traveler experience good traffic, nothing needs to be paid except a premium $-P(E(T))$. The payoff is defined as follows:

1. His payoff in the transportation system is good quality of service (QOS) $T_{\text{good}}$
2. His derivative payoff is $-P(E(T))$, as in Figure 5.1.

When traveler experience bad traffic, a gain/compensation $G$ is received while paying the premium $-P(E(T))$, as in Figure 5.2.

1. His payoff in the transportation system is bad QOS $T_{\text{bad}}$
2. His derivative payoff is \( G - P(E(T)) \), where \( G \) is in proportion to the experienced extra travel time from a predefined level \( K \).

Based on the two scenario analysis above, comparisons between conventional congestion pricing methods and travel time derivatives are given below. Traditionally, there are two categories of road toll: static road toll and dynamic road toll. With static road toll, the
traveler pays a fixed premium/toll $P$ to use the road, as in Table 5.2. The charge never changes. With the ordinary dynamic road toll, the traveler pays a fixed amount $P$ when the road is in a condition less favorable for travel (usually the prices are set higher during rush hours) and pays nothing when the road is in a favorable condition for travel, as in Table 5.3. With travel time derivatives (the table shows a call option on travel time), the traveler’s payment $P$ increases continuously when the expected traffic condition changes worse, and the traveler will then benefit from compensation payoff $G$, which is set in proportion to the quality of service received, as in Table 5.4. This comparison shows that the road tolls charged by travel time derivatives are directly linked to expected quality of service; hence, there are potentially more flexible and effective ways of providing insurance against any economic cost caused by the poor quality of traffic service.

Table 5.2: Payoff of traditional road toll, $P$ denotes the toll amount

<table>
<thead>
<tr>
<th>Traffic Condition</th>
<th>Traffic payoff</th>
<th>Derivative payoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good traffic</td>
<td>$T_{good}$</td>
<td>$-P$</td>
</tr>
<tr>
<td>Bad traffic</td>
<td>$T_{bad}$</td>
<td>$-P$</td>
</tr>
</tbody>
</table>

Table 5.3: Payoff of dynamic congestion pricing, $P$ denotes the toll amount

<table>
<thead>
<tr>
<th>Traffic Condition</th>
<th>Traffic payoff</th>
<th>Derivative payoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rush Hour</td>
<td>$T_{good}$</td>
<td>$-P$</td>
</tr>
<tr>
<td>Other time</td>
<td>$T_{bad}$</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.4: Payoff of a travel time derivatives, $P$ denotes its price

<table>
<thead>
<tr>
<th>Traffic Condition</th>
<th>Traffic payoff</th>
<th>Derivative payoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good traffic</td>
<td>$T_{good}$</td>
<td>$-P(E(T))$</td>
</tr>
<tr>
<td>Bad traffic</td>
<td>$T_{bad}$</td>
<td>$\alpha(T_{bad} - K) - P(E(T))$</td>
</tr>
</tbody>
</table>

Beyond reducing economic losses, this flexibility in payment makes the prices of travel time derivatives effective predictors for future travel times. The travelers can forecast the travel time of a path by reading the prices of the corresponding path. This market pricing and risk-return relation will change the travelers’ behavior. More specifically, there are two kinds of payoff relationships with differences in the length of forecasting time windows:

136
1. Some travel time derivatives are based on the future travel time value for a given
time at a given location. As the graph shows, in the short term, the price of travel
time derivatives predicts travel time. Given a certain type of travel time derivative,
the market participants believe the travel time is going to be lower or less volatile if
and only if the price of the derivative contract is higher. When making route choices,
travelers can select the paths with higher prices. This is shown in Figure 5.3.

Figure 5.3: The price of travel time derivatives predicts short term traffic conditions

2. Other travel time derivatives can be derived on the cumulative travel time for a long
time period in future. When considering a long term transportation plan and chooses
between driving more or taking buses more for the coming year, a traveler can check
the price of such derivatives. For certain products, the expected congestion is less
serious in the coming year if and only if the price is higher, and the traveler can then
select more public transportation if the derivative prices are generally low. This is
shown in Figure 5.4.

The payoff analyses above show the travel time derivative not only can provide a payoff
according to the travelers’ future experienced travel time to reduce potential economic cost
Figure 5.4: The price of travel time derivatives predicts long term traffic conditions

due to traffic delays, but it also helps them to predict the future travel time to reduce real time cost due to traffic delays. These features distinguish travel time derivatives from traditional insurance products on travel time (i.e., quality of service of the transportation system). They can be two ways for managing transportation risk. While both accomplish the same objective, their characteristics are not always the same. In general, derivatives are considered a more sophisticated but currently less regulated risk management tool than insurance. Given this situation, some purchasers prefer to use insurance, as most are familiar with this type of risk mitigation tool and may find comfort in its significant regulatory requirements. Insurance does, however, lack some of the flexibility associated with derivative-based solutions. Based on the comparison shown in ?, the following chart summarizes some of the potential differences between insurance and derivative travel time risk management products.
### Table 5.5: Comparison between travel time insurance and travel time derivatives

<table>
<thead>
<tr>
<th>Item</th>
<th>Travel time insurance</th>
<th>Travel time derivatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eligibility to Purchase</td>
<td>No minimum eligibility standards</td>
<td>May meet certain financial sophistication eligibility standards</td>
</tr>
<tr>
<td>Accounting and Tax Treatment</td>
<td>Premiums typically expensed over policy life, recoveries not typically a taxable gain</td>
<td>Treated as an investment, typically valuated using mark to market and recoveries generally create a taxable gain</td>
</tr>
<tr>
<td>Liquidity</td>
<td>Illiquid, effectively a buy and hold instrument</td>
<td>Greater liquidity than insurance, standard contracts are traded on an exchange, not necessarily a buy and hold instrument</td>
</tr>
<tr>
<td>Flexibility</td>
<td>Limited to the purchase of insurance covering measured transportation delays</td>
<td>Can be bought or sold, indexed on a virtually unlimited array of transportation variables</td>
</tr>
<tr>
<td>Regulatory Controls</td>
<td>Significant, including state insurance commissioners</td>
<td>Subject to general regulatory control of financial derivatives. However there may be exemptions similar to what is obtained for weather derivatives</td>
</tr>
<tr>
<td>Risk hedged</td>
<td>Only economic cost caused by traffic delay</td>
<td>Both economic cost and time cost caused by traffic delay (the price of travel time derivatives help to generate routing decisions)</td>
</tr>
</tbody>
</table>

### 5.1.3 Travel time derivatives can diversify risk for financial market

In portfolio theory, the diversification of investments into different asset classes is a recommended practice. For a basket of investments, the lower the correlation between assets, the less the total risk, according to Luenberger (1998). Most traditional asset classes (equity or bond) are derived from the capital of companies and thus are highly correlated in nature. The correlations between travel time and equity/bond are lower than the correlation between different equities/bonds, and the low correlation serves to diversify the portfolio. As investors recognize travel time derivatives as effective risk-reducing elements in their portfolios, they will invest money in the market. The basic risk diversification between the financial system and traffic system is shown in Figure 5.5. Moreover, travel time derivatives can hedge the

![Figure 5.5: Alternative risk transfer between transportation related industry and financial industry](image)

risk in other special markets: Since the weather condition is correlated with the experienced travel time, travel time derivatives will be a good hedging tool for the investors who invest in the weather derivative market. Likewise, CO2 emission levels have been traded in the
market, and since the CO2 emission level is highly correlated to the performance of the traffic system, travel time derivative will be a good tool to hedge against risk for the investors in the CO2 emission market. Compared to insurance on travel time experience, travel time derivatives are different in the following aspects:

1. Travel time derivatives can be traded in the market. One can hold the protection for an arbitrary party. Travel time derivative enables the exchange of risk.

2. Travel time derivatives can be

### 5.2 Potential participants and Market making

In this section, potential participants in travel time derivative markets are introduced, and other market making factors are addressed.

Generally speaking, there are two types of travel time derivatives for the two side of market. Type A: When the specified travel time is high, a leveraged reward is available to the buyer. Type B: When specified travel time is low, a leveraged reward is available to the buyer. Accordingly, participants with different risk profiles will buy different travel time derivatives to hedge their risk: Buyers of Type A are the participants who benefit (are hurt) in good (bad) traffic conditions; buyers of Type B are the participants who are hurt (who benefit) in good (bad) traffic conditions. The potential participants and their roles are summarized in Table 5.6. (QOS means quality of service)

In this market, the market making participants are the essential functional units. Ideally investment banks should take this role, as they do for the weather derivative and energy trading markets. They charge payments from market participants and provide protection according to contracts. With appropriate pricing schemes and suitable adjustments according to demand and supply, the total profit for investment banks will be positive, which motivates
Table 5.6: Different participants

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Hedging Motivation</th>
<th>Type</th>
<th>Other roles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual travelers</td>
<td>Traffic delay, business delay, extra charges</td>
<td>A</td>
<td>investors</td>
</tr>
<tr>
<td>Cargo transportation</td>
<td>Traffic delay and low QOS</td>
<td>A</td>
<td>investors</td>
</tr>
<tr>
<td>Tourism industry</td>
<td>Traffic delay and low QOS</td>
<td>A</td>
<td>investors</td>
</tr>
<tr>
<td>Event organizers</td>
<td>Traffic delay during events</td>
<td>A</td>
<td>investors</td>
</tr>
<tr>
<td>Municipal management</td>
<td>Traffic delay and congestion</td>
<td>A</td>
<td>investors</td>
</tr>
<tr>
<td>Insurance companies</td>
<td>Loss due to vehicle accidents</td>
<td>A</td>
<td>investors</td>
</tr>
<tr>
<td></td>
<td>Repackaging of vehicle insurance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gas company</td>
<td>Low overall gas consumption</td>
<td>B</td>
<td>investors</td>
</tr>
<tr>
<td>Owners of Toll roads</td>
<td>low profit due to good QOS on toll free roads</td>
<td>B</td>
<td>investors</td>
</tr>
<tr>
<td>Vehicle maintenance</td>
<td>Fewer accidents and business loss</td>
<td>B</td>
<td>investors</td>
</tr>
<tr>
<td>Auto companies</td>
<td>Fewer needs for new autos</td>
<td>B</td>
<td>investors</td>
</tr>
<tr>
<td>Public transportation</td>
<td>Less business</td>
<td>B</td>
<td>investors</td>
</tr>
<tr>
<td>Taxi companies</td>
<td>Less business</td>
<td>B</td>
<td>investors</td>
</tr>
<tr>
<td>Alternative transportation train</td>
<td>Less business</td>
<td>B</td>
<td>investors</td>
</tr>
<tr>
<td>Banks</td>
<td>Market Making</td>
<td>A/B</td>
<td>investors</td>
</tr>
<tr>
<td>Traffic detection agencies</td>
<td>Measurement providers</td>
<td>A/B</td>
<td>investors</td>
</tr>
<tr>
<td>GPS companies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portfolio managers</td>
<td>Risk diversification</td>
<td>A/B</td>
<td>investors</td>
</tr>
<tr>
<td></td>
<td>Speculation</td>
<td>A/B</td>
<td>travelers</td>
</tr>
<tr>
<td>Project management</td>
<td>Project financing</td>
<td>A/B</td>
<td>investors</td>
</tr>
<tr>
<td></td>
<td>Speculation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

them to operate the business. This scenario also follows the general mechanism in the current financial derivative markets. On the other hand, the buy/sell sides will seek protection from the market. Their total profit will be negative, which can be interpreted as the cost that they pay to avoid risk due to travel time uncertainties. Other factors that should be considered for market marking include the following:

1. Market microstructure will be crucial in determining the operation of the market. There are numerous links/paths in the transportation networks, and a large number of derivative contracts can be written on the experienced travel time on them. On the other hand, when the market starts initially, the trading activities will be small. Therefore, the market may encounter liquidity issues, where smaller trading amounts
can drive the prices, and so price volatility can be larger. Several measures can be taken, including restricting the number of products in the market, building temporary liquidity reserve, and so forth. A related discussion can be found in MacKenzie & Millo (2003), Dubil (2007) and Wolfers & Zitzewitz (2006).

2. Market scale is important for the sustainability of the derivatives markets. A survey conducted by the U.S. Department of Commerce in 2004 estimates that approximately 30% of the total U.S. GDP is exposed to some degree of weather risk Finnegan (2005). This considerable percentage leads to the necessary liquidity and prosperity of a weather derivative market. For the transportation industry, the percentage needs to be estimated. A larger percentage means more potential market participants.

3. Law enforcement. Regulation of such derivatives is a must. Unlike the weather, the traffic system is subject to people’s behavior. The great leverage provided by travel time derivatives can change people’s routing behavior greatly. This power may be needed to curb the potential for malicious inside trading or market manipulation. To regulate the laws, a series of corresponding policies or laws should be issued.

Based on the settings above, a market can potentially be set up for travel time derivatives. The major products are designed in the following section.

5.3 Product design

In this section, basic travel time derivatives are designed. Their payoffs are different functions of experienced travel time at a future time point or within a future time window at specific locations.

In order to define the products, a standard measurement of travel time must be defined for a given time point.
Definition 5.3.1 A standard measurement of travel time on a specific path at a specific time period on a specific day is the average travel time reported from specific travel time data providers on that path which starts within that time interval on that day.

Definition 5.3.2 A standard measurement plan of travel time on a specific path on a specific day is a set of standard measurements which are collected at pre-defined time of the day. The daily mean of a standard measurement plan is the mean value of such measurements.

All the time points in the following definitions are based on the standard measurement, and each observation is selected by specifying arrival time to the path. Here the average travel time is used, as it can be easier to obtain through loops, and data from loops provide abundant periodical measurements. When the coverage ratio of GPS devices is high enough, individual travel time measurements from GPS devices can be used as the standard measurement.

5.3.1 Standard country travel time index and equivalent return rate

First, national travel time index is designed, which is a weighted average of the latest travel time in downtown areas of major cities in the U.S. The national travel time index is an objective reference for trading and a good symbol for the transportation industry. The definition is given below, and local travel time indexes can be designed in a similar fashion.

Definition 5.3.3 Country travel time index

\[ T_{us} = \sum T_i \alpha_i \]

where \( T_i \) is the travel time in selected places in the country.
An illustrating index can be constructed by the weighted average of the real time travel time in downtown New York (a section of Fifth Avenue), downtown Chicago (a section of Maxwell Street), downtown Los Angeles (a section of Grand Avenue), and downtown Houston (a section of Main Street). This index can be viewed as an average traffic index on the quality of service of the urban transportation system in United States, which shows the national service level of urban traffic systems. The return of this index, volatility, and its sharp ratio are then used as references when pricing travel time derivatives. Sharp ratio is the ratio between access return and volatility of this derivative, and the access return is the extra return of this derivative compared to the risk-free bond.

5.3.2 Design of derivative product on travel time

Type 1 Basic options for a specific link at a given future time point

The simple derivative for the experienced travel time at a future time is defined as follows, and an example is given afterwards.

**Definition 5.3.4** Call option on a certain travel time. Consider the link l a specific time instant t in the future. If the travel time shown by the standard measurement at t (denoted as $T$) is higher than a given $K$, then there is a payment $\alpha(T - K)$ to the option buyer; if lower, there is no payment. $\alpha$ is the leverage coefficient.

**Definition 5.3.5** Put option on a certain travel time. Consider the link l a specific time instant t in the future. If the travel time shown by the standard measurement at t (denoted as $T$) is lower than a given $K$, then a payment $\alpha(K - T)$ is available to the option buyer; if higher, there is no payment. $\alpha$ is the leverage coefficient.

Consider Broadway in New York City from 20th to 60th Streets, if the travel time shown by the standard measurement on it entering at 10:00 a.m. on January 1, 2011, is equal to
70 minutes and the threshold value is set as 60, then there is a leveraged cash back to the buyer $10 \times (70 - 60)$; if the experienced travel time is lower than 60 minutes, the buyer gets nothing.

**Type 2 Futures on congestion-days**

The future written on the high congestion days (HCD) and low congestion days (LCD) are then designed. As a basic concept, the definitions of HCD and LCD are given below:

**Definition 5.3.6** *High Congestion Days (HCD) and Low Congestion Days (LCD) in discrete time settings*

Let $T_i$ denotes the mean of a standard measurement plan on day $d_i$ and $C$ as a specified reference value. The high congestion-days, $HCD_i$, and the lower congestion-days, $LCD_i$, on that day are defined as $HCD_i = \max(T_i - C, 0)$ and $LCD_i = \max(C - T_i, 0)$ respectively. In other words, $HCD$ is the extra amount of travel time spent on that day compared to the reference value $C$, and $LCD_i$ is the amount of travel time savings compared to the reference value $C$.

Then the HCD for a given time period $[t_1, t_2]$ is defined as the sum of the HCD in all the days in that period, given a fixed number of measurements.

$$HCD(t_1, t_2) = \sum_{i=1}^{n} HCD_i 1_{d_i \in [t_1, t_2]}$$

the LCD for a given time period $[t_1, t_2]$ is defined as the sum of the LCD in all the days in that period, given a fixed number of measurements.

$$LCD(t_1, t_2) = \sum_{i=1}^{n} LCD_i 1_{d_i \in [t_1, t_2]}$$

, as the HCD/LCD for the time period.
In a continuous setting, the payoff functions should be defined as follows:

**Definition 5.3.7** High Congestion Days (HCD) and Low Congestion Days (LCD) in continuous time

Given a threshold $T$, the HCD for a given time period $[t_1, t_2]$ is defined as

$$HCD(t_1, t_2) = \int_{t_1}^{t_2} \max(T_t - T, 0) dt$$

the LCD for a given time period $[t_1, t_2]$ is defined as

$$LCD(t_1, t_2) = \int_{t_1}^{t_2} \max(T - T_t, 0) dt$$

This pair of products shows the cumulative performance of the path compared to some average reference. Its price will show people’s expectation of the quality of service on the path; hence, its price can predict the long term traffic status on the path.

**Type 3 Congestion-days options**

Congestion day options are the options based on the average performance of a path in a future time window. The definitions are given first, followed by an example.

**Definition 5.3.8** Call options on high congestion days. Denote $K$ as the strike value:

The payoff of a HCD call is

$$V = \alpha \max(H_n - K, 0)$$

The payoff of a LCD Call is

$$V = \alpha \max(L_n - K, 0)$$
**Definition 5.3.9** Put options on high congestion days. Denote $K$ as the strike value:

The payoff of a HCD put is

$$X = \alpha \max(K - H_n, 0)$$

The payoff of a LCD put is

$$X = \alpha \max(K - L_n, 0)$$

Consider Broadway in New York City from 20th to 60th Streets. If the mean travel time on it is above 60 minutes, then a surplus $T - 60$ is recorded as a congestion day; otherwise 0 surplus is recorded. Then all these surplus values are added together for one year with 365 days. If the sum $S$ equals 2000 and so is larger than $K = 1500$, then there is a leveraged cash back $10 \cdot (S - K)$ where 10 is the leverage ratio; if not, the buyer receives nothing. This is an example of HCD call option. This pair of products leverages the buyer’s gain according to the long term traffic status in future. Compared to the futures, the options provide further leverage, and the buyers can get more return/loss if the traffic conditions changes. In buying such products, a traveler will change travel patterns accordingly. In this sense, options on travel time are effective in changing a traveler’s behavior. **Type 4 Futures on cumulative travel time**

This product is the futures written on the cumulative travel time in a future time period. The cumulative travel time index is defined first.

**Definition 5.3.10** Cumulative travel time index (CTT) in discrete time settings.

The CTT index over a time interval $[t_1, t_2]$ is defined as the sum of the daily standard measurement plan in a given time period.

$$CTT(t_1, t_2) = \sum_{i=1}^{n} T_i I_{t_i \in [t_1, t_2]}$$
Definition 5.3.11  Cumulative travel time index in a future time period (CTT) in continuous
time settings.

The CTT index over a time window \([t_1, t_2]\) is defined as the integration of travel time in
that time window.

\[
CTT(t_1, t_2) = \int_{t_1}^{t_2} T_t dt
\]

The payoff of the futures on CTT is in direct proportion to the travel time that a traveler
experiences over a time period. It is an alternative measure of the long term quality of service
to HCD/LCDs. **Type 5 Options on cumulative travel time** These products are the
options written on the CTT index. Their payoff functions are given as follows:

**Definition 5.3.12**  Call options on cumulative travel time. Denote \(K\) as the strike value:
The payoff of a HCD call is

\[
V = \alpha \max(CTT - K, 0)
\]

**Definition 5.3.13**  Put options on high congestion days. Denote \(K\) as the strike value: The
payoff of a HCD put is

\[
X = \alpha \max(K - CTT, 0)
\]

Again, the options on CTT provide greater leverage than other forms of road tolls; therefore,
they can potentially change a traveler’s behavior more effectively.

5.4  Pricing the derivatives on travel time

To price the travel time derivatives, the underlying model is selected as a mean reverting
process with an unknown driven process first. As there are different advantages for different
models and the data in different links may show different characteristics, a family of alternative models is given below. The best driven process is selected according to statistical tests. Finally, the pricing methods are discussed.

5.4.1 Alternative processes for the travel time

Mean reverting processes are the stochastic processes for which high and low values are temporary and values tend to move back to their average over time. Mean reverting models are frequently used in the financial literature, especially when calculating the price for interest rates derivatives and weather derivatives\cite{HullWhite:1990}.

Let \((\Omega, F, F_{t\geq 0}, P)\) be a complete filtration probability space. A random variable is a mapping \(X : \Omega \rightarrow \mathbb{R}^d\), if it is \(F\)-measurable, whereas a family of random variables depending on time \(t\), \(X_{t\geq 0}\) is said to be a stochastic process. A process \(X_t\) is \(F\)-adapted if every \(X_t\) is measurable with respect to the \(\sigma\)-algebra \(F\). Then the travel time process can be modeled by different mean reverting processes:

A mean-reverting process driven by Brownian motion

\(T_t\) can be modeled as mean reverting process driven by Brownian motion, as follows:

\[
    dT_t = db_t + a_t(b_t - T_t)dt + \sigma_t dB_t
\]

(5.1)

\(\sigma\) is the volatility of the travel time process and \(B_t\) is Brownian motion, which has a more complex structure than the usual independent and identical distributed (i.i.d.) white noise series. The solution is

\[
    T_t = b_t + (T_0 - b_0)e^{-\int_0^t a_s ds} + e^{-\int_0^t a_s ds} \int_0^t e^{\int_s^t a_u ds} \sigma_u dB_u
\]
When the coefficients are constant, the solution is simplified to the following form:

\[ T_t = b + (T_0 - b)e^{-at} + \int_0^t e^{-a(t-u)}\sigma dB_u \]

Then the travel time process model can be fitted to the conditional probability surface of the empirical conditional distribution. For any time instant \( t \), the distribution is Gaussian with the following mean and variance:

\[ \mu_t = T_0e^{-\int_0^t a_s ds} + b_t - b_0e^{-\int_0^t a_s ds} \]

and

\[ \sigma_t^2 = e^{-2\int_0^t a_s ds} \int_0^t e^{2\int_0^u a_s ds} \sigma_u^2 du^2 \]

By equaling the theoretical mean and variance to those of the empirical conditional distribution with the same time lag, the parameters can be estimated. Since \( dB_t \) is normally distributed with variance \( t \), and this term is independent for different time intervals, this model is tightly related to the usual time series model with i.i.d normal noise.

A mean-reverting process driven by an integrated process and the formal derivatives

\( T_t \) can be modeled as a mean reverting process driven by integrated processes, as follows:

\[ dT_t = db_t + a_t(b_t - T_t)dt + \sigma_t dY_t \quad (5.2) \]
Here, the integrated process of $X_t$ is defined as $Y_t = \int_0^t X_u du$ or in other form $dY_t = X_t dt$. $X_t$ is a diffusion process in the form of $X_t = a_t dt + b_t dB_t$. This model can actually approximate the continuous counterpart for ARIMA(1,1,0) process.

5.4.2 Model the travel time processes

In this subsection, empirical data are used to select the model to describe the travel time processes, and corresponding parameters are estimated. The driving process is identified as ARMA models, and their corresponding continuous versions are given. The model is then used to price derivatives in later sessions. First, the mean reverting model is re-parameterized as follows:

$$T_t = r_t + s_t + Y_t, \ t = 0, 1, 2, \ldots$$  \hspace{1cm} (5.3)

where $r_t$ is the trend part, $s_t$ is the season part, and $Y_t$ is the driving process that shapes the noise term. Define $c_t = r_t + s_t$ as the sum of the trend and seasonal parts. The terms are explained separately as follows:

1. The trend part is a linear function over time

$$r_t = a + bt$$

2. Seasonal parts are as follows:

$$s_t = b_t + w_t$$
(a) The daily part is:

\[ b_t = k + \sum_{i=1}^{I} a_i (\sin 2i\pi (t - f_i)/T_d) + \sum_{i=1}^{J} b_i (\cos 2i\pi (t - g_i)/T_d) \]

(b) The weekly part is:

\[ w_t = k + \sum_{i=1}^{I} a_i (\sin 2i\pi (t - f_i)/T_w) + \sum_{i=1}^{J} b_i (\cos 2i\pi (t - g_i)/T_w) \]

(c) Another alternative is the 10-parameter model given in Schrader. & Kornhauser (2003):

\[ V(t) = \mu + \sum_{i=1}^{3} k_i \phi(t, \mu_i, \sigma_i) \]

However, the trigonometric functions are selected as the basis function, as they are orthogonal and suitable to describe the periodical pattern of travel time.

3. \( Y_t \) is the driving process. It is a stochastic process that may contain a mean reverting part and a simpler stochastic part, which is usually a Brownian motion. The type of process is determined based on empirical data.

The data set is the travel time data set from California PEMS. Data were collected in November 2010, on the 80-E path between 80-E/Lincoln Road and 80-E/Old Davis Road, as presented in the following graph. The travel time data are from loops for every 5 minutes. The calibration process is as follows:

1. Regression estimate is generated for the trend over the year in the table, and the estimate is not significant except as a stable mean function \( a = 41.19 \). This part is shown in the Figure ?.

2. The seasonal model is studied with a regression method. The parameters for the two seasonal models are obtained as follows with \( T \) as 2064 for the weekly pattern and 288 for the daily pattern. The seasonal parts are shown in Figure ?.
3. After removing the trend and seasonal part, the residual is identified as the following ARIMA(1,1,1) model. A Box-Ljung test suggests p-value = 0.3419. This model describes the time series well.

\[(T_t - T_{t-1}) = a(T_{t-1} - T_{t-2}) + b\sigma e_t + \sigma e_{t-1}\]

### Table 5.7: Seasonal effects in the 80E data

<table>
<thead>
<tr>
<th>Trend</th>
<th>(k)</th>
<th>(a_1)</th>
<th>(a_2)</th>
<th>(a_3)</th>
<th>(b_1)</th>
<th>(b_2)</th>
<th>(b_3)</th>
<th>(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly</td>
<td>-0.293</td>
<td>1.241</td>
<td>1.362</td>
<td>0.387</td>
<td>-0.268</td>
<td>0.377</td>
<td>0.631</td>
<td>2064</td>
</tr>
<tr>
<td>Daily</td>
<td>-0.206</td>
<td>-0.664</td>
<td>0.728</td>
<td>0.399</td>
<td>-2.236</td>
<td>0.0749</td>
<td>-0.141</td>
<td>288</td>
</tr>
</tbody>
</table>

### Table 5.8: ARIMA modeling of the residuals

<table>
<thead>
<tr>
<th>Trend</th>
<th>(a)</th>
<th>(b)</th>
<th>(\sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly</td>
<td>0.1197</td>
<td>0.2612</td>
<td>0.265</td>
</tr>
<tr>
<td>value</td>
<td>0.0281</td>
<td>0.0272</td>
<td>0.265</td>
</tr>
<tr>
<td>s.e.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The model is described as the ARIMA(1,1,1) model by comparing the test results.
Figure 5.7: Remove the trend, weekly, daily part of 80E data series

Table 5.9: Model selection for 80E data

<table>
<thead>
<tr>
<th>Model</th>
<th>Log Liklihood</th>
<th>AIC</th>
<th>df</th>
<th>p-value of Box-Ljung test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arma(1,0)</td>
<td>-750.16</td>
<td>1506.32</td>
<td>9</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>ARIMA(1,1,0)</td>
<td></td>
<td></td>
<td>9</td>
<td>0.001240</td>
</tr>
<tr>
<td>ARIMA(1,1,1)</td>
<td>-750.16</td>
<td>1506.32</td>
<td>9</td>
<td>0.3419</td>
</tr>
<tr>
<td>Arma(1,1)-Garch(1,1)</td>
<td>377.1066</td>
<td>874.21</td>
<td>9</td>
<td>8.09e-05</td>
</tr>
<tr>
<td>Arfima(1,0.49,1)</td>
<td></td>
<td></td>
<td>9</td>
<td>&lt; 2.2e-16</td>
</tr>
</tbody>
</table>
4. The estimation is demonstrated for a path above. For different paths, the travel time process may appear to be fit to different models. Pricing derivatives can be considered on this different possible form of $Y_t$. The simple option will be to similar to pricing an Asian option on the primary process $Y_t$.

(a) If $Y_t$ is an AR(1) model (ARIMA(1,0,0)), the underlying process will be a simple mean reverting process as follows:

\[
T_t = c_t + Y_t \\
Y_t = -aY_{t-1} + e_t
\]
Its continuous version is as follows

\[
T_t = c_t + Y_t \\
dY_t = -(1 + a)Y_t dt + \sigma dB(t)
\]

where \( e_t \sim N(0, \sigma) \)

(b) If \( Y_t \) is an ARMA(1,1) model (ARIMA(1,0,1))

\[
T_t = c_t + Y_t \\
Y_t = -kY_{t-1} dt + e_t + be_{t-1}
\]

Its continuous version is as follows

\[
T_t = c_t + Y_t \\
dY_t = -(1 + a)Y_t dt + \sigma e_t^{(2)} dt + (b + 1)\sigma e_t^{(1)} dt
\]

where \( e_t \sim N(0, \sigma) \) and the formal derivatives are defined for \( B_t \) as

\[
 dB_t = e_t^{(1)} dt
\]

and

\[
 de_t^{(1)} = e_t^{(2)} dt
\]
(c) If \( Y_t \) is an integrated processes defined above for ARIMA(1,1,0)

\[
T_t = c_t + Y_t \\
(X_t - X_{t-1}) = -(X_{t-2} - X_{t-3}) + e_t
\]

Its continuous version is as follows

\[
T_t = c_t + Y_t \\
Y_t = \int_0^t X_u du \\
dX_t = -(1 + a)X_t dt + \sigma dB(t)
\]

where \( e_t \sim N(0, \sigma) \)

(d) If \( Y_t \) is an integrated processes defined above for ARIMA(1,1,1)

\[
T_t = c_t + Y_t \\
(Y_t - Y_{t-1}) = -(Y_{t-1} - Y_{t-2}) + e_t + be_{t-1}
\]

Its continuous version is as follows

\[
dX_t = -(1 + a)X_{t-1} dt + \sigma e_t^{(1)} dt + (b + 1) \sigma e_t^{(2)} dt \\
T_t = c_t + \int_0^t X_u du
\]

where \( e_t \sim N(0, \sigma) \) and formal derivatives are defined in the same way.
In the following sections, the possible pricing schemes will be shown for these processes. For ARIMA(1,0,0), ARIMA(1,1,0) there are some explicit solution methods. For the process where there are formal derivatives, i.e., ARIMA(1,0,1) and ARIMA(1,1,1), Monte Carlo methods can be used after the pricing principles are introduced, as the formal derivatives are not rigorously defined in the stochastic calculus discipline.

5.4.3 Hedge using a portfolio of derivatives

Since travel time is a non-tradable asset, pricing a travel time derivative is similar to pricing interest rate derivatives and weather derivatives. In this section, no-arbitrage pricing is used to price a simple travel time option.

When the underlying process is a mean reverting process, i.e., can be modeled by the diffusion, then consider two travel time derivatives with different payoff functions: \( F \) and \( G \). A portfolio is defined \( P = \alpha F + \beta G \) of \( F \) and \( G \) with \( \alpha + \beta = 1 \). Suppose the underlying process of the travel time changes is \( dX = db_t + A(t, X_t)dt + B(t, X_t)dW_t \) where \( b_t \) is a function of time. If the portfolio is risk neutral, then its value should increase at the same rate as a risk-free rate. The deductions are as follows

\[
\alpha \left( \frac{\partial F}{\partial t} + \frac{\partial F}{\partial X} dx + \frac{1}{2} \frac{\partial F^2}{\partial X^2} d\langle X \rangle \right) + \beta \left( \frac{\partial G}{\partial t} + \frac{\partial G}{\partial X} dx + \frac{1}{2} \frac{\partial G^2}{\partial X^2} d\langle X \rangle \right)
\]

\[
\alpha \left( \frac{\partial F}{\partial t} + \frac{\partial F}{\partial X} A(t, X_t) + \frac{1}{2} \frac{\partial F^2}{\partial X^2} B^2(t, X_t) \right) + \beta \left( \frac{\partial G}{\partial t} + \frac{\partial G}{\partial X} A(t, X_t) + \frac{1}{2} \frac{\partial G^2}{\partial X^2} B^2(t, X_t) \right) dt +
\]

\[
\left\{ \alpha \frac{\partial F}{\partial X} + \beta \frac{\partial G}{\partial X} \right\} db_t + \left\{ \alpha \frac{\partial F}{\partial X} B(t, X_t) + \beta \frac{\partial G}{\partial X} B(t, X_t) \right\} dW_t
\]
To make the portfolio riskless, the volatility should be zero:

\[ \alpha \frac{\partial F}{\partial X} B(t, X_t) + \beta \frac{\partial G}{\partial X} B(t, X_t) = 0 \]

and we have

\[ \frac{\alpha}{\beta} = -\frac{\frac{\partial G}{\partial X}}{\frac{\partial F}{\partial X}} \]

This approach eliminates \( ds \) term. Furthermore, the increasing rate should be interest rate \( r \):

\[ \alpha \left( \frac{\partial F}{\partial t} + \frac{\partial F}{\partial X} A(t, X_t) + \frac{1}{2} \frac{\partial^2 F}{\partial X^2} B^2(t, X_t) \right) + \beta \left( \frac{\partial G}{\partial t} dt + \frac{\partial G}{\partial X} A(t, X_t) + \frac{1}{2} \frac{\partial^2 G}{\partial X^2} B^2(t, X_t) \right) = r(F + \beta G) \]

and we have

\[ \frac{\frac{\partial F}{\partial t} + \frac{\partial F}{\partial X} A(t, X_t) + \frac{1}{2} \frac{\partial^2 F}{\partial X^2} B^2(t, X_t) - rF}{\frac{\partial F}{\partial X}} = \frac{\frac{\partial G}{\partial t} + \frac{\partial G}{\partial X} A(t, X_t) + \frac{1}{2} \frac{\partial^2 G}{\partial X^2} B^2(t, X_t) - rG}{\frac{\partial G}{\partial X}} \]

By dividing the volatility part on both side we have:

\[ \frac{\frac{\partial F}{\partial t} + \frac{\partial F}{\partial X} A(t, X_t) + \frac{1}{2} \frac{\partial^2 F}{\partial X^2} B^2(t, X_t) - rF}{B(t, X_t) \frac{\partial F}{\partial X}} = \frac{\frac{\partial G}{\partial t} + \frac{\partial G}{\partial X} A(t, X_t) + \frac{1}{2} \frac{\partial^2 G}{\partial X^2} B^2(t, X_t) - rG}{B(t, X_t) \frac{\partial G}{\partial X}} = \lambda \]

where we define \( \lambda \) as the market value of risk.

For an individual travel time derivative with payoff \( F \), its price should satisfy the following PDE:

\[ \frac{\partial F}{\partial t} + \frac{\partial F}{\partial X} (A(t, X_t) - \lambda B(t, X_t)) + \frac{1}{2} \frac{\partial^2 F}{\partial X^2} B^2(t, X_t) = rF \]

where \( 0 \leq t < T \) and \( x \geq 0 \) with boundary conditions:
\[ v(t, x) = e^{-r(T-t)}(x - K)^+ \]
\[ \lim_{x \to -\infty} v(t, x) = 0, 0 \leq t < T, x \geq 0 \]
\[ v(T, x) = (x - K)^+, x \geq 0 \]

1. For a call option with payoff function \( F = \max\{0, X - K\} \), the terminal conditions are the payoff functions for the option. With appropriate boundary and terminal conditions, this PDE can be solved and yield a proper price for the options.

2. If \( Y_t \) is just a mean reverting process we can directly consider \( dT_t = dc_t + k(T_t - c_t)dt + \sigma dB(t) \)

\[ A(t, T_t) = \frac{dc_t}{dt} + k(T_t - c_t) \quad B(t, T_t) = \sigma \]

and the PDE can be converted correspondingly. Here, the trend and seasonal parts are considered in the underlying process.

3. If \( Y_t = \int_0^t X_t dt \) and \( X_t = A(t, X_t)dt + B(t, X_t)dW_t \). We can price the claims on \( Y_t \) as the claims on \( X_t \). So the vanilla call options on \( Y_t \) will be viewed as Asian option on \( X_t \), as the pay off is

\[ \max(T_T - K, 0) = \max(c_T + \int_0^T X_t dt - K, 0) \]

The above pricing PDE will be applied to \( X_t \). The trend and seasonal parts are considered in the payoff function \( F \).

4. The market value of risk \( \lambda \) can be estimated through the sharp ratio \( \left( \frac{\mu - r}{\sigma} \right) \) of the country travel time index. By doing so, the average risk level associated with the travel time derivatives is set to be as the same as the country travel time index.
5.4.4 Risk neutral pricing in an incomplete market

To get a general pricing expression for the more complex derivatives, the risk neutral pricing in the incomplete market is used, deploying the method in Benth & Benth (2007).

The risk neutral representations

The risk-neutral probability $Q$. Since travel time is not a tradable or storable asset, the futures contracts cannot be hedged, and the market is therefore incomplete. A risk-neutral probability is by definition a probability measure $Q \sim P$ such that all tradable assets in the market are martingales after discounting. Thus, all equivalent probabilities $Q$ will become risk-neutral probabilities. We specify a sub-family of probability measures $Q$ using the Girsanov transform: assume $\Omega_t$ is a real-valued measurable and bounded function. The stochastic process:

$$ Z^\omega(t) = \exp \left( \int \frac{\omega(s)}{\sigma(s)} dW_s - \frac{1}{2} \int \frac{\omega^2(s)}{\sigma^2(s)} ds \right) $$

is the density process of the probability measure $Q$. Under $Q$, the process

$$ dB_t = dW_t - \frac{\omega(t)}{\sigma(t)} dt $$

is a Brownian motion and the risk-neutral process for the underlying mean reverting model is given as

$$ dT_t = dc_t + \omega(t) dt + a_t(c_t - T_t) dt + \sigma_t dW_t $$

and the process is then
\[ T_t = c_t + (T_0 - c_0)e^{-\int_0^t a_s ds} + e^{-\int_0^t a_s ds} \int_0^t e^{\int_0^s a_r dr} \omega_u du + e^{-\int_0^t a_s ds} \int_0^t e^{\int_0^s a_r dr} \sigma_u dW_u \]

When the coefficients are constant, the general solution reduces to the following simpler form:

\[ T_t = c + (T_0 - c)e^{-at} + \int_0^t e^{-a(t-u)} \omega_u du + \int_0^t e^{-a(t-u)} \sigma dW_u \]

Then the prices of the products in continuous time are as follows:

**HCD futures**

\[ F_{HCD}(t_1, t_2, T, t) = E_Q \left\{ \int_{t_1}^{t_2} \max(T_t - T, 0) dt \middle| F_t \right\} \]

**LCD futures**

\[ F_{LCD}(t_1, t_2, T, t) = E_Q \left\{ \int_{t_1}^{t_2} \max(T - T_t, 0) dt \middle| F_t \right\} \]

Assuming a constant interest rate, HCD call options can be calculated as

\[ C_{HCD}(t_1, t_2, T, t) = \exp(-rt)E_Q \left\{ \max(F_{HCD}(t_1, t_2, T, t) - K) \middle| F_t \right\} \]

and

\[ C_{LCD}(t_1, t_2, T, t) = \exp(-rt)E_Q \left\{ \max(F_{LCD}(t_1, t_2, T, t) - K) \middle| F_t \right\} \]

The dynamics of these options can be complicated. The option price can be found by simulating the travel time process under the specific measure \( Q \).
CTT futures

\[ F_{CTT}(t_1, t_2, T, t) = E_Q \left\{ \int_{t_1}^{t_2} T_1, 0 dt | F_t \right\} \]

Assuming a constant interest rate, CTT call options can be calculated as

\[ C_{CTT}(t_1, t_2, T, t) = \exp(-rt)E_Q \{ \max(F_{CTT}(t_1, t_2, T, t) - K) | F_t \} \]

If the process is driven by a Brownian motion, then ordinary martingale measure can be used to price the derivatives by specifying the market value of risk process \( \Omega_t = \exp\left(-\frac{1}{2}\theta^2_t - \theta_t B_t\right) \) such that \( B_t = B_t - \theta_t \) is a Brownian motion. Here \( \theta(t) \) is a real-valued measurable and bounded function.

**Explicit solutions for specific processes**

Next, some explicit solution methods are specified via PDE for the processes defined without formal derivatives, i.e., those for ARMA(1,0) and ARIMA(1,1,0).

For vanilla options on the integrated process \( Y_t \), we further get the following PDE when specifying a risk neutral measure \( P \)

**Theorem 5.4.1**

\[ v_t(t, x, y) + A(t, X_t)v_x(t, x, y) + xv_y(t, x, y) + \frac{1}{2}B(t, X_t)^2v_{xx}(t, x, y) = rv(t, x, y) \]

\( 0 \leq t < T, x \geq 0, y \in R \)

with the following boundary conditions:

\[ v(t, 0, y) = e^{-r(T-t)}(y - K)^+ \]
\[
\lim_{y \to -\infty} v(t, x, y) = 0, 0 \leq t < T, x \geq 0
\]
\[
v(T, x, y) = (y - K)^+, x \geq 0, y \in \mathbb{R}
\]

proof: Similar to the following more general theorem.

**Theorem 5.4.2** For the Asian type option on integrated process \( Y_t \) we can further expand state as follows:

\[
v_t(t, x, y, z) + A(t, X_t) V(t, x, y, z) + xv_y(t, x, y, z) + yv_z(t, x, y, z) + \frac{1}{2} B(t, X_t)^2 v_{xx} = rv(t, x, y, z)
\]

\[0 \leq t < T, x \geq 0, y \in \mathbb{R}, z \in \mathbb{R}\]

**boundary conditions**

\[
v(t, 0, 0, z) = e^{-r(T-t)}(z - K)^+
\]
\[
\lim_{z \to -\infty} v(t, x, y, z) = 0, 0 \leq t < T, x \geq 0, y \in \mathbb{R}
\]
\[
v(T, x, y, z) = (z - K)^+, x \geq 0, y \in \mathbb{R}, z \in \mathbb{R}
\]

Proof: Consider the claim \((Z_t - K)^+\), under the risk neutral measure \( Q \), we have:

\[
dZ_t = Y_t dt \\
dY_t = X_t dt \\
dX_t = A(t, X_t) dt + B(t, X_t) dW_t
\]

We consider that this group of differential equation defines a three dimensional Markovian
The value of claim will be $de^{-rt}v(t, x, y, z)$ By Ito’s lemma

\[
dv(t, x, y, z) = v_t dt + v_x dx + \frac{1}{2} v_{xx} d < x > + v_y dy + v_z dz
\]

\[
= e^{-rt}(-rv + v_t + \frac{1}{2} v_{xx} B(t, X_t)^2 + v_x A^*(t, X_t) + xV_y + yv_z) dt + v_x B(t, X_t)e^{-rt} dW_t
\]

This process should be a martingale so we have

\[-rv + v_t + \frac{1}{2} v_{xx} B(t, X_t)^2 + v_x A^*(t, X_t) + xV_y + yv_z = 0\]

That is

\[v_t + \frac{1}{2} v_{xx} B(t, X_t)^2 + v_x A^*(t, X_t) + xV_y + yv_z = rv\]

Here $A^*(t, X_t)$ is the drift of $X$ under the risk neutral measure. As $X$ is not tradable, it is not necessarily $r$, the primary drift $A(t, X_t)$ can be used for certain pricing measure $Q$

Since the underlying is not tradable, there is no unique risk neutral measure. In general, the option pricing in an incomplete market relies in the positions in the underlying and position in the option itself Dennis (n.d.). Specifying the drift as $A(t, X_t)$ in the P.D.E means the actually physical drift is the view of the trader about the growth of the process. Since travel time process is almost independent from the interest rate process, so this assumption can be well justified. Q.E.D.

Finally, for the process where we have formal derivatives ARMA(1,1) and ARIMA(1,1,1), Monte Carlo methods can be used within the risk neutral framework.
Chapter 6

Conclusions and Future Work

This thesis has examined the methodology of estimating road travel time distribution using copula methods and introduced advanced value pricing schemes using financial derivatives based on travel time. The study highlights the dependent structure of travel times between different links, and the major contributions are achieved by recognizing the linkage between the transportation and the financial systems.

The major contributions of this thesis are listed below:

- First, this thesis proposes a new framework for generating the conditional density of travel time for a focused link set using copula methods. Travel time on a link is modeled as a stationary random variable, given a fixed traffic scenario on that link. The dependence between such random variables corresponding to different links is modeled with a suitable copula. If only one scenario is considered, the conditional density of observed travel time on the focused link is then calculated based on this copula, using the observed travel time values as inputs. Transportation risk measures are calculated based on the estimated conditional travel time density of the focused link, which helps to identify optimal routing choices for travelers.
• Second, this thesis develops Gaussian copula mixture models (GCMMs) to estimate the joint distribution of a group of random variables that describe the travel time of links on a path. For a long path, the dependence between its constructing links and the link travel time performance differs as the overall traffic condition along the path changes. Based on this logic, this research models the dependence with different Gaussian copulas for all possible scenarios of the path travel conditions. In more detail, in a given scenario, a Gaussian copula is used to describe the dependent structure between the constructing links, and the link travel time distributions are also modeled as stationary random variables. Conditional path travel time distribution is then generated by summing up the link travel times with this copula dependence. The overall path travel time distribution is a weighted sum of such conditional path travel time distributions. Therefore, the mathematical model is a GCMM. This model can be estimated using a two-step maximum likelihood method if the path traffic scenarios can be well defined. More generally, the expectation maximum method can be used to estimate the general GCMM model. Such a model approximates the complex dependent structure of travel time well and takes full use of the available data.

• Third, this thesis designs the financial derivative based on travel time. This new financial derivative is based on road travel times in the transportation systems. It is a new value pricing scheme based not only on the level of congestion but also on the uncertainties surrounding the congestion. The potential market is analyzed, and major products are designed. The mean reverting processes, driven by diffusion and integrated diffusion processes, are used to model the travel time as the underlying asset. Derivative prices are calculated using no-arbitrage models and risk-neutral pricing models under incomplete market conditions.

In the following part, this thesis proposes recommendations for future research on this topic, organized according to the study’s three core aspects. These of the aspects are drawn to better understand the behavior of the transportation system and the financial system.
• First, travel time distribution estimation is difficult to determine given the strong dependence between travel times. By using copulas and scenario analysis, this complexity can be addressed with quantified errors. Many variations of copulas can be used to improve the goodness-of-fit in further research.

• Second, for the estimation of the GCMM models, algorithm design can be further conducted, and corresponding convergence properties should be considered. A more clearly defined example is the theatrical properties of the Varying-margin-optimization algorithm. After assembling the two-step maximum estimator for copula with the expectation-maximum algorithm, the whole procedure may yield interesting theoretical properties.

• Third, factor models can be introduced to get simpler approximations. From the view of each link, the path scenario can be interpreted as a systematic factor. In the current model, the systematic factor determines the copula and the link travel time distributions. Conditional path travel time distribution is still generated by simulating the estimated copula. With a suitable factor model between link travel time and the systematic factor, loadings can be estimated, and path travel time distribution can be calculated more easily. Such simplification may act as an industry standard for travel time risk management.

• Fourth, other statistical models can be used to describe the complex dependent structure between travel times. A good alternative is kernel regression Fan et al. (1996). The general properties of the kernel-based method have been studied extensively in statistical literature, while specific properties should be studied for transportation applications.

• Fifth, more complex stochastic models can be used to describe the travel time process. For example, stochastic volatility models can be introduced to describe the complex variance structure of travel time over time. The price of travel time derivatives can be
determined by using appropriate machine learning schemes or quoting market implied volatility of the underlying travel time processes Kimmel et al. (2007). Moreover, various methods can be used to price travel time derivatives, such as partial hedging and indifference pricing methods. These methods may yield potentially more accurate prices and improve the performance of both the transportation and the financial systems.

- Sixth, due-equilibrium of the financial system and the transportation system can be studied after introducing travel time derivatives. Fundamentally, value pricing schemes, including travel time derivatives, link the performance of the traffic system to that of the financial system. The more complex the scheme, the tighter such linkage. Therefore, maintaining equilibrium in both the transportation and the financial systems after introducing the travel time derivatives is a challenging and promising research topic.
Bibliography


Steele, J. (2001), Stochastic calculus and financial applications, Springer Verlag.


