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. Kornhauser

September 2014
Abstract

Given the complexity of transportation systems, generating optimal routing decisions is a critical issue. This thesis focuses on how routing decisions can be computed by considering the distribution of travel time and associated risks. More specifically, the routing decision process is modeled in a way that explicitly considers the dependence between the travel times of different links and the risks associated with the volatility of travel time. Furthermore, the computation of this volatility allows for the development of the travel time derivative, which is a financial derivative based on travel time. It serves as a value or congestion pricing scheme based not only on 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 part one (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 the data from fundamental traffic conditions, a scenario-based GCMM is studied. A distribution of the path scenario representing path traffic status is first defined; then, the dependent structure between constructing links in the path is modeled as a Gaussian copula for each path scenario and the scenario-wise path travel time distribution is 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. Different estimation methods are employed based on whether or not the path scenarios are observable: An explicit two-step maximum likelihood method is used for the GCMM based on observable path scenarios; for GCMM based on unobservable path scenarios, extended Expectation Maximum algorithms are designed to
estimate the model parameters, which introduces innovative copula-based machine learning methods.

In part two (Chapter 5), travel time derivatives are introduced as financial derivatives based on road travel times - a non-tradable underlying asset. This is proposed as a more fundamental approach to value pricing. The chapter addresses (a) the motivation for introducing such derivatives (that is, the demand for hedging), (b) the potential market, and (c) the product design and pricing schemes. Pricing schemes are designed based on the travel time data captured by real time sensors, which are modeled as Ornstein-Uhlenbeck processes and more generally, continuous time auto regression moving average (CARMA) models. The risk neutral pricing principle is used to generate the derivative price, with reasonably designed procedures to identify the market value of risk.
Acknowledgements

To my adviser.

To my parents.

To my family.
Contents

Abstract .................................................................................................................. iii
Acknowledgements ................................................................................................. v

1 Introduction ........................................................................................................... 12
  1.1 Motivation ......................................................................................................... 12
  1.2 Problem statement and thesis objective ......................................................... 19
    1.2.1 Transportation networks ............................................................................ 19
    1.2.2 Definition and measurement of travel time ............................................... 20
    1.2.3 Routing Decisions based on predicted travel time distributions .......... 22
    1.2.4 Major issues of interest ............................................................................. 23
  1.3 Thesis outline ................................................................................................... 29

2 Literature Review .................................................................................................. 30
  2.1 Travel time estimation ...................................................................................... 30
    2.1.1 Link travel time estimation ....................................................................... 31
    2.1.2 Path travel time estimation ....................................................................... 34
2.2 Dependent structure and copula theory ........................................ 38
  2.2.1 Definitions ................................................................. 38
  2.2.2 Fundamental theorem of copula ........................................ 40
  2.2.3 Tail dependence of the copula ......................................... 41
  2.2.4 Nonparametric estimation of density and conditional density ...... 46
2.3 Derivative pricing ................................................................. 49
  2.3.1 Derivative fundamentals ................................................... 49
  2.3.2 Probability settings of derivative pricing .............................. 52

3 Link Travel Time Estimation and Routing Decisioning through Copula Methods ......................................................... 57
  3.1 Profile description of link travel time data .............................. 58
  3.2 Generating travel time distributions based on copulas ............... 60
    3.2.1 Copula models and two step maximum likelihood estimation .... 60
  3.3 Routing decisions based on estimated travel time distributions .... 65
  3.4 Numerical analysis ............................................................ 72
  3.5 Further development: reliable estimates and similarity-based analysis .... 85
    3.5.1 Generation of reliable estimation .................................. 85
    3.5.2 Similarity-based copula reconstruction ............................ 89

4 Path Travel Time Distribution Estimation through Gaussian Copula Mixture Models ................................................... 92
4.1 Estimating travel time using scenario-based GCMMs 94

4.1.1 Background 94

4.1.2 Observable and unobservable path scenarios 98

4.1.3 Definition of a Gaussian copula mixture model (GCMM) 99

4.2 A Gaussian copula mixture model based on observable path scenarios 101

4.2.1 Scenario decomposition and summation 104

4.2.2 Estimation of path travel time distributions conditioned on a specific path scenario 105

4.2.3 Properties of GCMM with observable path scenarios 112

4.3 GCMM based on unobservable path scenarios and extended expectation maximization algorithms 125

4.3.1 Fixed-marginal-distribution GCMM 126

4.3.2 Varying-marginal-distribution GCMM 130

4.4 Numerical analysis 141

4.4.1 Identification of Observable Path Scenarios via Cluster analysis 141

4.4.2 Sensitivity test for GCMM based on observable path scenarios 144

4.4.3 Simulation Experiment for GCMM based on unobservable path scenarios 149

4.4.4 Empirical Test for GCMM based on unobservable path scenarios 154

5 Travel Time Derivatives: Market Analysis and Pricing 157

5.1 Initiation and necessity analysis 158
# List of Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (\Omega, \mathcal{F}, P) )</td>
<td>A standard probability space</td>
</tr>
<tr>
<td>( \Omega )</td>
<td>The sample space</td>
</tr>
<tr>
<td>( \mathcal{F} )</td>
<td>The ( \sigma ) algebra</td>
</tr>
<tr>
<td>( P )</td>
<td>The ordinary probability measure</td>
</tr>
<tr>
<td>( t )</td>
<td>Time or the index set.</td>
</tr>
<tr>
<td>( {\mathcal{F}<em>t}</em>{t&gt;0} )</td>
<td>The filtration</td>
</tr>
<tr>
<td>r.v</td>
<td>A random variable</td>
</tr>
<tr>
<td>( X(\omega) )</td>
<td>The general notation for a random variable</td>
</tr>
<tr>
<td>( x )</td>
<td>A specific value taken by the random variables</td>
</tr>
<tr>
<td>( X_t )</td>
<td>The general notation for a stochastic process</td>
</tr>
<tr>
<td>( E(X) )</td>
<td>The expectation of a random variable ( X )</td>
</tr>
<tr>
<td>( \text{Var}(X) )</td>
<td>The variance of a random variable ( X )</td>
</tr>
<tr>
<td>( \text{ATT} )</td>
<td>Average Travel Time</td>
</tr>
<tr>
<td>( T_t )</td>
<td>Travel time for a link given a time ( t )</td>
</tr>
<tr>
<td>( T_m )</td>
<td>The time instant when a vehicle passes a given monument</td>
</tr>
<tr>
<td>( \mu )</td>
<td>The mean value of a random variable</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>The standard deviation of a r.v and a volatility of a process</td>
</tr>
<tr>
<td>( CV )</td>
<td>Coefficient of variation</td>
</tr>
<tr>
<td>( p )</td>
<td>A path</td>
</tr>
<tr>
<td>( l )</td>
<td>A link</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>m</td>
<td>A monument</td>
</tr>
<tr>
<td>L</td>
<td>A link set</td>
</tr>
<tr>
<td>L(p)</td>
<td>The links forming a given path p</td>
</tr>
<tr>
<td>H</td>
<td>A general univariate/multivariate functions</td>
</tr>
<tr>
<td>V</td>
<td>A value function</td>
</tr>
<tr>
<td>C'</td>
<td>A sub copula function</td>
</tr>
<tr>
<td>C</td>
<td>A copula function</td>
</tr>
<tr>
<td>F(x)</td>
<td>A univariate cumulative probability function for a r.v, similar to G(x)</td>
</tr>
<tr>
<td>λ_U</td>
<td>The upper tail dependence of a copula</td>
</tr>
<tr>
<td>λ_L</td>
<td>The lower tail dependence of a copula</td>
</tr>
<tr>
<td>f</td>
<td>The density function for the random variables</td>
</tr>
<tr>
<td>ρ</td>
<td>Spearman’s ρ, a dependent measure</td>
</tr>
<tr>
<td>τ</td>
<td>Kendall’s τ a dependent measure</td>
</tr>
<tr>
<td>ρ</td>
<td>Parameters of copulas, the same applies to θ, δ, γ, κ</td>
</tr>
<tr>
<td>K(x)</td>
<td>A univariate kernel function</td>
</tr>
<tr>
<td>W(x)</td>
<td>A univariate kernel function</td>
</tr>
<tr>
<td>π</td>
<td>The weight of mixture model</td>
</tr>
<tr>
<td>Ψ</td>
<td>The cumulative probability function of standard normal distribution</td>
</tr>
<tr>
<td>ψ</td>
<td>The density function of standard normal distribution</td>
</tr>
<tr>
<td>Q</td>
<td>The risk neutral measure</td>
</tr>
<tr>
<td>S_t</td>
<td>The stock price</td>
</tr>
<tr>
<td>N_t</td>
<td>The jump measure/poisson random measure</td>
</tr>
<tr>
<td>L_t</td>
<td>The levy process</td>
</tr>
<tr>
<td>B_t</td>
<td>The Brownian motion</td>
</tr>
<tr>
<td>ρ(X)</td>
<td>The transportation risk measure</td>
</tr>
<tr>
<td>U(x)</td>
<td>The utility function for a given value x</td>
</tr>
<tr>
<td>VaR(X)</td>
<td>Value of risk, given a random variable</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>$L$</td>
<td>Log likelihood function</td>
</tr>
<tr>
<td>$P$</td>
<td>The dependent parameter for a Gaussian copula</td>
</tr>
<tr>
<td>$\hat{\pi}$</td>
<td>The scenario frequency after considering the pseudo observations</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>The lagrange multiplier</td>
</tr>
<tr>
<td>$G$</td>
<td>Payoff when buying travel time derivatives</td>
</tr>
<tr>
<td>$P$</td>
<td>The price of travel time derivatives</td>
</tr>
<tr>
<td>$K$</td>
<td>The strike of options</td>
</tr>
<tr>
<td>$T_{us}$</td>
<td>The standard country travel time index for USA</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>The market value of risk</td>
</tr>
<tr>
<td>$v$</td>
<td>The value function or the value of a claim</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Predict the conditional distribution of travel time on downstream links . . . 24

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) . . . . . . . . . . 26

1.3 Assembly of the path travel time distribution from link data . . . . . . . . 26

1.4 Price of the travel time derivatives leads to optimal routing decisions: Between two alternative routing choices, a higher derivative price predicts a potentially lower travel time and hence the corresponding link is selected by the traveler. 28

3.1 Different number of observations for links and paths (In the figure, two travelers traverse Link AB and then Link BC and Link BD respectively due to different routing decisions. As a result, there are more data records corresponding to link AB(2) than any path going through it: Path ABC and Path ABD both have one data record (vector) respectively) . . . . . . . . . . . . 59

3.2 Test network N95-US 195, the red frames are the targeted monuments (Monument points A,B,C,D,F,G are defined. The links between these points are highlighted to represent the paths of interest in the analysis of this section) . 73
3.3 Extracted topology from the test network on N95-US 195, the red frames are the targeted monuments ................................. 75

3.4 Fit travel time distribution on AB .................................................. 77

3.5 Empirical distribution v.s. simulated data based on different copulas ...... 79

3.6 The comparison of tails of the joint structure. Red: lower tail; Blue: upper tail 81

3.7 Estimation of the copula and conditional probability function of AB based on BB1 copula ................................................................. 83

3.8 Estimated conditional density(Red) v.s. Empirical conditional density (Histogram: Blue; Kernel smoothed distribution: Black) given current observation of 67 mile/hour per hour at a loop detector ........................................ 86

3.9 Combined conditional probability density function(Green) compared to two original estimations using a T copula (Red) and a BB1 copula (Blue). .... 88

3.10 Differences in estimation between the borrowed copula (Red) and the original copula(Blue)). Empirical pdf: Blue); Estimated pdf: Red ...................... 90

4.1 Assign three path travel time vectors to predefined observable path scenarios: In this figure, the path contains three links and there are two different travel time distributions on each link(1 with lower mean value, 2 with higher mean value) for different path scenarios. Eight path scenarios can be defined in this setting, each of which contains a vector of three link travel time distributions. More specifically, 3-2 in the bigger triangle in the first row means the particular travel time value on Link 3 is high so that is categorized in the second travel time distribution on Link 3. ..................................................... 102
4.2 Estimate the path travel time distribution using the copula in each path scenario. The path travel time distribution in this path scenario is estimated based on three link travel time distributions using a three dimensional copula that is estimated based on the categorized data in Step 1.

4.3 Aggregate path scenarios according to their historical frequency. The overall path travel time distribution is a weighted sum of the path travel time distributions corresponding to all path scenarios. Two of them are displayed in this figure.

4.4 Generate pseudo path travel time observations (To collect data for path a-b-c, the trips for Traveler A and Traveler B are connected together under the condition that shortly after A completes route a-b, B starts his trip on Link c).

4.5 Comparison of GMM and GCMM base case: n: data index; m: iteration index; k: copula index; i: dimension index.

4.6 Comparison of GCMM base case and GCMM with unsynchronized data: n: data index; m: iteration index; k: copula index; i: dimension.

4.7 Hierarchy cluster for travel time data of a three-segment path (Top: Original data: Bottom left: top 30 clusters; Bottom right: Number of observations across clusters).

4.8 Estimated path travel time distribution: Based on Major Three Clusters: Green; With Other Outliers: Red; Empirical: Black.

4.9 The experimental network in New Jersey.

4.10 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).
4.11 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) ................................................................. 146

4.12 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 .......................... 147

4.13 Left: 2 scenarios per link; Right: 8 scenarios per link; Estimated pdf: Red; empirical: Green; scenario-specific cdfs: Blue ................................................................. 147

4.14 12 Path travel time based on the approximation; Path 1: Red; Path 2: Blue 148

4.15 Comparison of the marginal link travel time distributions on one selected link across different models ................................................................. 151

4.16 Comparison of the cluster structure between the two links across different models ................................................................. 152

4.17 QQ-plots of the estimated path travel time distributions to empirical distribution across different models ................................................................. 153

4.18 Top: Comparison of pdf (Red: GMM; Blue: GCMM base case; Cyan: GCMM with unsynchronized data; Black: Empirical); Bottom left: Green QQplot Empirical(x) v.s. GCMM base case(y); Bottom right: Black QQplot Empirical(x) v.s. GCMM with unsynchronized data(y) ................................................................. 156

5.1 Travelers and travel time (quality of service) protection - good scenario: Good traffic conditions lead to low High Congestion Days. Therefore, HCD call option is out of money and there is no payoff to the traveler. .................. 162
5.2 Travelers and travel time (QOS) protection - bad scenario: Bad traffic conditions lead to high High Congestion Days. Therefore, HCD call option is in the money and there are payoff to the traveler to compensate his economic loss due to traffic delay.

5.3 The price of travel time derivatives indicates short term profit and loss due to traffic conditions: Between two alternative routing choices, a higher derivative price predicts a potentially lower loss and hence the corresponding link is chosen by the traveler.

5.4 The price of travel time derivatives indicates profit and loss due to long term traffic conditions: Between two alternative travel plans, a higher derivative price predicts a potentially overall lower loss in a year and hence the traveler chooses to drive.

5.5 Alternative risk transfer between the transportation and financial industries

5.6 Experiment Network and Sample Travel Time Data

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

5.8 80E ARIMA model
List of Tables

2.1 Definition of copula functions ................................. 44
2.2 Tail dependence for typical copulas ............................. 44
3.1 Definition of distance measures ................................. 63
3.2 Tail dependence for typical copulas ............................. 64
3.3 Definition of monument links ................................. 73
3.4 Distance measures for parametric estimators .................. 77
3.5 Goodness-of-fit for Copulas ................................. 78
3.6 Decision statistics for different rules .......................... 83
3.7 Deviation of estimates in cross validation .................... 84
3.8 Square distance between estimated speed distribution and empirical speed distribution ................................. 85
3.9 Parameter estimation for the T copula .......................... 87
3.10 Decision statistics for mean-variance rule on AB by different estimation procedures .................. 88
4.1 Decision statistics for different rules for the two paths ......... 148
5.1 Payoff of typical weather derivatives ........................................ 160
5.2 Canadian Degree Days Index (HDD) Futures traded on Chicago Mercantile
Exchange ................................................................. 161
5.3 Payoff of traditional road toll, $P$ denotes the toll amount .................. 164
5.4 Payoff of dynamic congestion pricing, $P$ denotes the toll amount ........ 164
5.5 Payoff of a travel time derivatives. $P$ denotes its price, $ET$ is the expected
travel time and $K$ is strike price of the derivative contract. The formula
describes payoff as a function of realized travel time and the strike price; the
price of the derivative contract is not zero but is calculated in proportion to
market participants’ expectations regarding future traffic conditions......... 164
5.6 US Congestion Days Index (CDD) Futures traded on Chicago Mercantile Ex-
change ................................................................. 165
5.7 Comparison between insurance on travel time and travel time derivatives .. 167
5.8 Different participants (B means benefit in good quality of service(QOS); H
means hurt in good QOS; market makers and investors can hold both types
of derivatives.) .......................................................... 172
5.9 Seasonal effects in the 80E data .............................................. 188
5.10 Model selection for 80E data .................................................. 190
5.11 Models with best fit for different paths ...................................... 191
5.12 Coefficients mapping between discrete and continuous ARMA models .... 192
5.13 Coefficient estimation for discrete and continuous time ARMA models . . . . 193
Chapter 1

Introduction

1.1 Motivation

Transportation networks are complex systems with substantially varied performance, much of which results from the cumulative effects of many individual behaviors. When individuals compete for the use of limited traffic paths of varying capacities, 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, particularly when many people try to reach their destinations at about the same time. Drivers tend to be accustomed to everyday congestion, and so plan for it. They tend to leave home early enough to reach their destinations on time. Unexpected congestion, however, is more troubling to travelers. When travel time takes much longer than expected, people complain. Thus, the reliability of travel time is of greater concern. Travelers should make decisions based not only on expectations of travel time but also its uncertainties. This brings risk into the picture. Those who wish to avoid being late will pay greater attention to the tail of the distribution, but for all travelers the shape of the expected travel time distribution
is important, not just its deterministic expectation. Consequently, travel time reliability is of particular concern, Park et al. (2011) and Rakha et al. (2006).

The objective of this research is to develop analytical methods for the estimation of travel time reliability and risk faced by individual travelers on individual segments and paths of real roadway networks. In order to find methods to reduce uncertainty, it is necessary to characterize and to quantify the magnitude of the uncertainty. 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 the overall traffic system.

This thesis first focuses on modeling methods for the prediction of vehicle travel time in a transportation system. The model begins by characterizing the transportation system as a graph, in which, monuments \( m \) and monument-to-monument \( m2m \) links are defined to represent the road network in an abstract fashion. Specific steps are taken 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.

This section of the thesis focuses on the development of a methodology for predicting 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 on its extensions in neighbor-
ing sections. An approach based on the copula method is developed in this thesis, Nelsen (2006). With copulas, the dependent structure between links of a transportation network can be described and observations of real links can be used to estimate the interdependence of travel times on different links. The travel time distribution on individual links of a network can then be 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 locations. 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 assumes the travel times between different links are independent. However, this approach is unrealistic because an important feature of transportation networks is the dependent structure of the travel times of different links. In this thesis, a copula is used to model the dependent structure between the travel times of different links given specific traffic conditions along the path, which is referred to as a path scenario. Using the copula, a conditional path travel time distribution is estimated for the path scenario. Moreover, it is assumed there are several path scenarios for the overall path traffic conditions. The unconditional path travel time distribution can be obtained by taking the weighted sum of these conditional path travel time distributions.

Such path scenarios are either observable or unobservable. The traffic status along a link can be observed and defined through experience by considering time of day, weather and the occurrence of any event that may change the travel time through the path significantly. As a result, travel time observations can be classified into these pre-identified path scenarios to estimate corresponding model parameters. Alternatively, it is also reasonable to claim these path scenarios can be unobservable due to the complex nature of transportation network dynamics; a given path data vector may belong to different path scenarios in a probabil-
ity sense. Suitable machine learning methods are developed to extract such hidden path scenarios from a data set so that path travel time distributions can be estimated.

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 is 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 section of the thesis, the coupling between travel time and travelers’s economic benefits is studied. The focus is on the volatility of travel time and its implications 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 with discussions of its market making mechanisms, product design and pricing methods. The introduction of such financial derivatives based on travel time enables an effective hedging tool against uncertainties in travel time, an innovative value pricing scheme and additional funding resources for the transportation systems.

Traditionally, value pricing has been used as an incentive to change typical consumption of transportation resources by pricing the use of those resources according to changes in demand ranging from over-capacity to under-capacity, so as to improve the overall utility of the transportation system. There are many value pricing schemes in the literature, Xin & Levinson (2007), Vickrey (1992) and de Palma & Lindsey (2011), and these include 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 experienced travel times. A static price cannot change to address this concern. In typical dynamic pricing schemes, a higher price could be charged during rush hour, but this price is usually fixed a priori so it does not address either expected or experienced travel times. System optimal pricing addresses the concerns of system performance instead of the experience of individual travelers, which may not serve the purpose adequately.

2. Traditional road pricing schemes charge a premium without providing economic pay-off/protection to travelers according to their experienced travel times. The charge is the access fee to the traffic resources, which is irrelevant to the quality of service rendered. Therefore, the charge amount does not contain information about expected traffic conditions in the future. Furthermore, without providing pay-off to individual travelers, the amount of payment/premium is generally small and not flexible according to changes in experienced travel times. Therefore, the potential of traditional road pricing in changing traveler behaviors is limited.

To mitigate the limitations above, travel time derivatives are introduced in this thesis as a fair market value pricing scheme, which prices quality of transportation service according to the value actually received by the purchasers (travelers). In the context of roadway value pricing, such a scheme would provide cash payoffs/protection against adverse changes in experienced travel time on that segment of the roadway. The traveler’s toll payment would therefore fluctuate according to the changes of the value of such protection, which is determined by travelers’ experienced travel times during a trip: The price is then determined as the expectation of this payoff. As a result, travelers pay tolls associated with the expected
quality of service they enjoy and receive real economic protection according to the their experienced quality of service during the trip.

Furthermore, given the close relationship between current pricing of such travel time derivatives and experienced travel time in the future, travelers would alter their travel behaviors according to changes in the price. These behavioral changes would reflect market participants’ general concerns with future performance of the transportation system and such a new pricing plan would potentially lead to innovative routing schemes. In order to address such possibilities, travel time derivatives are introduced in this thesis.

First, some analysis of financial derivatives is introduced to explore related concerns. 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 derivatives, John (2000). A derivative is a financial instrument whose value is derived from one or more underlying assets, market securities or indices, which are referred to as 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.

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

In this context, 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 and it can be measured by computing the standard deviation of the one-step ratios of the historical travel time series. By measuring such dynamic structures 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) and Carmona & Danilova (2003).

Besides being a form of a flexible value pricing (congestion pricing) scheme that can potentially change traveler behaviors, there are other benefits to introducing such travel time derivatives: First, travel time derivatives can serve as alternatives to insurance against traffic service quality. Travelers purchase derivative contracts before their trips begin and obtain a payoff when they complete their trips. The payoffs are directly linked to the performance of the path travel time that they experienced. Second, introducing travel time derivatives is also motivated by the need to hedge traffic-related risks to other participants, such as logistic companies and various delivery services whose welfare is related to the transportation system. Third, travel time derivatives can further diversify risk for financial markets. As derivatives based on weather and energy are traded in the market, travel time as a stochastic process correlated to these entities should also be introduced into the financial market to enable further risk diversification.

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. While travel time
is the measure of interest, the findings in this thesis can be extended to other measures. In
the following sections, the research methods are summarized in greater detail.

1.2 Problem statement and thesis objective

This section introduces the framework for this research, including the transportation network
setting, definition of travel time and major issues of interest.

1.2.1 Transportation networks

Travel accommodated by transportation systems can be conceptualized as continuous enti-
ties, such as fluids in piping systems, or discrete entities, such as cars in roadway systems.
The latter is of concern here. For such systems, graph models with associated link and node
characteristics are the norm. This thesis is focused on the travel time characteristics of such
networks.

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 \( l \) studied together for their dependent
structure is called a link set \( L \). A path \( p \) is a special link set in which the links are ordered
by the sequence in which they are visited by a traveler on a trip.

According to 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 times across
each link and node traversed from origin to destination.
Traditionally, links are considered the physical segments spanning intersections, and nodes are the 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 time on paths going straight differ from those on paths turning left or right. This is an important practical consideration. 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(m) are defined as reference points 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 an m2m network is simply the sequence of m2m links 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 physical node to physical node).

### 1.2.2 Definition and measurement of travel time

Two different measures of road travel time are used in this thesis and the empirical travel time data demonstrate characteristics distinct from those measures, including: (a) the travel time of an individual traveler and (b) some assembled measures over an aggregation of individual travel times (e.g., average travel time).
In the first section of this thesis, individual travel time is examined. By definition, the travel time experienced by an individual traveler is termed individual travel time, and it is the difference in the arrival time between two specific points in a given trip. Historical observations of individual travel time are typically 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 positioned in vehicles are the data source for individual travel time in this part of the study. Most of the data used here were recorded every three seconds throughout each trip. As the data were collected using only certain cars with GPS devices, the travel time observations in the data set can be sparse in time and difficult to analyze by typical time series methods. However, such observations of actual travel time at different points in time 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 risks caused by uncertainties in travel time.

In the second section of this thesis, 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 $t$, denoted as $ATT$ is defined as the average of experienced travel times by all travelers in the link entering at time interval $t$, denoted as $T_t$. It is a summary description of the traffic conditions on that link at a given time. Ideally, such measurement can be obtained by taking a small time interval $\delta t$ and calculating the travel time experienced by the vehicles entering the link within $[t - \delta t/2, t + \delta t/2]$ assuming traffic conditions are 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 based on 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 reveal the immediate 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 than 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 for providing protection against travel time risks on that link, and the price should be more stable.

1.2.3 Routing Decisions based on predicted travel time distributions

Using the aforementioned terminology, the routing decisions of a traveler can be described as follows: A traveler traverses from one m2m link to another during a trip. At a given monument, the destination is chosen by comparing the estimated travel time distributions associated with all alternative paths. To make a 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. Different optimal paths can be identified if different risk measures are utilized. Detailed 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 at which the traveler enters the link. Such forward-looking data contain sufficient information for 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, the storage of this 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 dependent structure between links should be considered.

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 which describes the dependence 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 and leads to reasonable routing decisions.

### 1.2.4 Major issues of interest

This decision framework poses substantial challenges to the research due to various factors. To focus on the topic and reduce those challenges, this thesis emphasizes the uncertainties of travel time and the dependence inherent in the problem given a limited number of alternate paths. This assumption is suitable for a practical road network on which the number of
alternate 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?

On 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 a 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 decision time. To enable this calculation, a model should be constructed 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)

A simple case is to estimate the travel time distribution of a downstream link. To address this issue, copula methods are used to estimate the link travel time distribu-
tions conditioned on current observations. The dependent structure between links is modeled by only one copula function, and the marginal distribution of travel times 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 segment travel times through a general 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, and the empirical evidence for this is shown in Figure 1.2. Clearly, the true path travel time distribution has much heavier tails than the estimated distribution, which assumes links are independent.

To conduct the aggregation, the joint distribution of travel time on the links constituting a given path is estimated, and 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 scenario is uniquely specified by a series of factors, including time of day, weather, events along the path, and so forth. Once specified, the path scenario determines the vector of link travel time distributions on the m2m links and the dependence between them. Given the copula structure is
stationary within this path scenario, a Gaussian copula can be used to model this dependent structure. In this way, a path scenario specifies a joint distribution for the travel time of all links constituting the path and hence 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 of the constituting links. This path scenario can be pre-defined according to the traffic context. Alternatively, it can be extracted empirically using maximum likelihood algorithms that are 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 scenarios. Copula models are used to model the dependence within each path scenario. The weights are the frequency number of the traffic scenarios.

This thesis further generalizes the mathematic form of the model introduced above as a Gaussian copula mixture model (GCMM). A GCMM is defined as a weighted sum of finite joint distributions, each of which contains a Gaussian copula. If the path scenarios are observable, they can be predefined and each path travel time observation is classified into a specific path scenario for estimation; if the path scenarios are not observable, they can be updated during certain recursive learning schemes; the travel time observations can belong to different path scenarios with different probabilities.

Following the literature of statistical learning, Gaussian Copula Mixture Models can be viewed as extension of Gaussian Mixture Models (Pekka Paalanen (2006) and Ming-Hsuan Yang (2006)), which aim to address the following two concerns:

- Heavy-tailed data require increasing numbers of clusters to fit with GMMs. To control number of clusters, heavy tails on marginal distributions should not lead to significantly greater clusters given the same underlying dependence structure.
- GMMs are usually applied to a synchronized data matrix of dimension $M$ and number of observations $N$. In many problems, there are numerous unsynchronized data each dimension, the number of which is denoted as $n_m$ for the $m$-th dimension. Such data should be utilized to update the joint distribution shared by the different dimensions.

To address the concerns, copulas are introduced into mixture models and new Expectation Maximum type algorithms are developed to estimate their parameters.

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?
To clarify these issues, a typical travel time derivative contract is demonstrated below in Figure 1.4. The payoff of the derivative is a function of the experienced travel time in the future, and the price of the contract is the conditional expectation of payoff. Therefore, the price of the contract corresponding to each link illustrates the general market view of 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 the traveler receives at the end of the trip compensates the risks incurred by unexpected traffic conditions. To enable such a mechanism, market making, product design and pricing are factors that should be addressed. They are each discussed in detail in this thesis.

More specifically, Ornstein-Uhlenbeck process and more generally, continuous time auto regression moving average (CARMA) models are used to model travel time, and derivative prices are based on the risk neutral pricing principle under incomplete market conditions, Benth & Benth (2007) and Härdle & Cabrera (2012).

Figure 1.4: Price of the travel time derivatives leads to optimal routing decisions: Between two alternative routing choices, a higher derivative price predicts a potentially lower travel time and hence the corresponding link is selected by the traveler.
1.3 Thesis outline

In summary, the interaction and similarity between traffic systems and financial systems is highlighted in this thesis. Copula methods and stochastic models from the financial mathematics 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.

The rest of this thesis is organized as follows. Chapter 2 provides background and a literature review of travel time estimation methods and related mathematic theory. Chapter 3 describes the link travel time estimation through copula models. Chapter 4 describes the estimation of path travel time distribution. Path scenarios are defined to represent different traffic statuses along a path and Gaussian copula mixture models (GCMM) are used to model the dependence between different links. The GCMM based on both observable path scenarios and unobservable path scenarios are discussed. Chapter 5 introduces travel time derivatives and their pricing methods. Chapter 6 concludes this thesis and describes potential future work.
Chapter 2

Literature Review

This chapter reviews the literature on travel time estimation, dependent structures and derivative pricing methods. In this thesis, dependent structure models including copulas are used to model travel time in order to generate routing guidance, while asset pricing models are used to price travel time derivatives. The literature on these topics provides the theoretical background for this thesis.

2.1 Travel time estimation

As noted in Chapter 1, road travel time estimation has been an important topic in transportation research. Numerous authors have examined methods for accurate prediction of travel time on roads. In previous work, authors have introduced various models to describe the dependent structure of travel time processes; however, general dependent structure models (copulas) have not yet 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 forecasts of travel times along the links are essential aspects of this research. In this section, different methods for calculating link travel time estimations are reviewed, with a focus on the dependence model between different links.

1. Time series methods: In Vemuri et al. (1998) and Ho & Lee (2004), the problem of short-term forecasting of traffic delays is formulated as a time series evolution problem. Delays are predicted under a regression framework. Zhang & Rice (2003) propose 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) use the Kalman filter to analyze travel time changes. The system model is described with a state equation and an observation equation based on traditional traffic flow theory. Cathey & Dailey (2003) identify 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.

3. Neural networks: Park et al. (1999) employ a spectral basis artificial neural network (SNN) to predict link travel time. In Jiang & Zhang (2003) and Mark et al. (2004), artificial neural network models are applied to describe the relationship between mean travel time and flow. Neural networks reconstruct the relationship by simple functions, but usually the dependent structure between the travel times of different links is not explicitly modeled.
4. Flow theory model: The models presented in Van Grol et al. (1998) and Petty et al. (1998) are based on macroscopic hydrodynamic traffic flow theory. Nie & Zhang (2005) propose 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 the link travel time for a vehicle entering a link at time \( t \) 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 the literature to estimate link travel times, the underlying modeling of the dependence between links shares some common characteristics:

1. Limited model formulations for addressing the dependence between travel times of different links:

First, most research assumes that the travel times of different links are independent, and the focus is on predicting the future change in travel time on a single link over time.; Second, for research that does not assume independence, the typical assumption has been that travel time of different links is subject to a joint normal distribution, Dailey et al. (2000). This assumption makes the result tractable, but the limitation is that a single joint normal distribution cannot fully describe the real dependence between the travel times on different links in a realistic transportation network. Third, other
advanced methodologies including neural networks may produce travel time predictions by considering the interdependence of travel times along different links implicitly based on the network’s complex structure, while more intuitive models achieve better performance with parsimonious settings and direct transportation interpretations and potentially other improvements.

2. Inadequate modeling of the empirical travel time distribution:

In terms of output, most previous research focuses on the prediction of mean travel time, whereas the prediction of future empirical travel time distributions is not fully addressed. In a real traffic network, empirical travel time distributions usually have heavy tails, demonstrating the occurrence of extreme travel times. Such observations should be utilized when modeling travel time distributions.

3. Challenges of describing the tail dependence of travel times:

Compared to minor changes in travel time, travelers tend to care more about the extreme values of travel time. They need to be informed of the presence of congestion in downstream links given the traffic status involved in making their travel plans. In other words, the dependence in the tails of travel time distributions should be studied to describe travel time risks and correlations because the conventional Pearson correlation measure is not adequate for this purpose.

4. Variation in data sources:

Besides the difference in mathematical formulations, 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.

(a) 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. Av-
verage travel time (ATT) on the link can be calculated using these measurements. Usually, the data points are obtained periodically and therefore are suitable for regular time series analysis.

(b) Vehicle-based measurement:

Floating car data and GPS-based data are typical vehicle-based measurements. Usually, the onboard device transmits 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 using this method are recorded at random time points, and therefore they are not suitable for time series analysis unless aggregated into discrete sets by data matching techniques. The data used in chapters 3 and 4 are the travel time data collected by Copilot GPS devices, which require suitable models to process them.

In the context above, an innovative model that can effectively describe the flexible dependent structure of the travel times of different links, predict empirical travel time distributions and address the occurrence of extreme travel time values should be developed. To this end, copula models are selected and justified in Chapter 3.

2.1.2 Path travel time estimation

After the link travel time distribution is characterized, a further challenge is how to assemble link travel time distributions to characterize 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 a path travel time distribution, spatial dependence and temporal dependence should be considered. Here, spatial dependence means that changing travel times on a given link may depend on the travel times on upstream links. Temporal dependence means that changing travel times on a given link may depend on overall traffic system conditions, which
change primarily according to time of day. This section reviews the 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 times. The assumption of the research is that the travel times along individual links at a particular point in time are statistically independent and that 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, the methods of estimating estimating such a probability distribution of travel times by considering the underlying dependent structure between upstream and downstream links is the focus of this thesis.

In Pattanamekar et al. (2003), it is proposed that in order 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 authors suggest it is impractical to estimate this function and instead they use a three-point polynomial approximation to estimate mean travel time instead of studying the joint probability density function in detail.

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

\[ CV = \frac{\mu}{\sigma} \]
The author demonstrates several methods for the estimation of path travel time variance from its component segment travel time variances:

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

\[
\sigma_p^2 = (\sum \mu_j)^2 \text{med}_{l \in L(p)} \frac{\mu_j}{\sigma_j}
\]

\[
\sigma_p^2 = (\sum \mu_j)^2 \left( \frac{\max_{l \in L(p)}}{2} \frac{\mu_j}{\sigma_j} + \frac{\min_{l \in L(p)}}{2} \frac{\mu_j}{\sigma_j} \right)
\]

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 of each other.

Sherali et al. (2006) derive another model based on the CV to estimate the path travel time variance. The minimum and maximum 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:

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

This formula is a modification and extension of Rakha’s estimation.

For investigations based on assumptions about the joint distribution of the travel times of different links, the joint normal assumption is the typical assumption of most studies with respect to the correlated link travel time estimation prediction of Dailey et al. (2000). Gao & Chabini (2006) consider the simplified correlation model over time but not between different links.

Hunter et al. (2009) use Bayesian networks to estimate travel time changes by considering the day-to-day changes in travel time. The broad framework is applicable for general traffic
systems with dependence between link travel times. Furthermore, a practical Expectation Maximum type method is proposed to generate path travel time distributions by assuming each link is independent.

Typically, the state of a traffic system can be observable and expert judgment can be used to identify these states; on the contrary, hidden factor models can also be used for estimation of system structure. Fan et al. (2007) study the general properties of statistical factor models in the context of covariance matrix estimation. Bai & Li (2012a) study the identification properties of hidden factor models. By assuming the existence of underlying factors and states, complex dependence structures between the travel times of different links can potentially be modeled in a more effective manner.

Herring et al. (2010) proposes a Coupled Hidden Markov Model to model the dynamics of link travel time in urban network. It is assumed that given the state of a link, its travel time distribution is independent from all other traffic variables and state transition may be correlated to a small subset of neighboring links.

Ramezani & Geroliminis (2012) uses Markovian chains to predict path travel time distributions. It is assumed that the traffic status of the current link and that of next link forms a Markov chain. The path travel time distribution is calculated using the product of the sequence of these pair-wise Markov matrixes assuming the transitions between different link pairs are conditionally independent.

Gaussian mixture models have been used widely in various applications (Pekka Paalanen (2006) and Ming-Hsuan Yang (2006)) and the Expectation Maximum algorithm has been utilized for estimating their parameters (Dempster et al. (1977)). The convergence properties of such Expectation Maximum algorithms have been discussed in Xu & Jordan (1996). In Park et al. (2011), a multi-state model is employed to fit a mixture of three Gaussian distributions into travel time observations of an expressway corridor. Each normal distribution is associated with an underlying traffic state providing quantitative uncertainty evaluation.
The multi-state mixture model results in a better fit.

Zhan et al. (2013) estimates link travel time by using multinomial logistical regression to model routine choices, assuming travel time between different links are joint-normally distributed. Mixture model which is estimated by the Expectation Maximum algorithm is employed to model the posterior probability of the observed taxi trip travel time given link travel time parameters.

Although GMM is a generally well-recognized model, each component of a GMM is a multivariate gaussian distribution that cannot effectively capture heavy tails and the number of components become sensitive w.r.t heavy tails. The introduction of more flexible components may help to further reduce number of components when working with heavy-tailed data.

In summary, the research on the real time dependence of link travel times is far from complete, and the dependence between non-normally distributed link travel times should be addressed as the next development in the literature.

2.2 Dependent structure and copula theory

Because the dependence of non-normally distributed travel time is examined 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 is introduced below:

2.2.1 Definitions

Dependent structure is based on the dependent relationship between random variables. Mathematically, it is expressed using copula functions. The approach to formulating 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 is subject to 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 copulas is given below based on the definition of the 2-increasing functions, grounded functions, and sub-copulas, which are defined in U.Cherubini (2004) and Nelsen (2006).

**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 deconstructed into its $N$ separate marginal distributions and a copula function, which describes completely 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 $x, y \in \bar{R}$,

$$H(x, y) = C(F_X(x), F_Y(y))$$

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:

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 \( \mathcal{F} \) 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

\[
H(x, y|\mathcal{F}) = C(F_X(x|\mathcal{F}), F_Y(y|\mathcal{F})|\mathcal{F}), \forall x, y \in \mathbb{R}
\]

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 tail dependence, which provides a measure for extreme co-movements in the lower and upper tails 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 copula theory is given below:
1. The density of 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, 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 copula theory, Nelsen (1995) and Fredricks & Nelsen (2007).

3. To estimate the parameters of a copula model, the usual method is a two-step maximum likelihood estimator in which the marginal distribution is estimated first and the 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. Copula models can describe more flexible dependent structures than those implied by a joint Gaussian distribution. This increased scope means the model is capable of describing flexible marginal distribution and dependence structures.

2. Copula models can be used to estimate the dependence between extremely high travel times. This use shows the probability of the 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 marginal distributions (i.e., the univariate travel time distributions) and dependent structure between links (the copula) separately. The former
can be estimated with a larger volume of data because there are usually relatively many travel time observations on a single link. To estimate the copula functions between the travel times of two focused links, however, a pair of travel time observations on both links should be identified. Typically the number of such pairs is smaller than the number of travel time observations on either of the two focused links.

4. Copula models can be estimated through maximum likelihood methods based on time sparse data. If the correspondence between link observations is defined, a vector of such observations can be collected over time, and parameters can be estimated based on such data sets.

5. Copula parameters 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 structures.

Compared to the characteristics of GPS data (the empirical distribution is heavily non-Gaussian; the data collected are aperiodic and space sparse; data is insufficient for model parameter estimation) which is highlighted in section 1.2.2. Copula methods are suitable for conducting analysis of sparse data sets generated by GPS devices.

To employ copula methods for the estimation of travel time distribution, the copulas used in this thesis are defined in Table 2.1:

To compare different copulas, their theoretical tail dependence properties are listed in Table 2.2.

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(z) and Upper tail R(z)) are defined as follows:
Table 2.1: Definition of copula functions

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

Table 2.2: 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>$\lambda_L = 2^{\mu+1}(-\sqrt{\mu + 1}\sqrt{1+\rho})$</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>$\lambda_L = 2^{-\frac{1}{\pi\delta}}$</td>
<td>$\lambda_U = 2 - 2^{\frac{1}{\pi\delta}}$</td>
</tr>
</tbody>
</table>

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 distributions and copulas are used in this thesis, covariance matrix estimation methods are reviewed below: The most straightforward estimation method for the correlation matrix is to estimate 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 and semi-definite. Without being positive and semi-definite, such a pa-
rameter matrix leads to defaults in calculation. As a solution, the Lasso method is employed to estimate the correlation matrix in this thesis: The lasso algorithm is first proposed in Tibshirani (1996). It minimizes the sum of squares subject to the sum of the absolute value of the coefficient being less than a constant.

\[
(\bar{\alpha}, \bar{\beta}) = \text{argmin}_{\alpha} \sum_{i=1}^{n} (y_i - \alpha \sum_{j} \beta_j x_{ij})^2
\]

s.t \[ \sum_j |\beta_j| \leq t \]

The Lasso solution is obtained by conducting an iterative procedure that begins from an overall least squares estimate and solves a constrained least squares 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 and semi-definite. The problem of inverse covariance matrix estimates is then set as

\[
\text{argmin}_{\Omega: \Omega \succeq 0} \log \text{det} \Omega + tr(S_{\Omega}) + \rho |\Omega|_1
\]

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

\[
\text{min}_{\beta} \frac{1}{2} W^{1/2} \beta W^{1/2} - b_0^2 + \rho |\beta|_1
\]

, where \( \beta \) can be solved by lasso.
2.2.4 Nonparametric estimation of density and conditional density

In this section, some non-parametric estimation methods, particularly kernel smoothing methods for joint distribution estimation are reviewed. As powerful statistical models, non-parametric estimation methods can be used to estimate the empirical travel time distribution by smoothing the empirical data using certain kernels. Moreover, the conditional distribution of random variables can be estimated based on smoothed distributions. Some basic mathematical definitions are given first, followed by discussion of the integration of parametric and nonparametric methods.

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:

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

The normal kernel is defined as:

$$K(x) = (2\pi)^{-d/2}exp\left(-\frac{1}{2}x^T x\right)$$

Based on the formulas above, the methodology in this thesis is to estimate link travel time distributions through non-parametric methods and to estimate the dependent structure between them through parametric methods. After the non-parametric estimation of the link
travel time distribution, the multivariate dependent structure can be estimated through cer-
tain 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 Gauss-
sian copulas. This flexible parametric model is a powerful estimator of complex dependent
structures. There may be alternative ways of utilizing the non-parametric estimators during
this modeling process. Some of them are discussed below to justify the model selection in
this thesis.

1. Conduct non-parametric kernel estimation for joint distribution. High dimensional
kernel estimation can be conducted to approximate the joint distribution. Hansen
(2004) suggest both a one-step estimator and a two-step estimator. The correspond-
ing properties are discussed. Given that the copula method is selected as the basic
methodology and that the estimation of joint distribution or conditional distribution
for correlated travel times are the output, this approach may not serve the purpose as
it does not differentiate the estimation of marginal travel time distribution from that
of the copula.

2. Estimate both marginal distributions and dependent structures using non-parametric
methods. In the literature, high dimensional kernels can be applied to estimate copula
functions. 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,
as noted in Chen & Huang (2007) and Chen & Huang (2007), which is a meaningful
though separate research topic.

3. Estimate conditional density directly through nonparametric methods. This method
is used to estimate the derivatives of copula function by kernel methods. Fan et al.
(1996) propose 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 using a Taylor expansion:

\[
E(K_{h_2}(Y - y)|x = z) = g(y|x) \\
= g(x) + g'(x)^T (z - x) + \frac{1}{2} g''(x)(z - x)'(z - x) \\
= \beta_0 + \beta_1^T (z - x) + \beta_2 vec((z - x)'(z - x))
\]

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

\[
\sum_{i=1}^{n} (K_{h_2}(Y_i - y) - \beta_0 + \beta_1 T(X_i - x) + \beta_2 vec((X_i - x)'(X_i - x)))^2 W_{h_2}(X_i - x)
\]

This method based on local polynomial regression is a good alternative for the travel time modeling exercise. Recognizing such promising alternatives, this thesis would select more explicit models for the copula structure in order to obtain simpler transportation interpretations.

Given that the copula method is selected as the basic methodology and that the estimation of joint distribution or conditional distribution for correlated travel times are the output, the current modeling approach is most appropriate.
2.3 Derivative pricing

In this thesis, financial derivatives based on travel time are used to hedge against potential risks 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 an 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 in-
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 underlying assets. Black-Scholes theory introduces classical methods to price such financial derivatives based on tradable assets, Black & Scholes (1973). Contrasted with the derivatives above, the weather derivative is one of the major financial derivatives with values 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 temperatures. These derivatives may provide protections to local farmers 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 due to 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 was 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 are mutually beneficial for the parties involved.

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

1. Actuarial pricing

50
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 and computed.

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 it can be extended to price financial derivatives based on non-tradable assets, Luenberger (2004). Benth & Benth (2007) discuss the risk neutral pricing of weather derivatives under incomplete market conditions and Härdle & Cabrera (2012) further discuss the identification of risk neutral measures through market information.

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 cheap and selling high.", Ma (2011). This principle has been used to price interest rate derivatives, Sundaresan (2009) and can be applied to travel time derivatives, Delbaen & Schachermayer (1997) and Hull & White (1990). Different hedging strategies may lead to different pricing methods as shown in Carmona & Danilova (2003).

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 has been found to be an effective way to price weather derivatives in Carmona & Diko (2005) and Xu et al. (2008).

Of interest is the similar stochastic nature of 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 travel time changes are described with suitable stochastic models in later chapters.

### 2.3.2 Probability settings of derivative pricing

Financial derivatives need to be priced. In the context of travel time derivatives, the price of using a stretch of a roadway should be calculated based on expected traffic conditions along that link. To conduct such calculation, necessary probabilistic settings are introduced below: Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t>0}, P)$ be a complete filtration probability space. A random variable is a mapping $X : \omega \to \mathbb{R}$ if it is $\mathcal{F}$-measurable, whereas a family of random variables depending on time $t$, $\{\mathcal{F}_t\}_{t>0}$ is defined as a stochastic process. A process $X_t$ is $\mathcal{F}$-adapted if every $X_t$ is measurable with respect to the $\sigma$-algebra $\mathcal{F}_t$. If the paths $t \to X(t, \omega)$ 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 $\{\omega \in \Omega | \tau(\omega) \leq t\} \in \mathcal{F}_t$, 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 $t \geq s \geq 0 : E[M_t | \mathcal{F}_s] = M_s$, $M_t$ is a local martingale if there exists a sequence of stopping times $\tau_n < \infty$ where $\tau_n < \infty$, and
such that $M_{t\Delta n}$ 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 to describe the dynamic of the spot price evolution. In mathematical finance, classical 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 changes in travel time when pricing travel time derivatives. The general mathematical formula 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 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
distributions, and the corresponding continuous driven process which is a standard
Brownian motion. A more complex driving process can be introduced that leads to
different properties of \( X_t \).

Moreover, the continuous time autoregressive moving average (CARMA) process is de-
defined as the generalization of the discrete ARMA process and was first introduced by Doob
(1944). Phadke & Wu (1974) define formal derivatives and the high order S.D.E for the
CARMA\((p, q)\) models. Harvey & Stock (1988) introduce a multivariate continuous time
model in which an \( n \)-dimensional process is represented as the sum of \( k \) stochastic trends
plus an \( n \)-dimensional stationary term, assumed to obey a system of higher-order autore-
gressive stochastic differential equations (S.D.E).

Based on the specification of the model, the problem of fitting continuous time auto-
regressions (linear and non-linear) to closely and regularly spaced data is studied extensively
in the literature. For the linear case Jones (1981) and Bergstrom (1985) used state-space
representations to compute exact maximum likelihood estimators, and Phillips (1959) did
so by fitting an appropriate discrete-time ARMA process to the data. This method is also
discussed in Brockwell (1995). According to Härdle & Cabrera (2012), the coefficients of
a continuous AR model can be estimated based on those of the corresponding discrete AR
models by matching coefficients. Brockwell et al. (2007) use exact conditional maximum
likelihood estimators for the continuously-observed process to derive approximate maximum
likelihood estimators based on closely-spaced discrete observations. Brockwell et al. (2006)
further introduce the continuous time COGARCH\((p, q)\) model which is found to have the
autocorrelation function of a continuous-time autoregressive moving average process.

Aside from calibration issues, the stability and convertibility condition between discrete
and continuous ARMA models are also discussed. According to Williams & Rasmussen
(2006) and Koralov & Sinai (2007), the conversion between discrete ARMA models and
continuous ARMA processes is not trivial in general. When $p \leq q$ the covariance function is not well defined based on rational spectral density. As a result, the CARMA($p, q$) process was first defined under the condition $p > q$. However, the CARMA($p, q$) process can be generalized to the $p \leq q$ case using the concept of generalized random process, according to more recent research, Brockwell & Hannig (2010).

For the application of CARMA($p, q$) processes in pricing, Benth et al. (2010) present a multi-factor continuous-time autoregressive moving-average (CARMA) model for the short and forward interest rates. The model is calibrated to a panel of spot rates and the empirical volatility of forward rates simultaneously via a Kalman filter. Direct calibration implies a pricing measure which is consistent with both the spot rate dynamics and forward rate volatility structure. Espen Benth et al. (2012) present a new model for the electricity spot price dynamics, in which authors fit the non-stationary trend using futures data with long delay until delivery, and a robust L1-filter to find the states of CARMA($p, q$) process.

The processes above are continuous and under such models there is no jump in value of asset prices. To model the possible discontinuous changes in prices, more advanced models such as the Levy process are needed. Consider a finite time horizon $t \in [0, T]$ and let $Q$ be a probability measure equivalent to $P$. Let $Z_t$ be the density process of the Random Nikodyn derivative so that:

$$Z_t = \frac{dQ}{dP}|_{\mathcal{F}_t}$$

The Levy - Kintchine decomposition of the process provides 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 $C_t$. $N$ 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

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

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 \([0, \infty) \times \mathbb{R} \times \mathbb{W}\) with \( X_{1(t,z,q)}X_{2(t,z,q)} = 0 \), satisfying \( \int_0^t \int_{\mathbb{R}\setminus\{0\}} |X_{1(s,z)}|l(ds,dz) < \infty \) and \( \int_0^t \int_{\mathbb{R}\setminus\{0\}} |X_{2(s,z)}|\mathcal{N}(ds,dz) < \infty \), a.s.

As introduced in Todorov (2007), the 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 \( \mu(\mathbb{R}_d^0) < \infty \), and it is of infinite activity when \( \mu(\mathbb{R}_d^0) = \infty \). 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 \( \int_{|y| < 1} |y|\mu(dy) < \infty \). Brockwell & Lindner (2009) defines the uniqueness and stability of levy driven CARMA(\( p, q \)) process, which can model both the the auto correlation and jump in a continuous fashion.

In summary, the need for modeling travel time distributions and related value pricing schemes requires innovative modeling methods. To provide a solid theoretical basis for related research, this chapter reviewed fundamental theories in transportation network modeling and financial engineering. Among related models, copulas and asset pricing theory, which were developed in financial engineering research, have been identified as suitable candidates for such modeling efforts. By way of demonstration, copulas are used for modeling the dependent structure between travel times of different links and estimating travel time distributions in the next chapter.
Chapter 3

Link Travel Time Estimation and Routing Decisioning through Copula Methods

Building upon the literature review in Chapter 2, this chapter develops copula models for the representation and computation of travel time distributions across individual links of a transportation network. The description of travel time data is discussed first in Section 3.1. Copula models are defined and a two step maximum likelihood estimator is used to characterize the copula parameters in Section 3.2: First, parametric and non-parametric estimators are compared to model distributions of link travel time. Second, conditional copulas are used to model the dependent structure of the distribution of the travel time of different links. Once these conditional link travel time distributions are estimated, transportation risk measures are introduced that enable the development of risk-based routing decisions in Section 3.3. These measures define the basic route choice framework for this thesis. Numerical examples are employed to demonstrate the procedure in Section 3.4.

Two further enhancements are introduced based on the framework. First, a method for
more reliable prediction of travel time distributions is generated by aggregating the predictions from adjacent link sets. Second, a method is developed by which similarity principles are used to compute copulas for intersections lacking sufficient data. This method enables the computation of these copulas from copulas of similar intersections having sufficient data. These enhancements reveal the potential of the developed methods in practical transportation applications.

3.1 Profile description of link travel time data

The analysis conducted in this thesis utilizes floating data sources, which tend to be comprised of GPS data that is captured and archived by turn-by-turn navigation systems over long time periods. These data have the following characteristics that should be carefully considered for proper modeling of link travel times:

1. Travel time data collected by GPS receivers tend to be unevenly distributed over time. Because 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 few in number. Therefore, it is difficult to produce consecutive travel time observations containing fixed time intervals, a data structure that is presupposed by standard time series analysis.

2. The data set is composed of the travel times of individual vehicle trips. For a given path, the available data records - where each record is a vector of travel time observations for the constituent links - are equal to the number of travelers who traversed the entire path. On the other hand, one link in this path can be part of many other paths, so the number of available data records for a given link is equal to or greater than that of any path that goes through the link. To take advantage of such variability, it is preferable to employ models that can estimate the marginal link travel time distribution and the dependent structures of different links separately. This characteristic is illustrated in
Figure 3.1: Different number of observations for links and paths (In the figure, two travelers traverse Link AB and then Link BC and Link BD respectively due to different routing decisions. As a result, there are more data records corresponding to link AB(2) than any path going through it: Path ABC and Path ABD both have one data record (vector) respectively)

3. The data are unevenly distributed throughout the various links of the transportation network. There are relatively abundant observations for some links but few and sometimes none for others. For any path (sequence of links and intersections), there can be insufficient observations for a proper characterization of the underlying travel time structure over the entire path. In this chapter, the similarity between intersections are employed using copula methods, to properly characterize the travel time distribution for those links that lack sufficient observations. This then allows for the development of path travel time distributions that can be used to compare alternatives in the route choice process.

The challenges above led to the current research, which aims to produce new models to examine travel time, particularly as collected by GPS systems. Copula methods will be utilized because they offer the best fit given such a sparse and uneven data set. To relieve concerns over more general applications, the method selected for analyzing the GPS data can constructively be applied to general cases of realtime crowd-sourced data or fixed location sensor data: denser data improves the precision of the results. Corresponding theoretical properties are discussed and examples from other data sources are provided in this thesis.
The details are discussed in the following sections.

3.2 Generating travel time distributions based on copulas

A fundamental assumption imbedded in most models representing the distribution of travel times or other attributes of the links of a transportation network is the independence of the attribute from all other links in that network. If one is only interested in a deterministic measure of these attributes, this assumption is generally acceptable; however, when focused on the distribution of the attributes of individual links, the assumption of independence is more problematic. The latter case motivates the research that directly addresses the interdependence of link attributes.

There are many different ways to model the interdependence. Among existing methods, copula methods are selected in this thesis because they directly address the interdependence between different random variables. The copula framework is applied in this chapter to explicitly model the interdependence of travel time distributions across links of a transportation network, which enables the estimation of travel time distribution given the latest information.

3.2.1 Copula models and two step maximum likelihood estimation

Copula models focus on the dependence of probability distributions. Their application to link travel time distributions begins by addressing the dependence of two selected links. In theory, any two links in the transportation network can be studied together and the significance of the dependence between them usually has an inverse relationship to the distance between the two links. A bivariate case is described below to demonstrate the general modeling process.
Let \( P \) be a conditional bivariate distribution function with continuous margins \( F_X \) and \( F_Y \), and let \( \mathcal{F} \) be some conditioning set. There then exists a unique conditional copula \( C : [0, 1] \times [0, 1] \) such that

\[
P(x, y|\mathcal{F}) = C(F_X(x|\mathcal{F}), F_Y(y|\mathcal{F})|\mathcal{F}), \forall x, y \in R
\]  

(3.1)

and the conditional distribution of \( Y \) given \( X \) can be calculated as follows:

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

(3.2)

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

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

(3.3)

Employing the copula model, maximum likelihood estimators (MLEs) are used to characterize 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 \( Y \), and \( \kappa \) is the parameter for the marginal distribution of copula.

1. Initialization: The total likelihood is expressed as \( L_{XY}(\phi, \gamma, \kappa) = L_X(\phi) + L_{Y|X}(\phi, \gamma, \kappa) \)

2. Step 1: Maximize \( L_X(\phi) \) to get an estimation \( \hat{\phi} \) of \( \phi \).

3. Step 2: 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 further deconstructed in the following steps.
1. Initialization: The total likelihood is expressed as \( L_{Y|X}(\phi_0, \gamma, \kappa) = L_Y(\gamma) + L_c(\phi_0, \gamma, \kappa) \).

2. Step 2.1: Maximize \( L_Y(\gamma) \) first to get \( \hat{\gamma} \).

3. Step 2.2: Maximize \( L_c(\phi_0, \gamma_0, \kappa \mid \hat{\phi}, \hat{\gamma}) \) conditioned on \( \hat{\phi} \) and \( \hat{\gamma} \).

Copula parameters are estimated using a 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 asymptotically efficient, according to White (1994).

**Characterizing marginal distributions using parametric and non-parametric estimators**

In this section, both parametric and nonparametric estimators are used to characterize the marginal distribution of link travel times: Parametric estimators 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:

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

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 demonstrate quality of fit. They are defined in Table 3.1, where \( F_{emp} \) is the cumulative distribution function of the empirical travel time distribution, and \( F_{est} \) is the cumulative distribution function of the estimated travel time 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 X}</td>
</tr>
<tr>
<td>L2 statistics</td>
<td>$L2 = \sum_{x \in X} (F_{emp}(x) - F_{est}(x))^2$</td>
</tr>
<tr>
<td>Anderson-Darling statistic:</td>
<td>$AD = \max_{x \in \mathbb{R}} \frac{</td>
</tr>
</tbody>
</table>

Estimation of the copula

After estimating the marginal distribution of travel time, the copula function describing the dependence between the travel times of different links is constructed based on the uniformly distributed random variables which are transformed from the link travel time distributions using the corresponding cumulative travel time distribution functions. Please note, by plugging a random variable into its cumulative distribution function, the output is a uniformly distributed random variable. Each link travel time observation is transformed to a value between 0 and 1 in this way, then the uniformly distributed random variables corresponding to different links show the dependent structure (copula) between the travel time of these links. The study of dependence based on copulas is therefore not subject to any difference in marginal link travel time distributions, which potentially leads to stable estimation of the dependent structure.

Travelers are interested in the co-occurrence of high travel time on the focused links, that is, the probability of experiencing increased travel time in the unobserved links given increased travel time in the observed links. In copula theory, such dependence is modeled by upper tail dependence, as described in Definitions 2.3.5 and 2.3.6. Different copulas yield different upper tail dependence measures, which may describe different likelihoods of experiencing increased travel time on the unknown link given the same experience on the known link. Below, alternative copulas are proposed and compared according to their upper tail dependence in order to illustrate such differences.

The copulas compared in this chapter include Bivariate Gaussian, T, Clayton, Gumbel,
Frank, and BB1. They were defined in Table 2.1 in Chapter 2. Their tail dependence is given below in Table 3.2. According to theoretical results, there is no upper or lower dependence in the Gaussian copula or Frank copula, i.e., when two random variables which depend on each other through Gaussian copula both have extreme values, these extreme values tend to distribute independently. There is only lower tail dependence for Clayton copulas and only upper tail dependence for Gumbel copulas, i.e., the extreme low values distribute in a correlated fashion in Clayton copulas; the extreme high values distribute in a correlated fashion in Gumbel copulas. The upper and lower tail dependence for T copula and BB1 are non zero: in the case of T copulas, both tails are the same and nonzero; in the case of BB1 copulas, they can be different. The additional freedom in the BB1 copulas may lead to better fitting over empirical data.

<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>$\lambda_L = 2t_{\mu+1}(\sqrt{\rho + 1} - 1)$</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>

To fully capture the dependence between extreme travel times, copulas with both upper and lower dependence are preferable to others. In practice, the empirical tail dependence described in Definition 2.3.7 is compared to the theoretical tail dependence of typical copulas in order to select the best fit. Estimation of marginal distributions and copulas allows the travel time distribution of the dependent links to be generated, which enables routing decisions based on estimated travel time distributions.
3.3 Routing decisions based on estimated travel time distributions

Once the travel time distributions have been estimated, risk measures can be used to recommend the best route choice. In finance, risk measures are used to determine the amount of an asset or a set of assets that should be reserved for potential losses. In a transportation context, risk measures can be defined as follows:

**Definition 3.3.1** A 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. \( \forall a \in \mathbb{R}, \rho(a + X) = \rho(X) + f(a) \): 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 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 terms of mean-variance, then that link is preferred by travelers when making 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 terms of transferred mean-variance, then that link is preferred by travelers when making routing choices.

2. Rule 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 terms of FSD, then that link is preferred by travelers when making 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
\[ X \succeq_{SSD} Y \iff \int_l^u P(X > t)dt \geq \int_l^u P(Y > t)dt, \forall a \in [l, u] \]

If a candidate link is dominated by other candidate links in terms of SSD, then that link is preferred by travelers when making 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 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
\[ X \succeq_{AR} Y \iff \frac{\int_t^\infty F_A(x)dx}{\int_t^\infty F_B(x)dx} \geq \epsilon \]

If a candidate link is dominating other candidate links in terms of the area ratio rule, then that link is preferred by travelers when making routing choices.

As this rule was first constructed based on comparison of CDF curves, no direct interpretation in the transportation context was given, which may lead to difficulties when trying to justify the decisioning rationale based on this rule. This thesis contributes to the topic by introducing Theorem 3.3.3, which presents the probabilistic framework.
behind the area ratio rule and links it to the differing capacities of the two transportation links to allow traffic to pass in the same time period. This interpretation fills the previous gap between mathematical and transportation-based interpretations of rules.

4. Rule 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 travelers. $1/a$ is the value of travel time that the traveler will give a utility value of $\exp(-1)$. Then the random variable which yields a larger expected utility value dominates the other. That is

$$X \succeq_{EU} Y \iff E\exp(-aX) \leq E\exp(-aY)$$

If a candidate link is dominating other candidate links in terms of the expected exponential utility rule, then that link is preferred by travelers when making routing choices.

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

$$X \succeq_{VAR} Y \iff Q^\alpha_X \leq Q^\alpha_Y$$

If a candidate link is dominating other candidate links in terms of the value-at-risk rule, then that link is preferred by travelers when making routing choices.

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

**Theorem 3.3.3** Suppose travelers enter a path according to a Poisson process $\lambda$, and the distribution of travel time is $F$, then according to the properties of the Poisson process:

$$P(M_t = k) \sim \text{Poisson}(\lambda'),$$

where $M_t$ is the number of travelers that traverse the path completely by time $t$. Here, the $\lambda' = \lambda \cdot \frac{1}{t} \int_0^t F(s)ds$

Proof:

Denote the number of travelers entering the path by time $t$ as $N_t$. It is assumed that the arrival event of a traveler and his travel time on the link are independent.

First, the distribution of travel time conditioned on the number of travelers that entered the route at time $t$ can be calculated, which is denoted as $N_t$, then conditioned on the entering time $T_i$ is

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

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

Then since the behavior of travelers is independent, the occurrence of an event where a traveler 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 \( n \) travelers traverse the path completely within time \( t \) is subject to a binomial distribution with the following parameters:

\[P(M_t = k|N_t = n) = C^k_n P_t^k \ast (1 - P_t)^{n-k}\]
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 travelers 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 travelers 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 rule between two paths can be calculated as:

\[
\rho = \frac{\int_{t_l}^{t_f} F_p(s) \, dt}{\int_{t_l}^{t_f} F_q(s) \, dt}
\]

\[
= \frac{\int_{t_l}^{t_f} F_p(s) \frac{1}{t_l} \, dt}{\int_{t_l}^{t_f} F_q(s) \frac{1}{t_l} \, dt}
\]

\[
= \frac{P_p(T_{ps} + T_i <= t|N_t = n)}{P_q(T_{ps} + T_i <= t|N_t = n)}
\]

\[
= \frac{\lambda * P_{p-t_l} * t_l}{\lambda * P_{q-t_l} * t_l}
\]

\[
= \frac{E(X_{p-t_l})}{E(X_{q-t_l})}
\]

Hence, the area ratio signifies the ratio of the expected number of cars completing the route by time \( t_l \) when travel time on the paths is subject to two different travel time distributions, respectively. It measures the distinct capacity of the two paths to allow traffic to pass in the same time period. Hence, the area ratio rule implies that if the expected road capacity of a candidate link is larger than the capacity of others, 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 experimental 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 travelers with installed CoPilot turn-by-turn GPS navigation systems produced by [http://www.alk.com](http://www.alk.com) (2009). The test area is shown in 3.2, and the targeting monument links are shown in Table [3.2](#).
3.3 Travel time is measured in seconds. As discussed in the introduction, monuments are defined as reference points between which travel times are measured. Usually, a monument is located at the midpoint of a physical link. A "m2m" link is defined as the shortest path on the underlying physical network between two given monuments.

![Figure 3.2: Test network N95-US 195, the red frames are the targeted monuments (Monument points A,B,C,D,F,G are defined. The links between these points are highlighted to represent the paths of interest in the analysis of this section)](image)

<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 sources of vehicle position-time data the method used herein is captured in real-time by an on-board GPS turn-by-turn navigation system. Each vehicle-centric data record contains the sequence of position-time values experienced by the vehicle in its travels. Time values are strictly monotonic and each position can readily be map-matched to a location along a link of the corresponding digital map. It is straightforward to interpolate the data sequence to obtain a rather precise value of the "exact" time that the vehicle passed any monument. Passage of adjacent monuments by a unique vehicle provides an observation of the travel time for that m2m pair. More specifically, this value is obtained by:

(a) Map-matching the GPS points to appropriate links.
(b) Along a given recorded trip, find the monuments through which the traveler traversed. For each monument, record the two closest GPS locations to the monument: One is upstream from the monument, and the other is downstream from the monument. Denote the time the traveler arrived at the upstream location as \((T_l, L_l)\), and denote the arrival time 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 obtain 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 that usually covers multiple m2m links, all possible sub-trips through this path are enumerated. The travel time for each sub-trip is recorded by taking the difference of the \(T_m\) corresponding to the starting monument and ending monuments of the sub-trip. The procedures above build the data set that is used to validate the theory.
3. Experimental design.

Based on the experimental network displayed in Figure 3.2, two simple topologies are extracted in Figure 3.3. Suppose:

(a) A traveler starts at Monument A and needs to go to Monument X via either AB-BX or AF-FX;

(b) The travel time on the link BX and FX is equal and deterministic. In this context, a comparison of travel time distributions between AB and AF is sufficient to yield a routing decision.

(c) Without loss of generosity, it is assumed the travel time distribution on AF is known and copula models are used to estimate the unknown travel time distributions on AB.

(d) There exists partial observation of the network: there is no new travel time observation on the link AB, but there is a new travel time observation on a nearby link BE. Based on the four assumptions above, the experiment demonstrates how the travel time distribution on AB can be estimated considering the most recent
observation on link BE. This procedure can be applied to any pair of links of interest in practical road networks. To support such a decision, the conditional travel time density on AB is first estimated using the travel time observations on link BE, BC and CD by considering the link sets AB-BE (on U.S. 195) and AB-BC-CD (on Interstate 95).

Then, the estimated distribution for AB conditioned on the latest travel time observations on other links is compared with the latest travel time observations on AB as validation of that particular estimate. For out-of-sample test, a leave-one-out cross-validation is then conducted. Finally, the routing decision is generated by comparing the estimated travel time distribution for AB with the smoothed empirical travel time distribution on link AF. Furthermore, travel time is estimated for another link set HG-GK based on its similarity to the link set AB-AE, which illustrates the procedure of estimating travel time distributions based on the similarity between two intersections.

4. Data and outliers.

In Pattanamekar et al. (2003), two methods are used to remove outliers in a traffic data set: 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, method 2) is selected, i.e. extreme values beyond $\mu + 4\sigma$ are excluded from the analysis. The rationale for this modeling choice is that travel time is calculated by taking the difference of the arrival times to two monuments in the GPS data set so that a large travel time value results. This travel time value is based on several factors, including (a) congestion between the two monuments, (b) measurement errors, and (c) traveler’s non-traffic related stops between monuments. The occurrence of unreasonably extreme values is usually caused by reason (b) or (c), which should be removed to avoid estimation bias.

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 3.4.

![Figure 3.4: Fit travel time distribution on AB](image)

To select the best parametrical distribution, the distance measure between the empirical distribution $F_{emp}$ and the estimated distribution $F_{est}$ are calculated and displayed in Table 5.7

<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>≈0</td>
</tr>
<tr>
<td>Weibull</td>
<td>0.0055</td>
<td>169.7458</td>
<td>43.1215</td>
<td>329.6703</td>
<td>≈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>≈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 and the kernel smoothing method yields the best fit for marginal travel time distributions. The
Kolmogorov-Smirnov test (K-S test) is a nonparametric test for the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution (one-sample K-S test). A small p-value for the KS test suggests that the null hypothesis that the travel time is from the specified distributions can be rejected.

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 3.5, and the corresponding AIC criteria values and the parametric bootstrap test for Archimedean copulas are shown in Table 3.5. Here, a greater AIC value or a larger p-value implies that the selected copula fits the data better: the BB1 copula fits the data best of all the candidate copulas according to these two criteria.

<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</td>
<td>-</td>
<td>-</td>
<td>0.8110</td>
<td>0.1308</td>
<td>0.0058</td>
<td>0.0350</td>
</tr>
</tbody>
</table>

The theoretical properties of different copulas are displayed in Table 3.2; fitting statistics are compared in Table 3.5 and the empirical travel time data is compared with simulated data based on different copulas in Figure 3.5. 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 and its theoretical tail dependence is zero, therefore the appropriate fitting statistics are not satisfactory, and simulated joint Gaussian variables do not distribute similarly to empirical data. However, given that various estimators can be used to construct the parameter matrix for a Gaussian copula,
Figure 3.5: Empirical distribution v.s. simulated data based on different copulas
and given the relative ease of simulating a Gaussian copula, it is efficient in high dimensional cases, which is the topic in Chapter 4.

(b) A T copula fits the data better than a Gaussian copula, as it has both high and low tail dependence. However, for T copulas, both upper-tail and lower-tail dependence is the same, which can be found in Table 3.2 and Figure 3.5. 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 copulas may lack a certain flexibility when used to describe travel time dependence.

(c) A Frank copula does not reveal the upper or lower dependence relationships per the zero theoretical values in Table 3.2 and the fitting results also show differences from empirical data.

(d) A Clayton copula only reveals the lower tail dependence i.e. 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, it is not particularly useful for modeling travel time dependence.

(e) A Gumbel copula only shows upper tail dependence i.e. the occurrence of high travel time on different links. It can hence be a reasonably simple model, compared to the BB1 family, because the occurrence of high travel time is of more concern in transportation research compared to low travel time.

(f) A BB1 copula 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 2.2.7, the two tail indices for
this link set are calculated. Results are shown in Figure 3.6:

![Figure 3.6: The comparison of tails of the joint structure. Red: lower tail; Blue: upper tail](image)

Figure 3.6 shows that the shape of lower tail and upper tail of the empirical travel time distribution is not symmetrical about the center, which implies that the two tails of the copula between link travel times are asymmetric in nature, with the upper tail heavier than the lower tail. Based on this observation, it is preferable to use a copula with asymmetric tail structures to describe this difference. Therefore, the BB1 copula is the best choice of all the candidate copulas. In this chapter, the BB1 copula is selected for the analysis of two dimensional cases. For higher dimensional cases addressed in the next chapter, a Gaussian copula mixed model is used to approximate the empirical dependent structure between the travel times of several links, which may be both flexible in describing the dependence structure and efficient in estimation.

3. Copula parameter estimation

Using the two-step estimator introduced in Section 3.2.1, the parameters of the BB1 copula between links AB and BE are estimated. The parameters are $\theta = 0.7634$ and $\delta = 1.5154$, which implies the corresponding tail dependence parameters are $\Lambda_U = 0.4201$ and $\Lambda_L = 0.5493$. The results demonstrate that there is a level of tail dependence
in the travel time data which is generally more significant than that implied by the joint normal distribution, which is 0. Furthermore, the implied dependence structure is asymmetric. In other words, the probability of the simultaneous occurrence of low travel times on both links (approximately 0.4201) is different from that of simultaneous high travel times (approximately 0.5493).

4. Routing decisions based on the estimated copula

After copula parameters are estimated, conditional travel time distributions on AB can be calculated using formulas defined in Section 3.2. The estimated copulas 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 3.7.

The estimated mean travel time of AB is \( \mu_{AB} = 258.9 \) seconds, given that the most recent travel time observation on BE is 317 seconds. The estimated standard deviation is \( \sigma_{AB} = 48.07 \). On the other hand, the observed data when the observation on BE is 317 is 251 seconds, which is quite close to the estimated mean travel time \( \mu_{AB} \). Note this mean travel time is not the overall mean value but the specific expected travel time conditioned on the latest traffic conditions on nearby links, which depends on the estimated dependent structure between the travel time on AB and BE.

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

Routing decisions are generated based on the estimated travel time distribution of alternative links, following the settings specified in the experimental design part of this section. 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; under the expected exponential utility rule with parameter \(-1/250\); under the Value-at-risk rule with 5%
Figure 3.7: Estimation of the copula and conditional probability function of AB based on BB1 copula

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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exponential utility</th>
<th>$a = -1/250$</th>
<th>$t_U = 1500$</th>
<th>VAR(Quantile)</th>
<th>5%</th>
</tr>
</thead>
<tbody>
<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>

upper tail. AF is preferred to AB under the area-ratio rule 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. As a result, travelers will choose different routes according to different decision rules, and this flexibility is a promising improvement for current routing guidance service.

5. While the estimated result above fits the real observation, it is arguable that one point comparison may lack statistical significance. Therefore, a leave-one-out cross-validation is conducted to show the difference between the estimated mean travel time and the real observation. In this cross validation experiment, the data set is separated into one
calibration set and one test set. The copula estimated on the calibration data set is used to estimate the unknown travel time distribution based on the test data set. 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 estimate of the conditional mean, which shows that the BB1 copula fits the empirical dependent structure better than a Normal copula. The conditional distribution is then used to generate routing decisions by calculating the following decision rules, as in Table 3.6.

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

6. Validate the estimation of conditional distribution using an alternative data set from loop detectors.

Up to now, the validation is based on the comparison between mean travel time and observed travel time, but due to limited data, it is not possible to compare the estimated travel time distribution with the observed travel time distribution given the latest traffic status. To partially serve this purpose, a denser data set is used to conduct a similar experiment. In order to demonstrate the application of copula methods to other travel time related measures, an experiment is conducted on a 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 using the methods in this section.

(b) To evaluate the performance of the estimator, the estimated distribution is compared with an empirical conditional distribution. The empirical version is con-
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 in Figure 3.8. The calculated density yields a reasonably good estimate, as shown in Table 3.8.

Table 3.8: Square distance between estimated speed distribution and empirical speed distribution

<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>

3.5 Further development: reliable estimates and similarity-based analysis

Due to the high dimensionality of the road network and very limited GPS device coverage, to estimate the link travel time distribution via copula approaches may still suffer from two limitations: 1) The number of links which should be studied together is high and a high dimensional dependent structure should be estimated; 2) The data is not sufficient to estimate the copula for some intersections. In this section, two simple practical solutions are proposed as applications of the model in the previous section.

3.5.1 Generation of reliable estimation

In practice, there is always a trade off between dimensionality and availability of data. Ideally, in estimating a link travel time on a link A, one can gather all travel time data from
Figure 3.8: Estimated conditional density (Red) v.s. Empirical conditional density (Histogram: Blue; Kernel smoothed distribution: Black) given current observation of 67 mile/hour per hour at a loop detector.

all nearby links and construct data matrixes to estimate copula parameters. In practice, however, it is harder to construct such matrixes for a path with more links because finding travelers to travel on a larger number of links at the same time (or subject to other matching conditions) is simply harder. A practical method is to break down this high dimensional copula by examining at the link sets A B C, ADE and AFG separately. Additionally, three
estimations for the travel time on AB are generated separately and finally a weighted average of the three is taken as the primary estimates, which can serve as a reasonable approximation. The final estimation is a weighted sum of these estimates. Mathematically, this weighted scheme is formulated as follows:

\[
P(t_b < x) = \sum_i f_{b|i}(x) \alpha_i
\]

(3.5)

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

To illustrate this weighted 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, the T copula is selected as the multidimensional copula to conduct an 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.

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

Table 3.9: Parameter estimation for the T copula

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

\[
p(t_{AB} < x) = f_{AB|BE}(x) \alpha_1 + f_{AB|BC,CD}(x) \alpha_2
\]
The parameter $\alpha$ is the weight given to the individual estimation corresponding to a given link set. 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 based on T copula and the three-link path AB-BC-CD is shown in Figure 3.10. Then, a combined estimated conditional travel time distribution is obtained by calculating a weighted average between the T and BB1 estimators. Decision statistics for the two estimated distributions (Based on T and Combined) are calculated in Table 3.10.

![Figure 3.9: Combined conditional probability density function(Green) compared to two original estimations using a T copula (Red) and a BB1 copula (Blue).](image)

<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>

If travelers wish to consider the traffic conditions on many links simultaneously, then they need to study the dependence among them. However, due to data limitations, it is
usually difficult 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 distinct, relatively low-dimensional link sets considers the different dependent
structures between the link and its neighbors. The more data and links used to generate the
estimation, the more reliable the estimated travel time distribution in the focused link. In
this way, travelers can make trade-offs between accuracy and model complexity.

3.5.2 Similarity-based copula reconstruction

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

**Assumption 3.5.1** Factor Determination (FD): There are similar travel time dependence
structures 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 topological 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 in future research.
Below, a numerical example is provided to illustrate how to generate travel time estimations for a blind intersection using the copulas calibrated in a similar intersection.

Suppose 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 $\theta = 4.10E - 7$ 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 the AB-BE pair is shown in Figure 3.10.

![Estimated Copula](image1)

![Estimated Pdf](image2)

**Figure 3.10:** Differences 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 are displayed in Figure 3.10. Based on the estimated model, 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. The estimation is based on the similarity in upper tails despite the differences in lower tails of the two travel time distributions. This suggests that asymmetric fitting can address travelers’ concern about high travel times while reducing the requirement for similar conditions, hence leading to potentially less problematic real world applications.

In summary, the research in this chapter proposes a framework to forecast the travel time distribution for a given link. Copula methods are selected, as they can effectively model sparse GPS travel time data and marginal distribution and dependent structure separately. Moreover, copula parameters estimated in one link set can be used for other link sets 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 leading to his final destination. To address this issue, multiple scenarios of the multiple links in a given path should be studied; conditioned on the different levels of the overall traffic conditions along the path, a Gaussian copula is used to estimate the dependent structure. Therefore, multiple Gaussian copulas are used to describe the complex dependent structures of the travel times of multiple links, which demonstrates great potential in travel time estimation.
Chapter 4

Path Travel Time Distribution

Estimation through Gaussian Copula Mixture Models

This chapter proposes procedures which are used to compute path travel time distributions from non-independent travel time distributions on individual links. To facilitate this Gaussian copula mixture models (GCMM) are developed to model the interdependence of travel time distributions among neighboring links on a given path so that individual travel time distributions can be assembled. Path travel time distributions are then generated through simulation using the estimated dependent structure.

The chapter begins with a fundamental description of the challenges encountered in trying to estimate the time to traverse a path and how such estimates can be used to help choose a path among a set of several or many feasible near optimal alternatives. The path scenario is introduced as a state representation of the path traffic status in which the overall travel time distribution is stationary. Such path scenarios can either be observable and pre-defined or unobservable and in need of extraction from the data.
An observable-scenario-based approach to compute a path travel time distribution is developed in Chapter 4.2. In this model, travel time is assumed to be a function of the systematic traffic conditions along the path (the path scenario). A set of observable path scenarios is pre-defined and travel time observations are classified in each scenario before conducting parameter estimation. In each path scenario, a Gaussian copula is used to model the dependence between the constructing links, and the conditional path travel time distribution is estimated through the Lasso method based on the travel time observations that correspond to each scenario, Meinshausen & Bühlmann (2006). The overall path travel time distribution is then constructed by summing the conditional path travel time distributions over the set of path scenarios using appropriate weights.

In Section 4.3, an unobservable-scenario-based approach is developed in which path scenarios are not observable but can be extracted through an iterative learning algorithm from the data. Two versions of expectation maximum methods are designed to fit the model to the empirical data. This methodology enables more flexible application of GCMMs by relaxing the need to pre-identify path scenarios and to categorize empirical data into these pre-defined path scenarios. As an extension to Gaussian Mixture models, GCMMs can further reduce number of components by employing more flexible components and utilizing unsynchronized data in each dimension.

In the process of estimating these models, pseudo path travel time observations are introduced to address the problem of data insufficiency resulting from the scarcity of travel time observations and the high dimensionality of the problem (i.e., typically, many links assemble to form a path): these additional data vectors are constructed using the travel time experienced by different travelers who enter and exit the path at the same location at about the same time so that more path travel time observations can be obtained. The theoretical properties of the GCMM estimator with such a data synchronization technique are studied.
4.1 Estimating travel time using scenario-based GCMMs

This section describes the challenges encountered in trying to estimate the path travel time distribution and the important interdependence between the travel times of different links. To tackle such a challenge, the chapter develops the path scenario based framework and utilizes Gaussian copula mixture models (GCMM). Such path scenarios are observable in Section 4.2 and are unobservable in Section 4.3, and the differences in observability lead to different estimation methodologies. To link statistical estimation with practical applications, this section also delineates the procedures to generate a path travel time distribution from an estimated GCMMs.

4.1.1 Background

Travelers can traverse a given monument path under different conditions where various factors change their experienced travel time, leading to the empirical pattern of path travel time distribution. It would be naive to assume stationary conditions and, furthermore, Gaussian conditions hold for the overall travel time distribution. A more practical assumption may be the path travel time distribution that is stationary under more specific conditions. For example, if the roads are congested, a given traveler’s travel time between A and B may be subject to a distribution; if the traffic is good, a traveler’s travel time between A and B may be subject to another distribution. It is assumed that the probability of being good v.s. bad can be estimated so that the overall expected travel time distribution is a weighted average of the two cases.

Formally, the term path scenario is introduced to describe the different traffic conditions a traveler may encounter during his trip. The fundamental assumption is that a traveler may encounter different path scenarios on a trip and the probabilities of path scenarios encountered on the trip are used as the weights to aggregate scenario-specific path travel
time distributions for the overall path travel time distributions.

A path scenario, that describes the overall traffic conditions of a given route, can be defined as a specific state in which the travel time for sections along the route is subject to a specific joint distribution. In a path scenario, the marginal link travel time distributions are all stationary with respect to the traffic conditions implied by the path scenario, and the copula structure is stationary as well. In mathematical terms, a path scenario specifies the type and parameters of the following items:

- The travel time distribution in the links of the path: \( \{T_{t_1}^s, \ldots, T_{t_I}^s\} \);
- The dependent structure between these link travel times. This dependent structure is described by a copula \( C^s \).

Based on the definition above, a path scenario represents the traffic conditions along a given path, and therefore constitutes the basic unit for statistical description of the path travel time distribution. To justify this definition, one needs to understand the characteristics of the path scenario. To address this aim, the internal characterization and external determination of these path scenarios will be discussed below.

Internally, the joint travel time distribution between links is characterized by a range of arrival times to the links in the path. Because travelers experience different travel times as they arrive at links at different times, the experienced travel time is usually contingent upon the arrival time to a link. Following this logic, the joint travel time distribution between different links should be associated with the sequence of arrival times to these links.

In the GCMM based on observable path scenarios, a strict partition of the space of arrival times is taken to characterize the joint travel time distribution: It is assumed that a sequence of arrival times to these links, each of which is within a given range, should characterize the joint distribution of experienced travel times of the traveler. In mathematical terms, it is therefore assumed that path scenarios are specified by the partition of arrival times to the
links in the path, where a partition of the space of arrival times is composed of a group of subsets of this space, and: 1) it does not contain the empty set; 2), the union of these subsets is equivalent to the volume of the space; and 3), the intersection of any two distinct subsets in this group is empty. The joint travel time distribution is assumed to be stationary in each subset in this partition. To illustrate, consider two consecutive links AB and BC: if a traveler Jack enters Link AB around 10 am and enters BC around 10:10 am, the joint distribution of the travel times of AB and BC for the arrival time sequence (10:00,10:10) should be different from that if Jack enters AB around 10 am and enters BC around 10:20 am, i.e., the arrival time sequence (10:00,10:20). The additional delay on Link AB may characterize the changes to the traffic status and hence to the joint distribution between travel times of AB and BC. In other words, it makes sense to assume: if the arrival time to AB is around 10:00 and that to BC is between 10:02 (suppose the lower limit to the travel time on AB is 2 minutes) and 10:15, then the joint travel time distribution is stationary. If the arrival time to AB is around 10:00 and that to BC is after 10:15, the joint travel time distribution is stationary but different, which implies more delays. The travel time observation with arrival time sequence (10:00,10:10) is used to estimate the travel time distribution in the first path scenario while the alternative travel time observation with arrival time sequence (10:00,10:20) is used to estimate the travel time distribution in the second path scenario.

In the GCMM based on unobservable path scenarios, a probabilistic characterization is constructed: It is assumed that path scenarios can not be observed but can be discovered through studying the travel time observations that comprise them. There are no deterministic boundaries between two joint travel time distributions and one data point can belong to many path scenarios in terms of probability. Following the example above, the travel time observation with arrival time sequence (10:00,10:10) is assumed to belong to the two different path scenarios with different probabilities $p$ and $q$, as the two path scenarios are not observable. The characteristics of the two unobservable path scenarios and the two probability values are estimated using machine learning methods.
Externally, the attributes and number of possible path scenarios are determined by a set of factors, including time of day when the trip starts, weather conditions and the presence of any special events occurring along the path. The first example trip is defined as starting around 10:00 am, as illustrated in the previous paragraph. And then several sensitivity scenarios can be discussed.

Suppose that the trip starts around 8:00am. Because it occurs during rush hour, the two joint distributions for travel time on AB and BC can be characterized differently. If the arrival time at AB is around 8:00 and the arrival time at BC is between 8:02 and 8:10—relative to the 15 minutes difference in the first example trip in previous paragraphs, then the joint travel time distribution is stationary. If the arrival time to AB is around 8:00 and that to BC is after 8:10, then the joint travel time distribution is stationary but different, which implies more delays. If it is raining on a specific day, then the intervening time of the two joint travel time distributions AB and BC may be 15 minutes instead of 10 minutes. If there is a parade on the road around 10:10 that day, the intervening time of the two joint travel time distributions on AB and BC may also be around 10:15 instead of 10:10. If the overall volatility of traffic, weather and event schedules increases, then more path scenarios should be used to model the travel time distribution on the path AB-BC. In each example noted above, the parameters of the GCMM change according to the dynamics of external factors. In a probability characterization, the mean and variance of the travel time in each path scenario will change along with external factors. In this chapter, the estimation of path travel time distributions in light of such external factors is the focus: in the first half of the chapter, it is assumed that path scenarios can be clearly defined through experience; in the second half of the chapter, it is assumed that path scenarios are not explicitly defined but can be extracted from data through machine learning methods. The identification of the external factors and related factor models can be studied in subsequent research.

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 es-
timator assembles the unknown high dimensional joint distribution in parts. Comparatively, traditional recursive convolution of link travel time distributions cannot be calculated, since the joint distribution of link travel times in a path is unknown without effective statistical estimators.

4.1.2 Observable and unobservable path scenarios

Path scenarios can be observable. The observability here refers to the fact that, in most cases, people can evaluate the path traffic status through general sense or knowledge of traffic engineering. Through experience, modelers can categorize observed travel time data into predefined path scenarios. Additionally, the link travel time distributions and the copula that describes the dependence between these link travel times in each path scenario can be estimated using the data categorized in that path scenario. The model building is conducted in a straightforward manner, following the process of scenario identification, data categorization, link travel time distribution estimation and copula estimation. In this process, pseudo path travel time observations may be constructed to populate the data vector while Lasso methods are used to estimate copula parameters employed to tackle the problem of data insufficiency.

Alternatively, path scenarios can also be assumed to be unobservable and observers can only see the traffic data revealing its pattern, following the intuition of hidden factor models, as in Bai & Li (2012b). Due to such an assumption, unobservable path scenarios can therefore only be extracted from the data using a learning algorithm instead of being estimated directly using more explicit procedures. Following this intuition, expectation maximum algorithms are designed to extract the parameters of path scenarios from the data. In such unobservable path scenarios, a data vector can belong to different path scenarios with probabilities and path scenarios evolving in the extraction process. Such flexibility leads to potentially better model fitting. However, certain regulatory conditions should be imposed for the parameters
to be identifiable: by imposing the two-step estimation procedure of copulas and marginal distributions, the parameterization of both parts are orthogonal to each other, which leads to good model identifiability between marginal travel time distributions and copula structure. The expectation maximum algorithm further identifies the parameters of multiple copulas in the model. As a result, such parameterization leads to successful extraction of the hidden scenarios of path travel time conditions.

4.1.3 Definition of a Gaussian copula mixture model (GCMM)

To formally characterize the scenarios used to estimate travel time distributions, a Gaussian copula mixture model (GCMM) is introduced that consists of a weighted sum of a finite number of joint distributions whose dependent structures are modeled by Gaussian copulas. It is a generalization of the usual a Gaussian Mixture Model (GMM). When the marginal distributions are restricted to be Gaussian as well, the model reduces to a GMM. To begin, the multivariate Gaussian copula is defined by the following probability function:

\[
F(u|P) = \int_{-\infty}^{\Psi^{-1}(u_1)} \ldots \int_{-\infty}^{\Psi^{-1}(u_d)} \frac{1}{(2\pi)^{n/2}P^{1/2}} e^{\exp(-\frac{1}{2}v^TP^{-1}v)}dv
\]

(4.1)

whose density is given by

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

(4.2)

where

\[
\Psi \text{ is the one dimensional cumulative distribution function for a standard normal distribution with density } \psi;
\]

\[
P \text{ is the copula parameter matrix;}
\]
d is the number of dimension.

Notice that the parameter matrix $P$ is the correlation matrix if the random variables are normally distributed as in Nelsen (2006). Then, with the Gaussianization of original travel time observations, a GCMM for the joint distribution of a random vector $X$ can be defined as follows:

$$F(X|\pi) = \sum_{k=1}^{K} \pi_k \int_{-\infty}^{Y_{k1}} \ldots \int_{-\infty}^{Y_{kd}} \frac{1}{(2\pi)^{n/2}P_k^{1/2}} \exp(-\frac{1}{2}Y^T P_k Y) dY$$  \hspace{1cm} (4.3)$$

where

$x = [x_1 \ldots x_d]$ is the marginal observation;

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

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

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

$\pi_k$ is the weight to the k-th copula.

Its density is given by

$$f(X|\pi) = \sum_{k=1}^{K} \pi_k \frac{1}{(2\pi)^{n/2}P_k^{1/2}} \exp(-\frac{1}{2}Y_k^T P_k Y_k) \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(Y_{kd})^2)$$  \hspace{1cm} (4.4)$$

The estimation of a GCMM is discussed in detail in later sections. After a GCMM is estimated based on formula (4.3) and (4.4), which describes the dependent structure of travel time between links in a given path, the path travel time distribution can be generated using the Procedure 4.1:

1. Simulate a $N$-dimensional data record from Gaussian joint distribution, given one of
the copulas with parameter $P_k$.

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

3. Sum such marginal variables together to get a sample of the path travel time corresponding to that path scenario, and the path travel time distribution is then estimated using the histogram of simulated data points.

4. Conduct Steps 1 through 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 as explored in Damelin & Miller Jr (2011) can be used to generate the travel time distribution for the path, but the simulation procedure is more computationally efficient.

**4.2 A Gaussian copula mixture model based on observable path scenarios**

In this section, a GCMM based on observable path scenarios is examined, which is defined using the following three steps:

1. A path scenario is defined as a stationary joint distribution between the travel time of the constructing links of the path. Given a path scenario, the marginal travel time distribution on each link is therefore stationary; conversely by categorizing marginal travel time observations to different path scenarios, the model parameters can be estimated. 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 represents the travel time a particular traveler takes to completely traverse the link. 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 the path 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 by assumption. The scenario-specific path travel time distribution can be estimated using this structure. This step is illustrated in Figure 4.2:

A Gaussian copula structure exists between the link distributions for a given path scenario, and the scenario-specific path travel time distribution can be estimated using this structure. This step is illustrated in Figure 4.2: in the 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. The three distribu-
tions are aggregated using a Gaussian copula and the outcome is the path travel time
distribution corresponding to the path scenario.

Figure 4.2: Estimate the path travel time distribution using the copula in each path scenario.
The path travel time distribution in this path scenario is estimated based on three link travel
time distributions using a three dimensional copula that is estimated based on the categorized
data in Step 1.

3. The overall path travel time distribution can be estimated by integrating scenario-
specific path travel time distributions employing the appropriate weights. As stated
at the beginning of this section, the fundamental assumption here is that a traveler
may encounter different path scenarios along 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. This is illustrated in Figure 4.3.
4.2.1 Scenario decomposition and summation

Based on the discussion on scenario definition in Section 4.1, the estimation process of path travel time distribution can be summarized as: To begin, different path scenarios are constructed to represent different traffic conditions that occur along the path. For each path scenario, the link travel time distributions are stationary and there is a stationary dependent structure between the travel times of the links along the path, which is modeled using a Gaussian copula. 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 using the frequency of these scenarios. In other words, the overall path travel time distribution is estimated by integrating the conditional path travel time distributions over the distribution of all possible levels of path travel time conditions. In the discrete setting, 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.
4.2.2 Estimation of path travel time distributions conditioned on a specific path scenario

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

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

2. Estimation of the Gaussian copula parameters using the Lasso method to avoid singularity in parameter estimations resulting from data insufficiency.

4.2.2.1 Construction of the data vector for one path scenario

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

The challenge posed by gathering such observations is that the number of available observations along a long path is very limited. On the other hand, more observations should be used to estimate the high dimensional copula along a longer path relative to a shorter
path. Such data insufficiency is related to how the data set was constructed through a data synchronization scheme, which is further addressed below:

1. In the original data set, path travel times are measured by short trips of travelers within a certain time limit. Due to the time limit, the number of links the travelers can traverse in a given trip is limited. As a result, if Path A contains more links than Path B, it is very likely that there are less path travel time observations along Path A than along Path B.

2. As a path observation is constructed using travel time observations for the links in the path, the number of available path travel time observations (a vector which contains a sequence of link 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 to this number.

To address the data insufficiency, 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 $p$ during 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
An original path travel time observation is the path travel time experienced by a traveler. As the data set is sparse, the number of original path travel time observations is limited. Pseudo path travel time observations are those constructed using the experienced travel time of different travelers under the condition that the exiting time of the former traveler is adequately close to the entering time of the successor, both in time and location. That is, pseudo path travel time observations are constructed by connecting the experienced link travel times of different travelers along a given path. The construction of such pseudo observations produces more path 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 where 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) has a 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., $[T^A_1, \ldots, T^A_n, T^B_{n+1}, \ldots, T^B_N]$, this data vector is then added into the observation data set, and all the original observations of $p$ are added to 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 from the link where traveler B left the path, traveler C’s

$$\pi_i = \frac{N_i}{\sum_j N_j}$$

(4.5)

$N_i$ of all path scenarios.
travel time is recorded to 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 a pseudo count includes both the original path travel time observations and the pseudo path travel time observations.

\[
\hat{\pi}_i = \frac{\hat{N}_i}{\sum_j \hat{N}_j} = \frac{N_i + \Delta_i}{\sum_j (N_j + \Delta_j)}
\]

This procedure is illustrated in Figure 4.4

Figure 4.4: Generate pseudo path travel time observations (To collect data for path a-b-c, the trips for Traveler A and Traveler B are connected together under the condition that shortly after A completes route a-b, B starts his trip on Link c).

Some related theoretical properties for such construction are studied in Section 5.3.4. It is demonstrated that given appropriate assumptions, the estimated path travel time distribution employing pseudo path observations converges to the underlying true distribution as
the number of observations is adequate.

4.2.2.2 Lasso estimation of the Gaussian copula

Although pseudo path travel time observations are collected to obtain more observations, data vectors may be still very limited for a long path and thus yield potential data insufficiency. To tackle this problem, in addition to constructing the pseudo path travel time observations, special estimation methods should also be used to obtain reliable copula parameters. In this context, the Lasso method is used to estimate the Gaussian copula parameters within each path scenario. Related background and mathematical details are discussed below:

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

\[
\Sigma = \frac{1}{n}Z^{T}Z - \frac{1}{n}Z^{T}11^{T}Z
\]

where

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

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

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:

Intuitively, this monotonic transformation transforms any non-Gaussian variable \(X\) into standard Gaussian variable \(Z\). The transformed Gaussian variable \(Z\) shares the same Gaussian copula as \(X\) so that they are joint Gaussian distributions. Thus, estimating the covariance matrix of \(Z\) using Lasso methods would yield an estimator of the
Gaussian copula parameter matrix for the original random variable $X$. 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 a Gaussian copula for non-Gaussian marginal data whose dependent structure is subject to a Gaussian copula, under the condition that monotonic transformations 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 a Gaussian copula can be obtained by converting the estimated covariance matrix into the corresponding correlation matrix.

Q.E.D.

2. As stated above, after the monotone transformation, the parameter matrix of the Gaussian copula is obtained by estimating the correlation matrix of the transformed standard normal variables.

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

In more detail, the Lasso method is used to yield a sparse, invertible estimate of a 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 formulated below, as in Friedman et al. (2008), Banerjee & Natsoulis (2006) and Banerjee & El Ghaoui (2008):

$$
\max_X \log(\det X) - \text{tr}(\Sigma^T X) - v|X|_1 \tag{4.9}
$$

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 is formulated as a series of $L^1$ constrained problems called Lasso. Furthermore, Least Angle Regression (LAR) or other Lasso solvers can be used to obtain the corresponding solutions Efron et al. (2004).

This whole procedure is tractable because:

1. $X$ is the inverse covariance matrix of the transferred data. By estimating $X$ and keeping it invertible, $X^{-1}$ is the covariance matrix of the data and is always invertible. On the other hand, 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 and semi-definite. This property leads to parsimonious models for travel time.

4.2.3 Properties of GCMM with observable path scenarios

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

4.2.3.1 Main scenario analysis

By definition, the scenario based GCMM is composed of several Gaussian copulas for different path scenarios. An error analysis is first conducted to study the property of the estimator, when there is only one path scenario. In such a case, the scenario-specific mean travel time for each link is set as the unconditional mean travel time of the link, and only one Gaussian copula is used to model the dependence between the link travel times. Therefore, the estimated path travel time can be expressed \( \sum_{i=1}^{N} T_{t_i} \) and \( t_i = \sum_{j=1}^{i} ET_{t_j} \). 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

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

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

112
\[ \Delta^L_i \doteq E(T_{t_i})(k_i P(s_i = t_i)(t_i - \text{mins}_i) - 1) \]

\[ \Delta^U_i \doteq E(T_{t_i})(K_i P(s_i = t_i)(t_i - \text{mins}_i) - 1) \]

\[ \Gamma^L_i \doteq E(T^2_{t_i})(k_i P(s_i = t_i)(t_i - \text{mins}_i) - 1) \]

\[ \Gamma^U_i \doteq E(T^2_{t_i})(K_i P(s_i = t_i)(t_i - \text{mins}_i) - 1) \]

then the total mean error of the estimation is bounded by

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

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:

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

\[ = K_2(x) P(T_{s_2} = x|T_{s_1} = ET_{s_1}) P(T_{s_1} = ET_{s_1})(ET_{t_1} - \text{min}T_{t_1}) \]

\[ = K_2(x) P(T_{s_2} = x|T_{s_1} = ET_{t_1}) P(T_{s_1} = ET_{t_1})(ET_{t_1} - \text{min}T_{t_1}) \]

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

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

$$ET_{s_2} = \int xK_2(x)P(T_{s_2} = x|T_{s_1} = ET_{t_1})P(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{s_1})dx$$

$$= \int xP(T_{s_2} = x|X_{s_1} = ET_{t_1})K_2(x)dxP(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{s_1})$$

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

$$ET_{s_2} = \int xK_2(x)P(T_{t_2} = x)dxP(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{s_1})$$

Take $K_2 = sup_x(K_2(x))$

$$\Delta_2 = ET_{s_2} - ET_{t_2}$$

$$= \int xK_2(x)P(T_{t_2} = x)dxP(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{s_1}) - \int xP(T_{t_2} = x)dx$$

$$\leq \int xP(T_{t_2} = x)dx(K_2P(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{s_1}) - 1)$$

$$= E(T_{t_2})(K_2P(s_2 = ET_{t_1})(ET_{t_1} - minT_{s_1}) - 1)$$

$$= E(T_{t_2})(K_2P(s_2 = t_2)(t_2 - mins_2) - 1)$$

114
Similarly, for lower bound

$$\Delta_1 \geq E(T_{t_2})(k_2 P(s_2 = t_2)(t_2 - mins_2) - 1)$$

where $k_i = \sup_x K_i(x)$

Similarly when $n=i$,

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

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 $s_i = s_1 + \sum_{j=1}^{i-1} T_{s_j} = t_1 + \sum_{j=1}^{i-1} ET_{t_j} = t_i$

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

$$\quad = \int xK_i(x)P(T_{t_i} = x)dxP(s_i = t_i)(t_i - mins_i)$$
And then take $K_i = \sup_x K_i(x)$

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

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

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

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

$$\Rightarrow \Delta_i^L$$

Similarly, for lower bound

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

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

$$\Rightarrow \Delta_i^U$$

where $k_i = \sup_x K_i(x)$

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

$$\sum_{i=1}^{n} ET_{s_i} = \sum_{i=1}^{n} ET_{t_i} + \sum_{i=1}^{n} \Delta_i$$

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

116
Similarly for the variance, there is a bound of error in each of the $T_{s_i}$. Consider

$$
\Gamma_i = E T_{s_i}^2 - E T_{t_i}^2 \\
= \int x^2 K_i(x) P(T_{t_i} = x) dx P(s_i = t_i)(t_i - mins_i) - \int x^2 P(T_{t_i} = x) dx
$$

So

$$
\Gamma_i^L \leq \Gamma_i \leq \Gamma_i^U
$$

where

$$
\Gamma_i^L = E(T_{t_i}^2)(K_i P(s_i = t_i)(t_i - mins_i) - 1) \quad \Gamma_i^U = E(T_{t_i}^2)(K_i P(s_i = t_i)(t_i - mins_i) - 1)
$$

And then

$$
Var T_{s_i} - Var T_{t_i} = E T_{s_i}^2 - (E T_{s_i})^2 - (E T_{t_i}^2 - (E T_{t_i})^2) \\
= \Gamma_i - \Delta_i (E T_{s_i} + E T_{t_i})
$$

So

$$
\Gamma_i^L - \Delta_i^U (E T_{s_i} + E T_{t_i}) \leq Var T_{s_i} - Var T_{t_i} \leq \Gamma_i^U - \Delta_i^L (E T_{s_i} + E T_{t_i})
$$

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 in estimation of the GCMM based on observable path scenarios.
4.2.3.2 Over-all error analysis

In this section, the properties of the GCMM based on observable path scenarios when the number of path scenarios approaches infinity are examined: given appropriate assumptions, the estimator is proven to converge with the true underlying path travel time distribution if the number of path scenarios approaches 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 observations of different travelers together to constitute a path observation vector. After considering such pseudo path observations, the following conclusion holds:

Theorem 4.2.4 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 the number of data points increases to infinity. $\hat{\pi}_i \rightarrow \pi_i$ as $N \rightarrow \infty$ if one of the following three conditions is satisfied:

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

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

3. if the following set of assumptions hold:

(a) The arrival events for different scenarios form an independency, Çinlar (2011).

(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 is independent, including the starting time and the route choice to different links.

Proof to Theorem 4.2.4:

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 times on the path. Since the system is stationary, the occurring frequency of such path observations is in proportion to the stated probability given by the law of large numbers.

If there is an error \( \theta \) between the time when the previous traveler left the path and the time when the next traveler entered the link immediately downstream, and, furthermore, it is 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 traveler at a given link is a Poisson process with parameter \( \lambda \), then when Car 1 traverses Link 1, the probability that more than zero travelers arrive to Link 2 within the next \( \theta_0 \) seconds is \( 1 - e^{\lambda \theta_0} \)
By Assumption (c), the entrance of different travelers 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 of obtaining such a pseudo path observation is \( (1 - \exp(-\lambda\theta_0)) \). If three different sub-trips are needed, the probability of obtaining 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 a traveler continuing on 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 of obtaining a pseudo observation is:

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

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

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

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

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

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

120
while the overall distribution with the pseudo observations is

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

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, pseudo path travel time observations can be gradually excluded when original path travel time 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 \( \lambda_2 \). And suppose
the number of travelers entering the first link are $M_1$ and $M_2$ respectively, then the following result can be obtained:

Claim

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

as $n \rightarrow \infty$

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

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

$$E \left\{ \frac{1}{N_2} | M_1 \right\} = M_1 p_R$$

and by the approximation formula in Rempala (2004):

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

And therefore:

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

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

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

Similarly the following formula holds:
\[
E\left(\frac{\hat{N}_1}{\hat{N}_2} | M_1, M_2\right) \rightarrow \frac{M_1}{M_2} \text{ when } M_2 \rightarrow \infty
\]

So
\[
E\left(\frac{\hat{N}_1}{\hat{N}_2} | M_1, M_2\right) \rightarrow E\left(\frac{N_1}{N_2} | M_1, M_2\right) \text{ when } M_2 \rightarrow \infty
\]

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 \((\sum M_i, P_R)\). And then:

\[
E\left(\frac{\hat{N}_1}{\sum_j \hat{N}_j} | M_i\right) \rightarrow E\left(\frac{N_1}{\sum_j N_j} | M_i\right) \text{ when } M_i \rightarrow \infty
\]

Then by definition:

\[
\hat{\pi}_i = \frac{\hat{N}_i}{\sum_j \hat{N}_j}
\]

\(N_i \rightarrow \infty\) implies \(M_i \rightarrow \infty\)

\[
\rightarrow \frac{N_i}{\sum_j N_j} \text{ when } M_i \rightarrow \infty
\]

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.

From another angle, there is a connection between this infinite Gaussian copula mixture model with high dimensional kernel density estimation. Assume equal intervals are defined in this partition scheme of the high dimensional space and the same joint distribution \(K\) with
Gaussian copula is defined in any sub space with at least one data observation. Assume there are $N_m$ points in the $m$-th sub-space and the center of this sub-space is $M_m$, this simplified special case of GCMM is equivalent to usual high dimensional kernel density estimator for the joint distribution:

$$Y(x) = \frac{1}{nh} \sum_{i=0}^{N} K\left(\frac{x - x_i}{h}\right)$$

$$= \frac{1}{nh} \sum_{m=0}^{M} \sum_{j=0}^{N_m} K\left(\frac{x - x_{ij}}{h}\right)$$

$$\rightarrow \frac{1}{nh} \sum_{m=0}^{M} \sum_{j=0}^{N_m} K\left(\frac{x - M_m}{h}\right)$$

$$= \frac{1}{h} \sum_{m=0}^{M} K\left(\frac{x - M_m}{h}\right) \frac{1}{n} \sum_{j=0}^{N_m} 1$$

$$= \frac{1}{h} \sum_{m=0}^{M} K\left(\frac{x - M_m}{h}\right) \hat{\pi}_m$$

$$\rightarrow \frac{1}{h} \sum_{m=0}^{M} K\left(\frac{x - M_m}{h}\right) \pi_m$$

$$= \sum_{m=0}^{M} f_m \pi_m$$

The model in this section is a special GCMM 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 with the Lasso method using the data in the corresponding path scenarios. It is demonstrated that the error of such an estimator is bounded when the number of path scenarios is one, and the estimator is related to a kernel density estimator and has similar properties as the latter as the number of path scenarios approaches infinity.

An important feature of the GCMM based on observable path scenarios addressed in this section is that each path travel time observation is classified precisely into an observable path
scenario for further estimation of copula parameters. If the classification of each data vector is conducted in a probability sense, i.e. a path travel time observation can be classified into different path scenarios with different probabilities, then the GCMM based on observable path scenarios can be generalized to more flexible models. Such flexibility in classification helps to improve the properties of these estimators when the number of path scenarios is finite.

4.3 GCMM based on unobservable path scenarios and extended expectation maximum algorithms

In this section, the GCMM based on unobservable path scenarios is discussed, which has two important characteristics that contrast with the GCMM based on observable path scenarios: First, in a GCMM based on unobservable path scenarios, it is assumed that path scenarios can not be observed directly but can be fully discovered based on available travel time observations without first defining scenario characteristics: marginal travel time distributions and joint structure can be updated in an iterative algorithm and fixed at the end. Second, it is assumed that each path’s travel time observation can belong to different scenarios with different probabilities instead of classifying path observations strictly into distribution scenarios, then, the model automatically fits the empirical data in an integrated fashion, leading to more efficient usage of limited data.

Essentially, estimating a GCMM based on unobservable path scenarios involves optimizing the parameters of the GCMM (all weights $\pi_k$ and dependency parameter $P_k$ of the Gaussian copulas) using the maximum likelihood method. Compared to GCMM based on observable path scenarios, there are fewer constraints in the GCMM in this section attributed to pre-definition of path scenarios and pre-classification of observations in these scenarios. Therefore, given a finite number of path scenarios, optimization with fewer constraints may
lead to higher values for the likelihood function and better quality of fit.

To conduct the estimation of a GCMM based on unobservable path scenarios, two estimation algorithms are designed. The difference between the two algorithms concerns whether or not the link travel time distribution is fixed during the iteration. If the link travel time distributions for each path scenario are first defined according to transportation context and therefore fixed in the iteration, a fixed-marginal-distribution GCMM can be used to estimate the dependent parameters; if the link travel time distributions for each path scenario 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 identify the optimal parameters of Gaussian copulas, which can then be used to explain the pattern of historical path travel time observations. This is the specific expectation-maximum algorithm for the likelihood function of GCMM. The steps of this algorithm are organized as follows:

**Fixed-marginal-distribution GCMM**

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

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

3. Maximum step:
   
   (a) Convert the marginal observations into probability values using the marginal cumulative distribution functions $F_{X_i}$. 

126
(b) Conduct maximum likelihood estimation for the Gaussian copula mixture model given $r_{nk}$.

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 $P_{mk}^m$. 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 a linear combination of these $K$ Gaussian copulas.
  - The overall likelihood function is denoted as

$$L = \sum_{n=1}^{N} \ln(\sum_{k=1}^{K} \pi_k \frac{1}{(2\pi)^{n/2}} \frac{1}{P_{mk}^{m}} \exp\left(-\frac{1}{2} \Psi^T P_{mk}^{m} Y_{n,k}^{m}\right)) \prod_{i=1}^{D} \frac{Z_{n,ki}^{m}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (Y_{n,ki}^{m})^2\right) \tag{1.10}$$

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

127
• Step 2: Expectation step. By taking partial derivatives to the Likelihood function with respect to $\pi_k$, a ratio $r_{nk}^m$ can be defined

$$D_{nk}^m = \prod_{i=1}^{D} \frac{1}{\sqrt{2\pi}} \exp \left( \frac{1}{2} (Y_{nk,i}^m)^2 \right)$$

$$\frac{\partial L}{\partial \pi_k} = \sum_{n=1}^{N} \frac{r_{nk}}{\pi_k}$$

where

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

(4.11)

• Step 3: Maximum step:

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

$$\tilde{L} = L + \lambda \left( \sum_{k=1}^{K} \pi_k - 1 \right)$$

$$\frac{\partial L}{\partial \pi_k} = 0$$

128
Then

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

and

\[
\pi_{nk}^m = \frac{\sum_{n=1}^{N} r_{nk}}{N} \quad (4.12)
\]

Next, consider the partial for \( P_k^{-1} \)

\[
\frac{\partial L}{\partial P_k^{m,-1}} = \frac{\frac{\partial \pi_{nk}^m \pi_{nk}^m}{\partial P_k^{m,-1}}}{\frac{\partial P_k^{m,-1}}{\partial P_{k}^{m,-1}}} \exp\left(-\frac{1}{2}(Y_{nk}^m)^T P_k^{m,-1} Y_{nk}^m\right)
\]
\[
= \frac{D_{n,k} \pi_{nk} \exp\left(-\frac{1}{2}(Y_{nk}^m)^T P_k^{m,-1} Y_{nk}^m\right)\left[|P_k^{m,-1}|^{-1/2}|P_k^{m,-1}|(P_k^{m})^T + |P_k^{m,-1}|^{1/2}(-\frac{1}{2})Y_{nk}^m\right]}{\sum_{j=1}^{K} D_{n,j} \pi_{nk} \exp\left(-\frac{1}{2}(Y_{nk}^m)^T P_j^{m,-1} Y_{nk}^m\right)\left[|P_j^{m,-1}|^{-1/2}|P_j^{m,-1}|(P_j^{m})^T + |P_j^{m,-1}|^{1/2}(-\frac{1}{2})Y_{nk}^m\right]}
\]
\[
= \frac{1}{2} \sum_{n=1}^{N} r_{nk} (P_k^m - Y_{nk}^m (Y_{nk}^m)^T)
\]

By making the derivative to equal to zero:

\[
P_k^m = \frac{\sum_{n=1}^{N} r_{nk} Y_{nk}^m Y_{nk}^m}{\sum_{n=1}^{N} r_{nk}} \quad (4.13)
\]

- 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 to which the likelihood functions converge. 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.

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 begins. Further considerations of the algorithm structure indicate that it may be further improved by updating the link travel time distributions in each path scenario during each iteration of the expectation-maximum algorithm. During the iteration of this algorithm, both link travel time distributions and parameters of the Gaussian copulas are updated, which leads to potentially higher likelihood values. This algorithm allows path travel time scenarios to be fully discovered based on data and it is useful when the initial path scenario is generated by a statistical cluster analysis and the travel time scenarios for each link are not fixed.

Basic Properties of Varying-marginal-distribution GCMM

A GCMM is defined based on the separation of the mixture of copulas and marginal distributions, which may potentially lead to different behavior from a GMM. To understand the properties of the GCMM, its likelihood function is studied so that appropriate estimation algorithms can be designed. The major properties of the GCMM are discussed below:
A GCMM has a bounded likelihood function value on bounded domains and tractable
derivatives conditioned on the estimated marginal probability functions. The likelihood
function is given below:

\[
L = \sum_{n=1}^{N} \ln\left(\sum_{k=1}^{K} \frac{1}{(2\pi)^{n/2}} \pi_k \exp\left(-\frac{1}{2} (Y_{n,k})^T P Y_{n,k}\right) \prod_{i=1}^{D} \frac{Z_{n,ki}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (Y_{n,ki})^2\right)\right)
\] (4.14)

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** Under mild conditions (Given the \(P\) and \(x_n\), assume there are bounds
to the probability values on \(x_n\) for each of the marginal distribution \(F_{kd}\)), the likelihood
function is bounded above in bounded region; non-decreasing and negative semi-definite
w.r.t density \(Z_{n,ki}\); may contain both local minimum and local maximum w.r.t trans-
formed variables \(Y_{n,k}\).

Proof:

Consider maximizing the following function

\[
L = \sum_{n=1}^{N} \ln\left(\sum_{k=1}^{K} \frac{1}{(2\pi)^{n/2}} \pi_k \exp\left(-\frac{1}{2} (Y_{n,k})^T P Y_{n,k}\right) \prod_{i=1}^{D} \frac{Z_{n,ki}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (Y_{n,ki})^2\right)\right)
\] (4.15)

with the constraints:

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

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

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

with

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

and

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

Then

\[
\frac{\partial \hat{L}}{\partial Z_{n,kj}} = \frac{1}{S_n} \pi_k \left( \frac{1}{(2\pi)^{n/2} P^{1/2}} \right) \exp \left( -\frac{1}{2} Y_{n,kj}^T P Y_{n,kj} \right) \prod_{i \neq j} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} Y_{n,ki}^2 \right)
\geq 0
\]

Where

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

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

where

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

And

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

132
\[ \gamma_{n,k}^T Z_{n,k} = 0 \]

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

\[ \frac{\partial \hat{L}^2}{\partial Z_{n,kj}^2} \geq 0 \]

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

\[ Z_{n,k} = \frac{S_n(-\gamma_{n,kj} + \theta_{n,kj})}{D_{n,k} \frac{1}{2\pi} \exp\left(-\frac{1}{2}Y_{n,ki}^2\right)} \] (4.16)

\[ \gamma_{n,ki} \frac{S_n(-\gamma_{n,kj} + \theta_{n,kj})}{D_{n,k} \frac{1}{2\pi} \exp\left(-\frac{1}{2}Y_{n,ki}^2\right)} = 0 \] (4.17)

\[ \theta_{n,ki} \frac{S_n(-\gamma_{n,kj} + \theta_{n,kj})}{D_{n,k} \frac{1}{2\pi} \exp\left(-\frac{1}{2}Y_{n,ki}^2\right)} - C_i = 0 \] (4.18)

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 \( Z_{n,k}^{b_1} \); (2) \( \gamma_{n,kj} \neq 0 \) and \( \theta_{n,kj} = 0 \), the solution is denoted as \( Z_{n,k}^{b_2} \); (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 \( B_{n,k} = \pi_k \frac{1}{(2\pi)^{n/2} \mu_n/2} \exp\left(-\frac{1}{2}Y_{n,k}^T P Y_{n,k}\right) \) and \( D_{n,k} = \prod_{i=1}^{D} \frac{Z_{n,ki}}{\sqrt{2\pi} \exp\left(-\frac{1}{2}Y_{n,ki}^2\right)} \) as defined in the previous section.
\[
\frac{\partial L^2}{\partial Y_{n,k}^2} = \frac{1}{S_n^2} (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^2} (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 and 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 \(\frac{1}{S_n^2} B_{n,k}D_{n,k}\) 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 \(\Lambda_{n,k} \geq 0\), \(\frac{\partial L^2}{\partial Y_{n,k}^2} \geq 0\) and \(\frac{\partial L^2}{\partial Y_{n,k}^2} \leq 0\)

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

Then

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

(4.19)

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

(4.20)
\[ \beta_{n,k}^T \left( \frac{S_n}{B_{n,k} D_{n,k}} \right) (-P + I)^{-1}(\alpha_{n,k} - \beta_{n,k}) - 1) = 0 \] (4.21)

Then \( Y_{n,k} \) can be solved using the equations above. Furthermore, 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 \( Y_{n,k}^{1b1} \); (2) \( \alpha_{n,k} \neq 0 \) and \( \beta_{n,k} = 0 \) the solution is denoted as \( Y_{n,k}^{1b2} \).

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 \( Y_{n,k}^{2b1} \); (2) \( \alpha_{n,k} \neq 0 \) and \( \beta_{n,k} = 0 \), the solution is denoted as \( Y_{n,k}^{2b2} \); (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.

- The estimation algorithms for GCMMs combine the two-step maximum likelihood estimator for copulas and the Expectation Maximum algorithm for GMMs together. In each iteration, the value of its likelihood function is nondecreasing during iterations of Expectation-Maximum algorithms that are applied with GCMMs and the algorithms converge globally to local maximums under mild conditions Wu (1983). The design and properties of these Expectation-Maximum algorithms are discussed in the next section.

- Model selection is conducted using Akaike and Bayesian information criterion to balance the accuracy and model complexity, Jianqing Fan & Wu (2009). and cluster methods such as k-means or hierarchy clustering can be used to set the initial parameters of each component.
Figure 4.5: Comparison of GMM and GCMM base case: n: data index; m: iteration index; k: copula index; i: dimension index

**GCMM Varying-marginal-distribution Base Case**

The algorithm updates the mixture of copulas and the marginal distributions separately. Essentially when estimating GMMs, the weights $\pi^m_k$ & correlation matrixes of components $P^m_k$ and the sufficient statistics (mean $\mu^m_{ki}$ and standard deviation $\sigma^m_{ki}$) of the marginal normal distributions are updated (Dempster 1977 Dempster et al. (1977)) based on the posterior probability $\gamma^m_{nk}$. In GCMMs, the sufficient statistics of marginal normal distributions are replaced with non-parametric estimators to the marginal pdf $f^m_{ki}$ and cdf $F^m_{ki}$ to improve flexibility, see the red boxes in Figure 4.5.

The major challenge of algorithm design lies in how the marginal distributions should be updated considering the posterior probability. An updating formula is developed and given by the following theorem:

**Theorem 4.3.2** In the GCMM base case, the updating of the marginal distributions follows the following formula with necessary normalizations:

$$F'_{ki}(c) = \sum_n \gamma_{nk} 1_{x_{ni} \leq c}$$

Proof:

136
Denote $x_n$ as the observed synchronized data vector, $z$ are the complete data. Recall in the Expectation step we calculate the posterior probability $\gamma_{nk}$ for $n$-th data vector belong to $k$-th cluster such that the incomplete data likelihood function below is expressed explicitly.

$$Q(\pi', P', F'|\pi, P, F) = E(\log f(z)|x_n, \pi, P, F)$$

In the Maximum step we calculate $[\pi', P', F'] = \arg\max_{\pi', P', F'} Q(\pi', P', F'|\pi, P, F)$ to obtain new parameters based on such $\gamma_{nk}$.

In this process, the poster distribution $\gamma_{nk}$ is $\gamma_{nk} = p_k(\pi, P, F)

So the natural estimator for the marginal distribution for the $i$-th dimension of the $k$-th component is its histogram conditioned on the current weights: $F_{ki}(c) = p_{ki}(x_{ni} \leq c|\pi, P, F) = \sum_n p_k(x_{ni}|\pi, P, F)1_{x_{ni} \leq c} = \sum_n \gamma_{nk}1_{x_{ni} \leq c}$

Further normalization is used to maintain the properties of a cdf and other univariate non-parametric estimator can be used. Q.E.D

Based on the theorem, the algorithm is further developed below:

**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 $r^m_{nk}$.

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 given $r^m_{nk}$.

4. Update the marginal link travel time probability distribution functions $F_n$ according
to the posterior classification probability $r_{n,k}^m$.

$$F_{ki}^m(y) = \frac{\sum n r_{nk}^m 1_{x_{ni} \leq y}}{\sum n r_{nk}^m} , \forall k\text{th copula, ith dimension}$$

(kernel smoothing can be applied)

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


The issue here is that heavy tail phenomena may be categorized into two classes: the heavy tails in the marginal distribution and the heavy tails in the dependence structure. GCMM separates the estimation for them and control the number of clusters purely based on the complexity of heavy tails in dependence structure (the latter). In this manner, the number of clusters could be further reduced and the mixture of copulas are robust towards heavy tails on the marginal distributions (the former).

**GCMM with Unsynchronized Data**

Furthermore, GCMMs with unsynchronized data are developed based on the rationale that unsynchronized data in each dimension can be used to update the marginal distribution, given the estimation of marginal distribution is separated from the mixture of copulas. An additional posterior probability $\gamma_{n_i,k}^m$ is introduced to represent the probability of $n_i$-th unsynchronized data on the $i$-th dimension belonging to the $k$-th component. An additional loop is then inserted into the Expectation Maximum algorithm for GCMM base case which further updates $\gamma_{n_i,k}^m$ based on new information, see the orange loop in Figure 4.6.

The major challenge of algorithm design lies in how the marginal distributions should be further updated given unsynchronized data and the existing nonparametric estimator. An updating formula is developed and given by the following theorem:
Theorem 4.3.3 In the GCMM with unsynchronized data, the updating formula of marginal distribution follows by the following formula with necessary normalizations:

\[ r'_{n_i,k} = \frac{\pi_k f_{kl}(x_{n_i})}{\sum_{k=1}^{K} \pi_k f_{kl}(x_{n_i})} \]

\[ F'_{kl}(c) = \sum_{n} \gamma_{nk} 1_{x_n \leq c} + \sum_{n_i} \gamma'_{n_i,k} 1_{x_{n_i} \leq c} \]

Proof:

Denote \( x_n \) as the observed synchronized data vector, \( x_{n_i} \) as the \( n_i \)-th observed unsynchronized data on the i-th dimension and \( z \) as the complete data. Recall in the Expectation step of the likelihood function is to calculate the posterior probability \( \gamma_{nk} \) for \( n \)-th data vector belong to \( k \)-th cluster such that the incomplete data likelihood function below is expressed explicitly.

\[ Q(\pi', P', F'|\pi, P, F) = E(\log f(z)|x_n, x_{n_i}, \pi, P, F) \]

Moreover, we also calculate the posterior probability \( \gamma'_{n_i,k} \) for \( x_{n_i} \) (the \( n_i \)-th unsynchronized observation on the i-th dimension) to belong to \( k \)-th cluster based on \( F_{kl}(c) \).

In the Maximum step we calculate \([\pi', P', F'] = \arg\max_{\pi', P', F'} Q(\pi', P', F'|\pi, P, F) \) to obtain new parameters based on such \( \gamma_{nk} \) and \( \gamma'_{n_i,k} \).

The posterior distribution \( \gamma_{nk} \) is \( \gamma_{nk} = p_k(x_n|\pi, P, F) \)
The poster distribution $\gamma'_{n_i,k}$ is

$$\gamma'_{n_i,k} = p_k(x_n|\pi, P, F) = \frac{p_k(x_n|\pi, P, F, K=k)P(K=k)}{\sum_k p_k(x_n|\pi, P, F, K=k)P(K=k)} = \frac{\pi_k f_k(x_n)}{\sum_k \pi_k f_k(x_n)}$$

So the natural estimator for the marginal distribution for the $i$-th dimension of the $k$-th component is its histogram conditioned on the current weights for all data on that dimension. $F'_{ki}(c) = p_k(x_{ni} \leq c|\pi, P, F) + p_k(x_{ni} \leq c|\pi, P, F) = \sum_n p_k(x_n|\pi, P, F)1_{x_n \leq c} + \sum_n p_k(x_{ni}|\pi, P, F)1_{x_{ni} \leq c} = \sum_n \gamma_{nk}1_{x_n \leq c} + \sum_n \gamma'_{n_i,k}1_{x_{ni} \leq c}$

Further normalization is used to maintain the properties of a cdf and other univariate non-parametric estimators can be used. Q.E.D

Based on the theorem, the algorithm is further developed below (similar parts as the base case are ignored to save space):

**Varying-marginal-distribution GCMM employing un-synchronized data**

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:
   
   (a) update $r_{nk}^m$ for synchronized data;

   (b) update $r_{n_i,k}^m$ for un-synchronized data using the following Bayes formula:

   $$r_{n_i,k}^m = \frac{\pi_k f_k^{m-1}(x_{ni})}{\sum_{k=1}^K \pi_k f_k^{m-1}(x_{ni})}$$

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 given $r_{n,k}$.
4. Update the marginal link travel time cumulative distribution functions $F_n$ according to the posterior classification probability $r_{n,k}^m$ and $r_{n,i,k}^{m'}$.

$$F_{k_i}^m(y) = \frac{\sum_n r_{n,k}^m 1_{x_n \leq y} + \sum_n r_{n,i,k}^{m'} 1_{x_n \leq y}}{\sum_n r_{n,k}^m + \sum_n r_{n,i,k}^{m'}} \forall kth copula, ith dimension$$

(kernel smoothing can be applied)

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


The philosophical issue here is whether synchronized data truly represent the joint distribution adequately and whether the unsynchronized data may add to our understanding of it. To bring unsynchronized data into the whole Expectation Maximum algorithm enlarges the information set of the probability space $(\Omega, \mathcal{F}, P)$ so that deeper elaboration of the data is possible (Cinlar 2011 Çinlar (2011)). This is a significant improvement from GMM beyond the flexibility applied to the marginal distribution.

4.4 Numerical analysis

4.4.1 Identification of Observable Path Scenarios via Cluster analysis

In this section, an experiment is conducted to demonstrate the identification of path scenarios via cluster analysis.

- Path scenario identification is first conducted with the help of cluster analysis. The link travel times on the three links AB, BC and CD, as highlighted in Figure 3.3, are collected and displayed in the three dimensional plot on top in Figure 4.7; Hierarchy
clustering is run on these data points and the top 30 clusters are shown on the bottom left; The histogram for the number of data points across different clusters is shown on the bottom right, it demonstrates that only three clusters contain significantly more number of points than others. Hence these three clusters are identified as the major path scenarios.

- Path scenario identification are further confirmed with possible transportation factor analysis such as time of day and the correlation to these factors are not contribute
significantly to the path scenario identification, in this particular case.

- Estimation of the path travel time distribution using the selected path scenarios that is identified by cluster analysis is demonstrated in Figure 4.8. In this figure, the green pdf is generated using the three major path scenarios that are with the greatest numbers of observations, as in the histogram in Figure 4.7. Compared to the black empirical pdf, the green pdf does not capture the heavy upper tail well. An additional red pdf is generated by further categorizing the remaining data points into three path scenarios (hence six path scenarios are used); it is found that the red pdf approximates the heavy tails noticeably better. These remaining points may contain outliers and correct treatment for them lead to better estimation of tails.

![Figure 4.8](image.png)

**Figure 4.8: Estimated path travel time distribution: Based on Major Three Clusters: Green; With Other Outliers: Red; Empirical: Black**

- Although GCMMs based on observable paths scenarios are straightforward to estimate, there may be several limitation to them:
  - Identification of Path scenarios requires experience. Cluster analysis is a data
driven process while more expert judgement need to be imposed to confirm the identified path scenarios;

- As shown in the tests above, treatment of outliers have noticeable impact to the estimation result;
- Factor sensitivity may not be high based on regression analysis;

4.4.2 Sensitivity test for GCMM based on observable path scenarios

In this section, an experiment is conducted examining the path travel time estimation methodology based on observable path scenarios proposed in this chapter. To this end, 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.

![Experimental network in New Jersey](image)

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

Figure 4.9: The experimental network in New Jersey

First, the sensitivity of the estimation with respect to $\Delta$ is studied in Figure 4.10, which is the bound of the time difference when different travelers assemble pseudo path observations.
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, relatively few new observations are included. In later experiments, $\Delta$ is set as a small portion of the standard deviation.

Figure 4.10: Change of estimation as $\Delta$ changes (Empirical: Red dots; $0.5 \sigma_i$: Cyan; $\sigma_i$: Red; $2 \sigma_i$: Blue; $3 \sigma_i$: Black)

Second, the sensitivity of the estimation with respect to Lasso penalty $v$ is studied while fixing $\Delta$. The estimates based on different Lasso penalties are shown in Figure 4.11. 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 path scenarios is studied. Path scenarios are identified using the overall congestion level on each link: all travel time observations within a range are categorized into one scenario for one link and it is assumed that the number of possible travel time scenarios in each link can be one (the
Figure 4.11: 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)

main scenario), two, four, eight, or sixteen so that the number of possible path scenarios is the square of these numbers as only two links are studied together due to constraints on dimensionality (part of the original 12-link path). The estimated cumulative distribution functions are given in Figure 4.12. The experiment shows that (1) as the number of possible travel time scenarios in each link increases, the difference between estimated distribution and empirical distribution tends to be smaller; (2) 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); (3) there is bias when the number of path scenario is limited; (4) This bias can be further reduced by better identification of path scenarios using expert judgment or using the GCMM based on unobservable path scenarios introduced in Section 4.3.

The probability density function is shown in Figure 4.13. When the number of states on each link is two, the upper tail of the empirical distribution is not well described. When the number of possible travel time scenarios on each link is eight, the upper tail is better
Figure 4.12: 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 described by the model.

Figure 4.13: Left: 2 scenarios per link; Right: 8 scenarios per link; Estimated pdf: Red; empirical: Green; scenario-specific cdfs: Blue

Fourth, to illustrate the comparison of the two competing paths using different decision
statistics, the path travel time distribution on two paths are compared (the lasso penalty is \( v = 0.001 \)). The estimated path travel time distributions are displayed in Fig 4.14.

Figure 4.14: 12 Path travel time based on the approximation; Path 1: Red; Path 2: Blue

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.

Table 4.1: Decision statistics for different rules for the two paths

<table>
<thead>
<tr>
<th>Parameter</th>
<th>((\mu, \sigma))</th>
<th>(\mu + r\sigma^2) (r=0.05)</th>
<th>FSD first violation</th>
<th>SSD first violation</th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

Exponential utility

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Exponential utility</th>
<th>AR</th>
<th>VAR(Quantile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a=-1/1000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Path 1</td>
<td>0.2476</td>
<td>251.5</td>
<td>1832</td>
</tr>
<tr>
<td>Path 2</td>
<td>0.3509</td>
<td>559.5</td>
<td>1394</td>
</tr>
<tr>
<td>Preference</td>
<td>Path 2</td>
<td>Path 2</td>
<td>Path 2</td>
</tr>
</tbody>
</table>

Overall, the following advantages are achieved using the GCMMs introduced in this section:
1. The non-normal distribution of travel time in each path scenario can be uncovered and
the heavy tail in the overall path travel time distribution can be modeled.

2. The number of parameters can be reduced further compared to Gaussian Mixture
models.
   (a) The number of components (copulas) necessary to describe the whole joint dis-
   tribution can be fewer as heavy-tailed marginal distribution could explain more
   deviation in data;
   (b) The mean of the marginal distribution are eliminated as the marginal distribution
   is estimated directly using posterior probability via nonparametric methods.

3. A variation of the GCMM model can take advantage of the extra link travel time
observation data which cannot be used in GMM fitting, as the estimation of copula
and the marginal distributions is separate.

4.4.3 Simulation Experiment for GCMM based on unobservable
path scenarios

In this section, a simulation experiment is conducted to demonstrate the performance of the
fixed-marginal-distribution GCMM and the varying-marginal-distribution GCMM: two link
travel time samples are simulated and the dependent structure of the two link travel time
is determined using several copulas. The marginal distributions in each cluster can contain
heavier tails than those in Gaussian joint distribution, which is difficult to described using
the one copula mentioned in Chapter 3. Different alternative models are used to estimate the
dependent structure and the path travel time distribution. Finally, the estimated distribution
is compared to the empirical path travel time distribution.

In this experiment, four methods are compared: Model (1): a Gaussian mixture model
(GMM); Model (2): a fixed-marginal-distribution GCMM (based on unobservable path
scenarios); Model (3): a varying-marginal-distribution GCMM using only the path travel
time observations (based on unobservable path scenarios); Model (4): a varying-marginal-
distribution GCMM 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 constructed to esti-
mate the dependent structure of the two links. There are additional travel time observations
on each link that are not matched to a path travel time observation vector, because some
travelers may just have traversed one of the two links at the observation time.

First, the link travel time distribution for each path scenario is compared and Figure
4.15 displays the results for the second link: For GMM (Model 1), the estimated link travel
time distribution for each path scenario is a Gaussian distribution. For GCMM based on
fixed marginal link travel time distributions (Model 2), the fixed marginal link travel time
distribution is determined via one dimensional GMM classification. For GCMM with varying
marginal distributions (Model 3, 4), the link travel time distributions in each path scenario
is estimated using weighted kernel smoothing based on all path travel time observations
where the weights are the posterior probability that a travel time observation vector belongs
to a given path scenario. Again, one data point may be associated with different path
scenarios/clusters in a probability sense instead of a strict categorization in GCMM.

Second, the cluster structure is estimated using different models and compared in Figure
4.16: For GMM, the centers of joint Gaussian distributions in GMM are displayed; For
GCMM (Model 2, 3, 4), the mean link travel time in each path scenario is displayed. Here
the number of clusters is selected via AIC information criteria and the number of copulas in
GCMM is smaller than that in GMM.

Third, the travel times on the two links are simulated using the estimated model; their
sum is calculated and its distribution is generated. The distribution of their sum is com-
pared to the empirical path travel time distribution, which is the distribution of the sum
of the travel times on the two links in the empirical data set via QQ-plot and two sample
Figure 4.15: Comparison of the marginal link travel time distributions on one selected link across different models

Kolmogorov-Smirnov test. The QQ-plots are displayed in Figure 4.17.

The results of the two sample Kolmogorov-Smirnov tests are displayed in Table 4.2. The results show that the two models based on varying marginal GCMM fit the original path travel time distribution reasonably well.
Figure 4.16: Comparison of the cluster structure between the two links across different models

Table 4.2: p-values of two-sample K-S test compared with the empirical distribution

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0002</td>
<td>0.0252</td>
<td>0.1304</td>
<td>0.1003</td>
</tr>
</tbody>
</table>

Overall, these results may suggest that:
Figure 4.17: QQ-plots of the estimated path travel time distributions to empirical distribution across different models

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

- The tail of the estimated path travel time distribution based on fixed-marginal-distribution GCMM is lighter than that of the true tail possibly because the pre-defined link sce-
nario should be improved to match the true scenarios on the links;

- The estimated path travel time distribution based on varying-marginal-distribution GCMM displays good risk properties by estimating the empirical path travel time distribution more closely. The tails are slightly heavier than the latter as well.

- The additional marginal data brings slight changes to the estimated GCMM, which employs the unused information in traditional cluster models. It is the separation of marginal distribution and copula that renders such data useful.

4.4.4 Empirical Test for GCMM based on unobservable path scenarios

The real data set from the transportation system using the travel time of individual drivers in New Jersey which is captured from GPS devices is employed for model testing and the data is displayed in Figure 4.7. The same procedure is used as the simulation test in the previous section except the calibration data set is real and only GCMM with varying marginal distributions are studied to further demonstrate their flexibility. The results are summarized below, to save space the three-dimensional clusters are omitted:

- The comparison to the empirical path travel time distribution is shown in Table 4.3 for a three segment path. Akaike information criteria indicates both the GMM and the GCMM need three components to describe the data well and p-values of the K-S tests for GCMM are noticeably larger.

Table 4.3: p-values of two-sample K-S test compared with the empirical distribution

<table>
<thead>
<tr>
<th></th>
<th>GMM</th>
<th>Base Case</th>
<th>Extra-Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0518</td>
<td>0.9646</td>
<td>0.1157</td>
</tr>
</tbody>
</table>

- Estimated distributions are compared in Figure 4.18: GCMM base case fit the empirical distribution best and GCMM with unsynchronized data captures heavier tails as
there are some higher values in the unsynchronized data. The heavier tail in GCMM with unsynchronized data is caused by differences in marginal distributions due to new information in the synchronized data, but not by material changes of the mixture of copulas (Please note the black empirical distribution is estimated based on the synchronized data and it does not account for potential heavy tails in the unsynchronized data).

In summary, the research in this chapter expands upon the framework in the previous chapter to predict path travel time distributions. The research herein introduced a GCMM based on observable path scenarios in which the Lasso method is used to obtain reliable estimates of copula structures. Moreover, GCMMs based on unobservable path scenarios are also introduced; the Expectation Maximum algorithm is extended to estimate the GCMMs. Overall, GCMMs first add more flexibility to fit heavy tails on marginal distributions while remaining relatively robust against it; GCMMs further incorporate unsynchronized data into estimation, both of which improve the approximation to the complex dependence structures given limited number of components. 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 utilized in disciplines other than transportation research.
Figure 4.18: Top: Comparison of pdf (Red: GMM; Blue: GCMM base case; Cyan: GCMM with unsynchronized data; Black: Empirical); Bottom left: Green QQplot Empirical(x) v.s. GCMM base case(y); Bottom right: Black QQplot Empirical(x) v.s. GCMM with unsynchronized data(y)
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, for use by transportation facilities. A travel time derivative is a useful financial product to (a) hedge against transportation-related risk by users of the transportation facility, (b) manage the demand for those facilities through derivatives-based dynamic tolling, (c) contribute to risk mitigation through portfolio diversification and (d) provide an additional source of revenue for owners of transportation facilities. In this chapter, potential market participants are analyzed first, and then major products designed for a travel time derivatives market are presented. Alternative models for describing underlying travel time changes are discussed, together with corresponding pricing methods.
5.1 Initiation and necessity analysis

5.1.1 Derivatives and weather derivatives as hedging tools

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 upon 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 elements of derivatives can be further classified as tradable and non-tradable. Most items listed in the preceding paragraph as bases for derivatives can be traded in a market and so are called tradable underlying assets. A few others, such as temperature, precipitation, and travel time, are non-tradable. This difference in the tradability of an underlying asset triggers differences in market making mechanism, trading strategy and pricing methods. To provide more background, market making mechanism refers to the practice of a broker-dealer firm accepting the risk of holding a certain number of shares of a particular derivative contract in order to facilitate trading in that derivative contract in the financial market; a trading strategy is a fixed plan that is designed to achieve a profitable return by going long or short of the derivative contract in markets; pricing method implies the determination of an appropriate price for the derivative contract so that market participants such as travelers and firms in the transportation industry can gain profile and control risk appropriately, Hull (2009).

Derivatives based on tradable assets are common in the current financial markets. Stock options are typical financial derivatives. A European call option on stock is a contract in
which the buyer gets a payoff if the price of underlying stock is higher than a given value on a given future date. The buyer has to pay a premium to enter this contract, which is the value of this contract. As the prediction of future stock value changes, the value of this contract changes accordingly and this stock option contract is hence a derivative based on the underlying stocks. These products are ordinary financial products traded in market and major investment make markets for all participants. Usually, investors purchase such derivatives to hedge the risk of their positions in the corresponding stocks and more advanced trading strategies are available for these derivatives. The Black-Scholes pricing model is a classic pricing method based on the idea that the risk of these derivatives can be hedged by dynamic trading the corresponding underlying instruments in the market.

Derivatives based on non-tradable assets were first introduced in 1999, when 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) analyzed participants and roles in that futures market and found that weather derivatives act as alternative and more flexible ways of insuring against weather related risk. Industries subject to weather risk participate in the buy/sell side of the market, while speculators, who trade purely for profit, provide an important source of liquidity.

Weather derivatives provide insurance to farmers and agriculture companies against bad weather and low crop output. The payoff for one farmer 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. This risk-protection mechanism is shown in Table 5.1. In the table, $G$ represents the gain on the derivative, and $P$ represents the premium that the farmer pays for the contract.
Table 5.1: Payoff of typical weather derivatives

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

To see some typical weather future contracts traded in the market, refer to Table 5.2, CME (2013). In this table, the settings of Canadian Degree Days index HDD futures are displayed. In this contract, 18 degrees Celsius ($C$) minus the average daily temperature floored at zero is summed over a given calendar month for several given measurement stations in Canada. Suppose the buyer purchases the contract at price $P$; initially, if temperatures in the specified measurement stations are generally lower and this sum is potentially higher in a given month, the price of the contract would increase. In this case, the buyer profits from the price increase and increased payoff $G - P$ to compensate potential losses due to cold weather.

As weather derivatives are used to hedge risk related to temperature, precipitation, and other factors, the pricing of weather derivatives should primarily be based upon prediction of weather conditions. Based on accurate prediction of future weather changes, the contract is priced using different methods other than the Black-Scholes pricing model, considering the fact that weather conditions are not traded in the market. These pricing methods are not founded on dynamic hedging of the un-tradable underlying instruments but on other more general pricing schemes such as risk neutral pricing under incomplete market conditions, indifferent pricing principles and etc, which are introduced in greater detail in later sections.

5.1.2 Travel time derivatives are a flexible value pricing scheme

Formally, a travel time derivative contract is a financial instrument whose prices are derived from the value of travel time measurements. The introduction of such a contract may bring several benefits to the transportation and financial system, which are addressed in this and
Table 5.2: Canadian Degree Days Index (HDD) Futures traded on Chicago Mercantile Exchange

<table>
<thead>
<tr>
<th>Contract Size</th>
<th>CAN $20 times the respective CME Canadian Degree Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index Product Description</td>
<td>Heating Degree Days (HDD) for Canadian Cities</td>
</tr>
<tr>
<td>Measurement definition</td>
<td>The temperature for a particular city is reported from a specific automated weather station: Calgary International Airport (WMO 71877) Edmonton International Airport (WMO 71123) Montreal/Pierre Elliot Trudeau Airport (WMO 71627) Toronto Pearson International Airport (WMO 71624) Vancouver International Airport (WMO 71892) Winnipeg International Airport (WMO 71852)</td>
</tr>
<tr>
<td>Pricing Unit</td>
<td>Canadian Dollars (CAN$) per index point</td>
</tr>
<tr>
<td>Tick Size</td>
<td>1 index point</td>
</tr>
<tr>
<td>(minimum fluctuation)</td>
<td>(= CAN$20 per contract)</td>
</tr>
<tr>
<td>Trading Hours</td>
<td>CME Globex (Electronic Platform)</td>
</tr>
<tr>
<td>(All times listed are Central Time)</td>
<td>SUN 5:00 p.m. - FRI 3:15 p.m. Daily trading halts 3:15 p.m. - 5:00 p.m.</td>
</tr>
<tr>
<td>Last Trade Date/Time</td>
<td>Fifth Exchange business day</td>
</tr>
<tr>
<td></td>
<td>after the futures contract month, 9:00 a.m.</td>
</tr>
<tr>
<td>Contract Months</td>
<td>HDD: Nov, Dec, Jan, Feb, Mar plus Oct and Apr</td>
</tr>
<tr>
<td>Settlement Procedure</td>
<td>Daily Settlement Procedures for Monthly HDD Futures</td>
</tr>
<tr>
<td></td>
<td>Final Settlement Procedures for Monthly HDD Futures</td>
</tr>
<tr>
<td>Position Limits</td>
<td>All months combined: 10,000 contracts</td>
</tr>
<tr>
<td></td>
<td>See CME Rule 42102.D.</td>
</tr>
<tr>
<td>Ticker Symbol</td>
<td>Calgary = A2 Edmonton = A4 Montreal = A5 Toronto=A7 Vancouver = A8 Winnipeg = A9</td>
</tr>
<tr>
<td>Exchange Rule</td>
<td>These contracts are listed with, and subject to, the rules and regulations of CME.</td>
</tr>
</tbody>
</table>

Following sections. Temperature changes at a given location, on the one hand, and travel times along a given path, on the other, both share similar stochastic patterns. Similar to farmers, travelers could usefully be insured against the economic costs of low-quality traffic service. This insurance can be generated by using financial derivatives based 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’ route choice, by which their true time cost caused by traffic delays can be
reduced.

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

1. The payoff in the transportation system is good quality of service (QOS): $T_{good}$

2. The derivative payoff is $-P(T)$, as in Figure 5.1.

![Figure 5.1: Travelers and travel time (quality of service) protection - good scenario: Good traffic conditions lead to low High Congestion Days. Therefore, HCD call option is out of money and there is no payoff to the traveler.](image)

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

1. His payoff in the transportation system is bad QOS $T_{bad}$

2. His derivative payoff is $G - P(T)$, where $G$ is in proportion to the experienced extra travel time from a predefined level $K$
Based on the two scenario analyses above, comparisons of traditional congestion pricing methods and travel time derivatives are given. Traditionally, there are two categories of congestion pricing schemes: static road toll and dynamic road toll (toll by time of day and hence by congestion levels).

With static road tolls, the traveler pays a fixed premium/toll $P$ to use the road, as in Table 5.3. The toll is constant no matter when the traveler enters the link. With dynamic road tolls, the traveler pays a fixed amount $P$ when the road has less favorable conditions for travel (usually the prices are set higher during rush hours) and pays nothing when the road has favorable conditions for travel, as in Table 5.4.

With travel time derivatives the traveler’s payment $P$ increases continuously in tandem with expected traffic conditions, which allows the traveler to benefit from compensation payoff $G$, which is set in proportion to the quality of service received, as in Table 5.5. This comparison shows that the road tolls charged through 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 exacted by poor quality of traffic service in the future.

Table 5.3: 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 5.4: 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 5.5: Payoff of a travel time derivatives. $P$ denotes its price, $ET$ is the expected travel time and $K$ is strike price of the derivative contract. The formula describes payoff as a function of realized travel time and the strike price; the price of the derivative contract is not zero but is calculated in proportion to market participants’ expectations regarding future traffic conditions.

<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>$-P(T)$</td>
</tr>
<tr>
<td>Bad traffic</td>
<td>$T_{bad}$</td>
<td>$\alpha(T_{bad} - K) - P(T)$</td>
</tr>
</tbody>
</table>

In this way, travelers’ tolls are based on expected future traffic conditions. If future traffic conditions are expected to be good (bad), the potential future loss is less (more), the payoff is less, and the toll will be less (more). Therefore, travelers could consider taking alternate routes to avoid congestion and reduce travel costs. This flexibility in payment makes the prices of travel time derivatives effective predictors of future travel times. Travelers can forecast the travel time of a path by researching the prices of the corresponding path. This would change traveler behaviors and help them to reduce real-time costs due to traffic delays.

To provide a more practical example, a U.S. Congestion Day index futures contract can be defined in a similar fashion as the Canadian Degree Days index HDD futures contract displayed in Table 5.6. In this contract, the average daily travel time (which can be measured using certain sampling schemes) minus a predefined travel time value floored at zero is
summed over a given calendar month for several urban transportation routes in the U.S. Suppose the buyer purchases the contract at price $P$; initially, if travel times on the specified urban routes are generally higher and this sum is potentially higher in a given month, the price of the contract would increase. In this case, the buyer profits from the price increase and increased payoff $G - P$ to compensate potential losses due to high travel time. This contract can be designed and traded in a similar fashion as that for weather derivatives. Note that related mechanisms are discussed in more detail in the subsequent sections, beginning with more basic products.

Table 5.6: US Congestion Days Index (CDD) Futures traded on Chicago Mercantile Exchange

<table>
<thead>
<tr>
<th>Contract Size</th>
<th>US $20 times the respective CME USA Congestion Days Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index Product Description</td>
<td>Congestion Degree Days (CDD) for U.S. Cities</td>
</tr>
<tr>
<td>Measurement definition</td>
<td>The travel time in a particular city is reported based on a specific route: along Fifth Avenue between 59th st and Washington Square, New York along Michigan Avenue between South Lake Drive and Roosevelt Road, Chicago along Sunset Boulevard between Prospect Ave and Harbar Highway, Los Angeles along Main Street between Bissonnet St and Commence St, Houston</td>
</tr>
<tr>
<td>Pricing Unit</td>
<td>US Dollars (US $) per index point</td>
</tr>
<tr>
<td>Tick Size (minimum fluctuation)</td>
<td>1 index point (= US $ 20 per contract)</td>
</tr>
<tr>
<td>Trading Hours (All times listed are Central Time)</td>
<td>CME Globex (Electronic Platform) SUN 5:00 p.m. - FRI 3:15 p.m. Daily trading halts 3:15 p.m. - 5:00 p.m.</td>
</tr>
<tr>
<td>Last Trade Date/Time</td>
<td>Fifth Exchange business day after the futures contract month, 9:00 a.m.</td>
</tr>
<tr>
<td>Position Limits</td>
<td>All months combined: 10,000 contracts See CME Rule 42102.D.</td>
</tr>
<tr>
<td>Exchange Rule</td>
<td>These contracts are listed with, and subject to, the rules and regulations of CME.</td>
</tr>
</tbody>
</table>
5.1.3 Travel time derivatives can serve as alternatives to insurance against traffic service quality

Different from traditional congestion pricing schemes, travel time derivatives not only impose a road toll but also provide a corresponding payoff to travelers. For typical travelers, the payoff is larger when the experienced travel time is high, which is similar to insurance against bad quality traffic. Ideally, to hedge against the risk of experiencing high travel times, travelers can purchase insurance contracts; however, travel time derivatives have special characteristics over insurance contracts:

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 be reassured by its significant regulatory requirements. Insurance, however, lacks some of the flexibility associated with derivative-based solutions. The comparison shown in Table 5.7 summarizes some of the potential differences between insurance and derivative-based travel time risk management products.

Besides the general characteristics above, travel time derivatives can provide more flexible protection than insurance products. To enable such flexible protections, there are two kinds of payoff functions for travel time derivatives, which differ in the time span covered by underlying travel time measures.

1. The first type of travel time derivatives can be based on the instantaneous travel time measures at given locations in the future. As the graph shows, for some travel time derivatives of this type, the market participants believe the economic loss due to high travel time is going to be lower or less volatile if and only if the price of the derivative contract is higher. Therefore, the price of travel time derivatives indicates economic loss due to travel time in the short term and travelers can select the paths with higher prices
Table 5.7: Comparison between insurance on travel time and travel time derivatives

<table>
<thead>
<tr>
<th></th>
<th>Insurance on travel time</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 valued using mark to market and recoveries generally create a taxable gain</td>
</tr>
<tr>
<td>Liquidity</td>
<td>Illiquid, it is effectively a buy and hold instrument. Coverage is non-cancelable</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 element or combination of elements mainly for the buyer</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>Economic cost caused by traffic delay</td>
<td>Economic cost caused by traffic delay with more flexible payoff and leverage</td>
</tr>
</tbody>
</table>

when making routing decisions. This category of travel time derivative is demonstrated in Figure 5.3.

2. The second type of travel time derivative can be derived based on the cumulative travel time measures for a long time period in the future. For some products of this type, the economic loss due to congestion is less in the coming year if and only if the price is higher. Therefore, the price of travel time derivatives indicates economic loss associated with long-term traffic conditions and a traveler can then plan to use alternative paths or use public transportation if derivative prices are generally low. When considering a long-term transportation plan, travelers can check the price of such derivatives. This category of travel time derivatives is demonstrated in Figure 5.4 and the CDD index option cited in Table 5.6 belongs to this category.
Travel time put option:
If the experienced travel time on the
link is lower than 4 minutes, the
buyer can get $2*(4-T_t)$

Routing choice

Put option price $5$

Put option price $6$

Less economic loss due to congestion

Put option payoff more
$2*(4-T_t)$

Put option is more valuable

In summary, as flexible payoff functions can be defined, travel time derivatives can provide a payoff according to travelers’ experienced travel time in future to reduce potential economic
costs due to traffic delays, which is more advantageous over traditional insurance.

5.1.4 Travel time derivatives can provide protections to firms or businesses whose profit is related to traffic conditions

Travel time derivatives are useful for businesses whose profits are related to traffic conditions. Firms can be distinguished into two categories based on whether or not they benefit from good traffic conditions.

For cargo transportation companies, profit are derived from daily transportation service. If overall traffic conditions are bad, there are more delays for trucks, therefore overall service to clients is worse and operation costs are increased. The profits of the company are reduced and it can lose competitive advantage in the marketplace. With travel time derivatives, the company can invest in travel time derivatives and hedge potential losses due to bad traffic conditions in the future. If the overall traffic conditions are good, the service is good and the firm can obtain potentially increasing profits. The cost is just a premium which is used to purchase derivative contracts, if overall traffic conditions are poor the following year.

On the other hand, some firms would profit if traffic conditions are worse, including toll road owners, public transportation companies, companies offering alternative transportation services, etc. When traffic conditions are good, fewer travelers will select toll roads, therefore toll road owners tend to have less profit when overall travel time is low, i.e., they are hurt by good traffic conditions. Similarly, fewer people would select public transportation or alternative transportation including trains over driving themselves when traffic conditions are good, therefore related firms are also hurt from good traffic conditions. Travel time derivatives provide methods for them to hedge their risks when traffic conditions are good.

As different businesses have different payoffs based on the performance of traffic systems, they have incentives to hedge their risk over the market. Different risk appetites between
different market participants lead to diversified trading activities. The firms who do not profit from good traffic conditions can trade against those that benefit from good traffic conditions. These hedging activities provide strong incentives for introducing travel time derivatives.

5.1.5 Travel time derivatives can diversify risk for financial markets

In portfolio theory, the diversification of investments into different asset classes is a recommended practice. For a given set of investments, the lower the correlation between assets, the less the total risk, as in 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 classes are lower than the correlations among 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 the traffic system is shown in Figure 5.5.

Figure 5.5: Alternative risk transfer between the transportation and financial industries

Moreover, travel time derivatives can hedge the risk in other special markets. For example, since weather conditions are correlated with experienced travel time, travel time derivatives will be a good hedging tool for investors who invest in the weather derivative
market. Likewise, CO2 emission levels have been traded in the market, and since CO2 emissions levels are highly correlated with the performance of the traffic system, travel time derivatives will be a good tool to hedge against risk for the investors in the CO2 emissions market.

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 sides of the market. Type B: When the specified travel time is expected to be high, a leveraged reward is available to the buyer. Type H: When the specified travel time is expected to be 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 B are the participants who benefit (are hurt) from good (bad) traffic conditions; buyers of Type H are the participants who hurt (who benefit) by good (bad) traffic conditions. The potential participants and their roles are summarized in Table 5.8.

As a newly introduced market, the market making of travel time derivatives is critical and challenging, Panayides & Charitou (2004). Market makers match buyers with sellers to enable smooth trading activities; they also provide liquidity to the market by holding short term positions; due to their efforts, the market price of financial derivatives is determined and maintained. To make a profit, market makers quote both a buy and a sell price in derivative contracts, which differ on the bid-offer spread, and they use hedging strategies to control their risk. Investment banks are typical market makers for financial derivatives. With appropriate pricing methods and suitable trading exercise, the total profit for investment banks is positive, which motivates them to operate the business, following the general
mechanism in the current financial derivative markets. On the other hand, the counter-parties to the market parties will seek protection from the market and their total profit are negative, which can be interpreted as the cost that they pay to hedge risk due to travel time uncertainties in the future. Important factors that should be considered for market making include the following:

1. Market microstructure will be crucial in determining the operation of the market. There are numerous links/paths in transportation networks, and a large number of derivative contracts can be written based on their experienced travel time. Conversely, when this new market begins operating, the trading activity will be low. Therefore, the
market may encounter liquidity issues, where smaller trading amounts can drive the prices, and thereby increase price volatility. Several measures can be taken to minimize potential liquidity issues, including restricting the number of products on the market, building temporary liquidity reserves, and so forth. Related discussions for other types of derivatives can be found in MacKenzie & Millo (2003), Dubil (2007), Wolfers & Zitzewitz (2006) and 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 estimated 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. In the transportation industry, the percentage needs to be estimated and a larger percentage means more potential market participants. There is significant amount of research on the cost of travel time, which can roughly be measured in the annual revenue raised by road tolls. For example in Sugiyanto et al. (n.d.), it is stated "there were $63.2 billion in actual congestion costs in the 85 urban areas [in the U.S.] in 2002. It is estimated that public transportation saved an additional $20 billion in congestion costs for this group." The billion-dollar congestion costs imply a significant impact of traffic delays to individual travelers, which motivates them to hedge their risks. More profoundly, the companies, the profits of which are changed by traffic service, may have more freedom to purchase and trade travel time derivatives, which should be further estimated. The sum of all related profit and costs add to the potential for travel time derivative markets, Harford (2006) and Nash & Sansom (2001).

3. A healthy market for travel time derivatives also requires appropriate legal regulations. As observed in traditional financial markets, malicious insider trading or market manipulation can occur if the participants know additional information, which may change future travel time through illegal sources. Moreover, unlike traditional un-
derling assets, travel time is the aggregate effect of traveler behavior nearby, so the potential of travel time derivatives in changing traveler behavior may lead to possibility of manipulating future travel times and hence price of travel time derivatives through intentional routing guidance. To prevent such undesired cases and regulate the travel time derivatives market, appropriate policies or laws should be issued.

Based on the settings above, a market can potentially be established for travel time derivatives. The major products for this potential market are presented in the following section.

5.3 Design of Travel Time Derivatives

In general, the underlying asset of travel time derivatives is some measure of future travel times. Investors receive cash flows in proportion to travel time related measures as their payoff and in order to purchase such derivatives, investors have to pay a price. Based on the introduction to travel time derivatives in previous sections, several classifications can be applied to travel time derivatives:

1. By contract type, travel time derivatives can be classified into futures, options, etc;
2. By measurement places used in the derivative, they may be classified as derivatives based on one path or several paths, which is hence based on an index of travel time;
3. By the time span of the underlying travel time measures, travel time derivatives can be classified as instantaneous or long term based;
4. By whether the buyer gets a payoff when traffic is good or bad, travel time derivatives can be classified as Beneficial (B) versus Hurting (H).
This section continues to introduce more mathematics for describing the travel time derivatives.

### 5.3.1 Standard travel time measurements

In order to define products based on travel time measures, a standard measurement of travel time must be defined.

**Definition 5.3.1** A *standard measurement of travel time on a specific path and time is the average travel time reported from specific travel time data providers on that path within a small time interval around that time.*

**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 a pre-defined time of day. The daily mean of a standard measurement plan is the mean value of such measurements.*

In the above definitions of travel time derivatives market products, all travel time values for a given time in a day are based on a standard measurement, and all travel time values for a given day are based on the standard measurement plan. Each observation is selected by specifying the arrival time to the path. To provide adequate measurements of travel time to support the trading and pricing of travel time derivatives, a loop detector is recommended. The reasons for this choice include:

1. Pricing of travel time derivatives should be based on periodical travel time measurements so that classical stochastic analysis can be used to model travel time and corresponding models can be calibrated. As is summarized in Section 2.1, site-based measurement such as loop detectors can yield such periodical estimations based on occupancy and flow, and hence data from them are suitable for the study of travel time derivatives.
2. Pricing of travel time derivatives should be based on average travel time on the given path to prevent individual measurement error from introducing instability in derivative prices. Loop detectors yield estimation of average travel time based on occupancy and flow, which satisfies this requirement and relieves related concerns.

5.3.2 Standard spatial travel time index and equivalent return rate

First, a spatial travel time index should be designed. This index can be a weighted average of the latest travel time in downtown areas of major cities in the U.S. A 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 Spatial Travel Time Index**

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

where \( T_i \) is the travel time in selected places within a given area.

For example, a Spatial Travel Time Index could be constructed as the weighted average of the realtime travel time in downtown New York (a section of Fifth Avenue), downtown Chicago (a section of Michigan Avenue), downtown Los Angeles (a section of Sunset Boulevard), 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 the United States, which shows the national service level of urban traffic systems. The return of this index, volatility, and its sharpe ratio can then be used as references when pricing travel time derivatives. Note the Sharp ratio is the ratio between access return and volatility of
this derivative, and excess return is the extra return of this index relative to the risk-free bond.

Travel time indexes can also be defined for a given area, if taking the average travel time in the major avenues of Manhattan can be used to indicate the general traffic conditions on Manhattan island. Travelers can trade over such indexes to compensate for the waste of time and economic loss due to traffic delays.

5.3.3 Design of derivative products on travel time

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

The simple derivative based on 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 its standard measurement entering at 10 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 \ast \ast(70 - 60); if the experienced travel time is lower than 60 minutes, the buyer gets nothing.
Type 2 Futures on congestion-days

After establishing basic options for a specific link at a given future time point, the futures 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 \( LCD \) 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 on all the days in that period, given a fixed number of measurements.

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

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

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

, 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 contin-
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 reflect market participants’ expectations 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 days options are 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 greater than 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 \( 10 \times (S - K) \) where 10 is the leverage ratio; if not, the buyer receives nothing. This is an example of an HCD call option. This pair of products leverages the buyer’s gain according to long term traffic status in the future. Compared to the futures, the options provide further leverage, and the buyers can get more return/loss if traffic conditions change. 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 contract 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 given 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 derivatives on travel time

To price travel time derivatives, the underlying travel time series is first selected as a continuous time mean reverting process with trend and seasonality adjustments. Due to variation
in traffic conditions across different links, the travel times on different links may be fitted to models with different orders. To provide such flexibility, a family of alternative models are introduced in this section. Model selection is conducted based on empirical data according to statistical principles and pricing methods are discussed.

5.4.1 Alternative stochastic processes for modeling 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, particularly when calculating the price for interest rates derivatives and weather derivatives, Hull & White (1990).

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

**A mean-reverting process driven by Brownian motion**

The travel time process \(T_t\) 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)

where

\(b_t\) is the trend and seasonality component, \(\sigma\) 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 to the S.D.E is

\[ T_t = b_t + (T_0 - b_0)e^{-\int_0^t a_s ds} + e^{-\int_0^t a_s ds} \int_0^t e^{\int_0^s a_u du} \sigma_u dB_u \]

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:

\[ \mu_t = T_0 e^{-\int_0^t a_s ds} + b_t - b_0 e^{-\int_0^t a_s ds} \]

and

\[ \sigma^2_t = e^{-2\int_0^t a_s ds} \int_0^t e^{2\int_0^s a_u du} \sigma_u^2 du^2 \]

By equating 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 closely related to the usual time series model with i.i.d normal noise.

If defining \( X_t = b_t - T_t \), the residual process after removing trend and seasonality component in the model above is subject to the following S.D.E:

\[ dX_t = -a_t X_t dt + \sigma_t dB_t \]
A more general version is to define \( Y_t = \int_0^t X_t dt \) and further modeling of the travel time process is subject to the following equation:

\[
T_t = b_t + Y_t
\]

This generalized continuous time process can then be used to approximate the case in which the error term after removing trend and seasonality factor is an integrated process of order 1, such as ARIMA\((p, 1, q)\).

**Continuous autoregressive moving average (CARMA) process**

More generally, a continuous-time Gaussian autoregressive and moving average process (CARMA) can be used to fit the travel time process. The process \( X_t \) in previous section is the simplest CARMA\((1, 0)\) process. By definition, a CARMA process \( Y_t \) is defined symbolically to be a stationary solution of the stochastic differential equation:

\[
a(D)C_t = b(D)B_t
\]

with coefficients \( a_1, a_2, \ldots, a_p \) and \( b_0, b_1, \ldots, b_q \) for \( p > q \)

\[
a(z) = a_0 z^p + a_1 z^{p-1} + \cdots + a_p
\]

\[
b(z) = b_0 + b_1 z + \cdots + b_q z^q
\]

The operator \( D \) denotes differentiation with respect to \( t \), which is in the formal sense for the Brownian Motion Brockwell (2001). Due to the fact the derivative of \( B_t \) does not exist with probability 1, the process is represented further in the following state space representation.
\[ C_t = BX_t \]

and

\[ dX_t = AX_t dt + edB_t \]

with

\[ B = [b_0, b_1, \ldots, b_q, 0, \ldots, 0] \]

\[ X_t = (X_{t,0}; \ldots; X_{t,p-1}) \] and the first \( p - 1 \) element of \( X_t \) is defined as

\[ X_{t,j} - X_{0,j} = \int_0^t X_{u,j+1} du \]

\[ A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \cdots & a_1 \end{bmatrix} \]

\[ e = [0, 0, 0, 0, 0, 0, 1]^T \]

when \( p = 1 \), \( A \) is defined as \(-a_1\)

By applying the multidimensional Ito Formula, the solution to the S.D.E above is below:

\[ X_t = e^{At}X_0 + \int_0^t e^{A(t-u)}edB_u \]
The process above is stationary and well defined when \( p > q \). If \( p \leq q \), the covariance function does not exist and the spectral density does not exist. However, the process is further defined as a general random process (GRP) by Brockwell & Hannig (2010) for the study of the derivatives of \( \text{CARMA}(p,q) \) processes. In the following sections, the thesis focuses on the continuous time derivative pricing based on \( \text{CARMA}(p,q) \) process for the \( p > q \) case, while it is recommended that Monte Carlo simulation be conducted based on the fitted discrete time models to approximate the derivative prices for if \( p \leq q \).

5.4.2 Modeling travel time processes

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

\[
T_t = r_t + s_t + Y_t, \quad t = 0, 1, 2, \ldots
\]

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:

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

(b) The weekly part is:

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

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

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

However, the trigonometric functions are selected as the basis function, because 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 and the type of process is determined based on empirical data. For the data in this section, it is defined that \( C_t = dY_t \) and \( C_t \) is modeled as a CARMA\((p, q)\) process.

The data set is the travel time data set from California PEMS. Data were collected in November 2012, on the 80-E path between 80-E/Cummings Skwy and 80-E/Maritime Academy, as presented in Figure 5.6. The travel time data are gathered from loops captured every 5 minutes. The calibration process is as follows:

1. A 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 \( a = 2.23 \). This part is shown in the Figure 5.7.
2. The seasonal model is studied using 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 5.7.

| Table 5.9: Seasonal effects in the 80E data |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Weekly trend    | $k$             | $a_1$           | $a_2$           | $a_3$           | $b_1$           | $b_2$           | $b_3$           | $T$             |
|                 | 0.0927          | -0.0135         | 0.0794          | -0.0902         | 0.0477          | -0.0873         | -0.0228         | 2064            |
| Daily trend     | 0.0002          | 0.0240          | 0.1143          | -0.0795         | -0.1509         | 0.0127          | -0.0204         | 288             |

3. After removing the trend and seasonal part, the residual is used to estimate the CARMA$(p, q)$ process. In the literature, the methods to estimate CARMA$(p, q)$ models can be based either directly on the continuous-time process or on a discretised version. The latter relates the continuous-time dynamics to a discrete time ARMA process. The advantage of this method is that standard packages for the estimation of ARMA processes may be used in order to estimate the parameters of the corresponding CARMA process. However, not every ARMA$(p, q)$ process is embeddable in a CARMA$(p, q)$ process. Brockwell and collaborators devote several papers to the
embedding of ARMA processes in a CARMA process, Brockwell (1995) and Brockwell & Brockwell (1999). In the study below, this approach is employed by assuming an appropriate class of CARMA processes to work with.

Following the intuition above, the residual is identified as the following ARIMA(1,1,0) model (Table 5.10). A Box-Ljung test suggests that the p-value = 0.1928. This model
describes the time series well.

![Standardized Residuals](image1)

![ACF of Residuals](image2)

![p values for Ljung–Box statistic](image3)

Figure 5.8: 80E ARIMA model

Table 5.10: Model selection for 80E data

<table>
<thead>
<tr>
<th>ARIMA model</th>
<th>Log Likelihood</th>
<th>AIC</th>
<th>p-value of Box-Ljung test</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(1, 0, 0)</td>
<td>14467.36</td>
<td>-28928.72</td>
<td>0.0035</td>
</tr>
<tr>
<td>ARIMA(1, 1, 0)</td>
<td>14438.3</td>
<td>-28872.61</td>
<td>0.1928</td>
</tr>
<tr>
<td>ARIMA(1, 1, 1)</td>
<td>14438.31</td>
<td>-28870.61</td>
<td>0.1322</td>
</tr>
<tr>
<td>ARIMA(2, 1, 1)</td>
<td>14438.31</td>
<td>-28868.61</td>
<td>0.0840</td>
</tr>
</tbody>
</table>

The model is described as the ARIMA(1, 1, 0) model by comparing the test results, and additional variance modeling can be conducted to describe the volatility process.

4. The estimation is demonstrated for a selected path above, in which the model follows ARIMA(1, 1, 0). However, travel time processes for different paths may appear to fit ARIMA models with different orders (Table 5.11). To address such variations, the
pricing of travel time derivatives is discussed considering all possible orders of ARIMA models. Particularly, distinct treatments are applied conditioned on whether \( p > q \) holds or not. In more detail, the spectral density for a CARMA\((p, q)\) process is defined as

\[
f(\omega) = \frac{\sigma^2 \beta(i\omega) \beta(-i\omega)}{2\pi \alpha(i\omega) \alpha(-i\omega)}
\]

The stationarity condition of the CARMA processes requires the roots of the polynomial \( \alpha(z) \) to have negative real-parts and that \( p > q \), Tómasson (2011). Below the estimation of CARMA\((p, q)\) processes are conducted based on fitted ARMA model considering different relations of \( p \) and \( q \):

<table>
<thead>
<tr>
<th>Path Name</th>
<th>Model</th>
<th>p-value of Box-Ljung test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cummings Skwy - Maritime Academy</td>
<td>ARIMA(1, 1, 0)</td>
<td>0.1928</td>
</tr>
<tr>
<td>Berkeley-Davis</td>
<td>ARIMA(2, 1, 1)</td>
<td>0.0503</td>
</tr>
<tr>
<td>Lincoln-Davis</td>
<td>ARIMA(1, 1, 1)</td>
<td>0.3419</td>
</tr>
</tbody>
</table>

(a) If \( p > q \), the discrete solution for the CARMA\((p, q)\) state space representation can be given as the following equation

\[
x_{t+\delta t} = e^{A\delta t} x_t + \int_t^{t+\delta t} e^{t+\delta t-u} e_p \sigma_u dB_u
\]

The noise term is of variance \( \int_0^{\delta t} e^{Au} e_p e^{At} du \). If first order expansion of the matrix exponential \( e^{Au} \) is taken, a set of formulas can be derived to map the parameters of the CARMA\((p, q)\) processes \((\alpha_n, \beta_n)\) approximately to the parameters of the corresponding discrete ARMA\((p, q)\) models \((a_n, b_n)\), Härdle & Cabrera (2012). Correspondingly, the parameters of the CARMA\((p, q)\) processes can be backed out using estimated ARMA model parameters: Assuming the grid size is \( h \), the mapping for the auto regression parts of the model in the first three orders
is displayed in Table 5.12.

<table>
<thead>
<tr>
<th>Model</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAR(1)</td>
<td>$a_1 = 1 - \alpha_1 h$</td>
</tr>
<tr>
<td>CAR(2)</td>
<td>$a_1 = 2 - \alpha_1 h$</td>
</tr>
<tr>
<td></td>
<td>$a_2 = \alpha_1 h - \alpha_2 h - 1$</td>
</tr>
</tbody>
</table>

The estimation of the moving average parameters can be based on auto correlation function Espen Benth et al. (2012) and a least absolute deviation algorithm can be used to estimate the MA part based on the empirical and theoretical autocorrelation functions of the CARMA processes.

$$\gamma(s) = b' e^{A|s|} \Sigma b$$

where

$$\Sigma = \int_0^{\infty} e^{Au} e_p e_p^t e^{A^t u} du = -A^{-1} e_p e_p^t$$

In this representation, $A^{-1}$ is the inverse of the operator $A : X \rightarrow AX + XA'$, Pigorsch & Stelzer (2009) and Barndorff-Nielsen et al. (2007).

With these mapping formulas, the coefficients of the continuous ARMA processes can be solved using the parameters of the fitted discrete ARMA model. These mappings ignore higher order terms and can hence be applied to low order CARMA processes for engineering purposes. These values can also be used as the initial values for more precise estimation of the CARMA parameters.

To obtain such precise estimations, a Kalman filter is used to further extract the unobserved states of the CARMA $(p, q)$ processes based on the state-space representation and parameters are estimated by minimizing the error between observed and estimated output series of the state space model. In more detail, given
a travel time series whose trend and seasonality are removed, first order differentia-
tion is taken to remove the integrated part of the model. The remaining process
\(X_t\) is then processed using a Kalman filter to extract the unobserved higher order
states in the CARMA process, Brockwell et al. (2011) and Schlemm & Stelzer
(2012). The filtered model produces an output series \(X_t'\) and the mean square
error of \(X_t\) and \(X_t'\) are minimized to yield the minimum mean-squared-error lin-
ear predictors for the parameters of CARMA\((p, q)\), which is found by numerically
minimizing the sum of squares. To obtain a warm start, the parameters obtained
from the previous first order approximation are used as initial values. The esti-
mated parameters are displayed in Table 5.13 and the estimated processes need
to be calibrated further to the market to determine the market value of risk \(\theta\),
using the methods discussed in next section.

<table>
<thead>
<tr>
<th>Path</th>
<th>Parameter</th>
<th>ARMA</th>
<th>Initial Value</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cummings Skyway - Maritime Academy</td>
<td>AR-1</td>
<td>-0.0326</td>
<td>0.2065</td>
<td>0.2512</td>
</tr>
<tr>
<td>Berkeley-Davis</td>
<td>AR-1</td>
<td>0.9646</td>
<td>0.2071</td>
<td>0.1780</td>
</tr>
<tr>
<td></td>
<td>AR-2</td>
<td>-0.1275</td>
<td>0.03258</td>
<td>0.0685</td>
</tr>
<tr>
<td></td>
<td>MA-1</td>
<td>-0.6575</td>
<td>0.1356</td>
<td>-0.6296</td>
</tr>
</tbody>
</table>

(b) For the processes corresponding to the case of \(p \leq q\), for example, the ARIMA\((1, 1, 1)\)
on path Lincoln-Davis, it is recommended that Monte Carlo method based on the
discrete ARIMA model be used to price the travel time derivative after the pricing
principles and risk neutral measure are specified. Although CARMA \((p, q)\) has
recently been extended into generalized random processes for cases when \(p \leq q\,
Brockwell & Hannig (2010), the processes under such settings are not well defined
in the usual sense. In terms of rationale, the general random process yields a way
to represent the discrete approximation of the high order derivatives of Ornstein-
Uhlenbeck process with respect to selected linear functionals, but this representa-
dion does not lead to typical continuous time stochastic models. The application
of a generalized random process for pricing travel time derivatives is still contingent upon further studies of such processes and hence the Monte Carlo method is a reasonable approximation for modeling and pricing travel time derivative.

In summary, travel time processes are fitted to discrete ARMA($p, q$) models with trend and seasonality adjustment first. Then the fitted models are converted into stable CARMA($p, q$) processes for further calculation when $p > q$. When $p \leq q$, it is recommended that Monte Carlo simulation based on the discrete ARMA models be used to price travel time derivatives. In the following sections, potential pricing schemes are discussed based on fitted CARMA($p, q$) processes.

5.4.3 Risk neutral pricing in an incomplete market

To obtain a general pricing expression for travel time derivatives, risk neutral pricing principle in incomplete market conditions is employed in this section, deploying the method in Benth & Benth (2007).

Risk neutral representations

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. A sub-family of probability measures $Q$ is specified using the Girsanov transformation: assume $\omega_t$ is a real-valued measurable and $\frac{\omega_s}{\sigma_s}$ is a bounded function. The stochastic process:

$$Z^\omega(t) = \exp\left(\int_0^t \frac{\omega_s}{\sigma_s} dW_s - \frac{1}{2} \int_0^t \frac{\omega_s^2}{\sigma_s^2} ds\right)$$
is the density process of the probability measure $Q$. Under $Q$, the process

$$dB_t = dW_t - \frac{\omega(s)}{\sigma(s)}dt = dW_t - \theta_t$$

is a Brownian motion and $\theta_t$ is called the market value of risk process. Based on this measure change, the dynamics of the underlying process under the risk neutral measure are given by the following S.D.E:

$$dT_t = dc_t + \omega(t)dt + \alpha_t(c_t - T_t)dt + \sigma_t dW_t$$

and the solution to this S.D.E is then

$$T_t = c_t + (T_0 - c_0)e^{-\int_0^t \alpha_s ds} + e^{-\int_0^t \alpha_s ds} \int_0^t e^{\int_u^t \alpha_s ds} \omega_u du + e^{-\int_0^t \alpha_s ds} \int_0^t e^{\int_u^t \alpha_s ds} \sigma_u dW_u$$

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 different travel time derivative contracts in continuous time are calculated as the discounted expectation of their corresponding payoff function under this measure, which is given below:

The price of a HCD futures contract can be calculated as:

$$F_{HCD}(t_1, t_2, T, t) = E_Q \left\{ \int_{t_1}^{t_2} \max(T_s - T, 0) ds | F_t \right\}$$
The price of a LCD futures contract can be calculated as:

\[ F_{LCD}(t_1, t_2, T, t) = E_Q \left\{ \int_{t_1}^{t_2} \max(T - T_s, 0) ds | \mathcal{F}_t \right\} \]

Assuming a constant interest rate, the price of a HCD call options contract can be calculated as:

\[ C_{HCD}(t_1, t_2, T, t, K) = e^{-rt} E_Q \left\{ \max(\int_{t_1}^{t_2} \max(T_s - T, 0) ds - K, 0) | \mathcal{F}_t \right\} \]

and the price of a LCD call options contract can be calculated as

\[ C_{LCD}(t_1, t_2, T, t, K) = e^{-rt} E_Q \left\{ \max(\int_{t_1}^{t_2} \max(T - T_s, 0) ds - K, 0) | \mathcal{F}_t \right\} \]

The price of a CTT futures contract can be calculated as:

\[ F_{CTT}(t_1, t_2, t) = E_Q \left\{ \int_{t_1}^{t_2} T_s ds | \mathcal{F}_t \right\} \]

Assuming a constant interest rate, the price of a CTT call options contract can be calculated as

\[ C_{CTT}(t_1, t_2, t, K) = e^{-rt} E_Q \left\{ \max(\int_{t_1}^{t_2} T_s ds - K, 0) | \mathcal{F}_t \right\} \]

Since travel time is neither a tradable nor storable asset, the derivatives contracts cannot be hedged using travel time itself in the financial markets, and the market of the travel time derivatives is therefore incomplete. Under such incomplete markets, the risk neutral measure is not unique. To obtain the prices, the risk neutral measure should first be specified considering the characteristics of the incomplete market setting. Moreover, the expectation can be calculated using two alternative methods: it can be calculated using Monte Carlo simulation of the underlying processes under the specific risk neutral measure, the average
discounted payoff of the financial derivative in all paths yield the price of the contract; alternatively, some explicit formulas can be obtained by considering the property of the discounted price processes under the risk neutral measure.

**Determination of the risk neutral measure**

The incompleteness of the travel time derivative market requires the estimation of the market price of risk (MPR) for pricing and hedging travel time derivatives. The market price of risk adjusts the underlying process representing travel time so that the implied price is arbitrage free. As it is stated in Section 5.4.3, $B_t = W_t - \theta_t$ is a Brownian motion under the risk neutral measure. Since the underlying asset is not tradable, there is no unique risk neutral measure. The drift of the asset price process is the view of the trader about the growth of the process. There are different ways of specifying this measure and some of them are discussed below:

1. Market value of risk can be inferred from market traded products. Härdle & Cabrera (2012) suggests inferring the market price of risk (MPR) from traded CAT futures by minimizing the mean square error between the modeled contract prices with the market traded prices. Once the MPR for temperature futures is known, it is used to price other derivatives. According to Meyer-Brandis & Tankov (2008), within incomplete markets, there may exist many equivalent risk-neutral measures; it is the job of the market as a whole, via trading of derivatives, to decide which measure prevails at any one given point in time. A class of equivalent martingale measures can be identified which maintains the structure of real-world dynamics for asset prices. These measures can then be used to obtain forward prices and value spread option. Moreover, the differences in pricing measures leads to risk premium which can be calibrated using market prices: Lucia & Schwartz (2002) showed a negative market price of risk associated to the non-stationary term in their two-factor models, when analyzing data
from energy market. In the proposed two-factor model, where the non-stationary term is a drifted Brownian motion, the negative market price of risk appears as a negative risk-neutral drift. Benth et al. (2008), showed that using the certainty equivalence principle that the presence of jumps in the spot price dynamics will lead to a positive risk premium in the short end of the futures curve. Bessembinder & Lemmon (2002) explain the existence of a positive premium in the short end of the futures market by an equilibrium model.

In the context of travel time derivatives, the choice of $\theta_t$ uniquely determines the equivalent martingale measure under which derivatives pricing is performed. One way of defining the market price of risk is to extrapolate from option prices. This technique resembles recovery of the implied volatility in the Black-Scholes model. A chosen objective function can be minimized to find $\theta_t$, such as the mean absolute percentage error between the market and model option prices. The market prices can be chosen as averages of the bid and ask offers and options with different strikes can be used to calibrate $\theta_t$ for a given day. Alternatively, calibrating it to futures prices is also feasible. The procedure is analogous to that used with options, and the model can be calibrated to one futures price. Futures have more liquidity than options and hence allow for a more frequent and precise calibration of $\theta_t$. The value of $\theta_t$ is subject to the incentive for hedging on the demand side relative to the supply side. For a concrete example, the market value of risk process $\theta_t$ can be calibrated by a set of HCD futures contracts based on the common underlying travel time series by minimizing the mean squared difference between modeled price and traded price below

$$\theta_t = \arg\min \left( \sum_i (F_{i,\text{market}} - F_{i,\text{HCD}}(t_1, t_2, T, t, \theta_t))^2 \right)$$

Different contracts can be used to conduct such calibration and contracts with the most liquidity in the market are the best instrument for such purposes.
2. Suitable hedging strategies result in a price, that suggests a risk neutral measure. Different hedging strategies leads to varying derivative prices. The following example demonstrates how to obtain the price of a travel time derivative contract by hedging it with another travel time derivative contract.

Suppose the two derivatives have two 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:

The P&L changes of this portfolio are obtained below through Ito Lemma:

\[
dP = \alpha(\frac{\partial F}{\partial t} dt + \frac{\partial F}{\partial X} dX + \frac{1}{2} \frac{\partial^2 F}{\partial X^2} d<X>_t) + \beta(\frac{\partial G}{\partial t} dt + \frac{\partial G}{\partial X} dX + \frac{1}{2} \frac{\partial^2 G}{\partial X^2} d<X>_t)
\]

The equation further leads to the following relationship if the definition of the underlying diffusion is applied:

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

To make the portfolio risk free, 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

\[
\frac{\alpha}{\beta} = -\frac{\partial G}{\partial X} / \frac{\partial F}{\partial X}
\]
This approach eliminates $ds$ term. Furthermore, the increasing rate should be interest rate $r$:

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

and we have

$$\frac{\partial F}{\partial t} + \frac{\partial F}{\partial X} A(t, X_t) + \frac{1}{2} \frac{\partial F^2}{\partial X^2} B^2(t, X_t) - rF = \frac{\partial G}{\partial t} + \frac{\partial G}{\partial X} A(t, X_t) + \frac{1}{2} \frac{\partial G^2}{\partial X^2} B^2(t, X_t) - rG$$

By dividing the volatility part on both side we have:

$$\frac{\partial F}{\partial t} + \frac{\partial F}{\partial X} A(t, X_t) + \frac{1}{2} \frac{\partial F^2}{\partial X^2} B^2(t, X_t) - rF = \frac{\partial G}{\partial t} + \frac{\partial G}{\partial X} A(t, X_t) + \frac{1}{2} \frac{\partial G^2}{\partial X^2} B^2(t, X_t) - rG = \theta_t$$

where we define $\theta_t$ as the market value of risk.

For an individual travel time derivative with value $F$, its price should satisfy the following P.D.E:

$$\frac{\partial v}{\partial t} + \frac{\partial v}{\partial X} (A(t, X_t) - \theta_t B(t, X_t)) + \frac{1}{2} \frac{\partial v^2}{\partial X^2} B^2(t, X_t) = rv$$

where $0 \leq t < T$ and $x \geq 0$ with boundary conditions:

$$v(t, x) = e^{-r(T-t)}(x - K)^+$$

$$\lim_{x \to -\infty} v(t, x) = 0, 0 \leq t < T, x \geq 0$$

$$v(T, x) = (x - K)^+, x \geq 0$$

(a) 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 P.D.E can be solved and yield a proper price for the options.

(b) If \( Y_t \) is just a mean reverting process we can directly consider \( dT_t = dc_t + k(T_t - c_t)dt + \sigma dB(t) \)

\[
A(t, T_t) = \frac{dc_t}{dt} + k(T_t - c_t) \ B(t, T_t) = \sigma \text{ and the P.D.E can be converted correspondingly. Here, the trend and seasonal parts are considered in the underlying process.}
\]

(c) 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 an Asian call 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 P.D.E are applied to \( X_t \) while the trend and seasonal components are considered in the payoff function \( F \).

(d) The pricing P.D.E yields a price for the derivative contract and hence implies a risk neutral measure. This price corresponds to the strategy in which one hedges the travel time derivative with another similar derivative contract based on the same underlying assets but with different payoff functions. The results imply a risk neutral measure that corresponds to the hedging strategy.

In the example above, the two derivative contracts are assumed to be both liquidly traded and hence the market is completed by introducing a second derivative contract for hedging purposes. This method follows standard pricing schemes for derivatives based on non-tradable underlying assets. Additional care can be taken when such a derivative contract is not available or the liquidity of the derivative market is not adequate. For example, according to Carmona & Danilova (2003), a stock that is potentially correlated with travel time changes can be used to hedge the travel time derivative contracts. The stock of logistic companies or
alternative transportation providers, whose business are related to changes in travel time can be used to hedge travel time derivatives. It is also possible to hedge travel time derivatives with energy derivatives or travel time derivatives, the underlying processes of which are correlated with travel time. The liquidity of the stock market may be better in certain cases than derivative market

3. Some characteristics of the risk neutral measure can be specified based on certain optimality conditions. For example, the minimum entropy measure has been studied in Frittelli (2000), as it can yield reasonable asset prices and there is a huge literature on the use of maximum entropy measure for calibration purposes. In Follmer & Schweizer (1991), a minimal martingale measure based on local variance minimization provides a strategy that penalizes over-hedging. Xu (2006) introduces pricing methods based on suitable risk measures and partial hedging is used when certain risk measures are introduced to control the residual risk at expiration. Such methods can be applied to identify the best pricing measure for pricing travel time derivatives.

These methods may lead to different prices due to the non-tradable nature of travel time. This discussion again shows that the prices of travel time derivatives are subject to the choices of specific hedging strategies, under incomplete market conditions with non-unique risk neutral measures. In the following section, it is assumed that a risk neutral measure can be calibrated using the price of traded derivatives.

**Risk neutral pricing for travel time derivatives**

In this section, pricing P.D.Es are further derived based on the pricing measure that is identified using methods in the previous section. The rationale is that derivative prices are functions of underlying processes and these prices should be martingales under the specified risk neutral measure. To provide some background, a martingale is a stochastic process for which, at a particular time in the realized sequence, the expectation of the next value in the
sequence is equal to the present observed value even given knowledge of all prior observed
values at the current time. Based on this rationale, Proposition 8.1 of Steele (2001) introduces
the martingale P.D.E condition for a stochastic process, which suggests the drift term should
be zero for a martingale process. As the mathematic derivatives of the price function can
be calculated, the martingale condition above leads to partial differential equations (P.D.E)
which yields the analytical solution for the prices of travel time derivatives. In the following
analysis, the prices based on simple CARMA(1,0) processes with first order integration
are first discussed and then extended to general CARMA(p,q) processes with first order
integration and with \( p > q \); \( Y_t \) is obtained by removing trends and seasonal adjustments
from the original travel time series.

**Theorem 5.4.1**  For European call options based on the process \( Y_t \), where \( Y_t = \int_0^t X_u du \) and
\( X_t \) is a CARMA(1,0) process, the price can be found via the following P.D.E, after specifying
a risk neutral measure \( P \):

\[
v_t(t, x, y) + A(t, X_t) v_x(t, x, y) + x v_y(t, x, y) + \frac{1}{2} B(t, X_t)^2 v_{xx}(t, x, y) = r v(t, x, y)
\]

\( 0 \leq t < T, x \in \mathbb{R}, y \in \mathbb{R} \)

with the following boundary conditions:

\[
v(t, 0, y) = e^{-r(T-t)}(y - K)^+, 0 \leq t < T, y \in \mathbb{R}
\]

\[
\lim_{y \to -\infty} v(t, x, y) = 0, 0 \leq t < T, x \in \mathbb{R}
\]

\[
v(T, x, y) = (y - K)^+, x \in \mathbb{R}, y \in \mathbb{R}
\]

**proof:** Due to the existence of integration, i.e. \( Y_t = \int_0^t X_u du \), to price a European call
option on \( Y_t \) is similar to pricing an Asian call option on the primary process \( X_t \). The proof
follows from Theorem 7.5.1 of Shreve et al. (2004). As $Y_t$ is obtained by removing trends and seasonal adjustments from the original travel time series, it can be negative and the boundary conditions are slightly different.

**Theorem 5.4.2** For the Asian type call option on the process $Y_t$, where $Y_t = \int_0^t X_u du$ and $X_t$ is a CARMA(1,0) process, its price can be solved via the following P.D.E by further expanding states as follows:

$$v_t(t, x, y, z) + A(t, X_t)v_x(t, x, y, z) + xv_y(t, x, y, z) + yv_z(t, x, y, z) + \frac{1}{2} B(t, X_t)^2 v_{xx} = rv(t, x, y, z)$$

$0 \leq t < T, x \in \mathbb{R}, y \in \mathbb{R}, z \in \mathbb{R}$

with boundary conditions:

$$\lim_{z \to -\infty} v(t, x, y, z) = 0, \quad 0 \leq t < T, x \in \mathbb{R}, y \in \mathbb{R}$$

$$v(T, x, y, z) = (z - K)^+, x \in \mathbb{R}, y \in \mathbb{R}, z \in \mathbb{R}$$

$$v(t, 0, 0, z) = e^{-r(T-t)}(z - K)^+, t < T, z \in \mathbb{R}$$

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 process. The value of the derivative contract will be $P = de^{-rt}v(t, x, y, z)$. By Ito’s lemma, it
is subject to the following dynamics:

\[
\begin{align*}
dv(t, x, y, z) &= v_t dt + v_x dx + \frac{1}{2} v_{xx} d < x > + v_y dy + v_z dz \\
dP &= e^{-rt}(-rv + v_t + \frac{1}{2} v_{xx} B(t, X_t)^2 + v_x A^*(t, X_t) + xv_y + yv_z) dt + v_x B(t, X_t) e^{-rt} dW_t
\end{align*}
\]

The discounted price process should be a martingale under the risk neutral measure, 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.

For the boundary conditions, as \(Y_t\) is obtained by removing trends and seasonal adjustments from the original travel time series, it can be negative. All the state variables corresponding to the CARMA(1, 0) process can be in \(R\).

If \(Y(t)\) approaches \(-\infty\), then the probability that the call expires in the money approaches zero and the option price approaches zero. This leads to the first boundary condition. The second boundary condition is just the payoff of the call option.

Q.E.D.
Noticing the similarity in defining the state space representation of the CARMA\((p, q)\) and the Asian option pricing formula above, additional state expansion is employed to price the travel time derivatives based on CARMA\((p, q)\) processes. For example, consider the Asian option based on \(Y_t\), where \(Y_t = \int_0^t X_u du\) and \(X_t\) is a CARMA\((2, 0)\) process, the state space dynamic can be described by the following set of S.D.E:

\[
\begin{align*}
    dZ_t &= Y_t dt \\
    dY_t &= X_0^0 dt \\
    dX_0^0 &= X_1^1 dt \\
    dX_1^1 &= A^\ast(t, X_0^0, X_1^1) dt + B(t, X_0^0, X_1^1) dW_t
\end{align*}
\]

where \(X_0^0\) and \(X_1^1\) construct the CAR\((2)\) process, and \(Y_t\) represents the CARMA\((2, 0)\) process, and \(Z_t\) characterizes the integrated price process in the payoff function of the Asian option. This group of differential equations defines a four dimensional Markovian process. The value of derivative contract will be \(de^{-rt}v(t, x_1, x_0, y, z)\). By Ito’s lemma, it is subject to the following dynamics:

\[
\begin{align*}
    dv(t, x_1, x_0, y, z) &= v_t dt + v_{x_1} dx_1 + \frac{1}{2} v_{x_1 x_1} d < x_1 > + v_{x_0} dx_0 + v_y dy + v_z dz \\
    dP &= e^{-rt} (-rv + v_t + \frac{1}{2} v_{x_1 x_1} B(t, X_0^0, X_1^1)^2 + v_{x_1} A^\ast(t, X_0^0, X_1^1)) \\
        &\quad + x_1 v_{x_0} + x_0 v_y + y v_z) dt + v_{x_1} B(t, X_0^0, X_1^1) e^{-rt} dW_t
\end{align*}
\]

Using the martingale condition, the pricing P.D.E is obtained as follows:

\[
v_t + \frac{1}{2} v_{x_1 x_1} B(t, X_0^0, X_1^1)^2 + v_{x_1} A^\ast(t, X_0^0, X_1^1) + x_1 v_{x_0} + x_0 v_y + y v_z = rv
\]
For the Asian option based on \( Y_t \), where \( Y_t = \int_0^t X_u du \) and \( X_t \) is a CARMA(2,1) process, the following set of S.D.E holds in a similar fashion except that the moving average coefficients lead to different representation of \( Y_t \):

\[
\begin{align*}
  dZ_t &= Y_t dt \\
  dY_t &= (b_0 X_1^t + b_1 X_0^t) dt \\
  dX_0^t &= X_1^t dt \\
  dX_1^t &= A^*(t, X_1^t, X_1^t) dt + B(t, X_1^t, X_1^t) dW_t
\end{align*}
\]

This group of differential equations defines a four dimensional Markovian process. The value of derivative contract will be \( de^{-rt}v(t, x_1, x_0, y, z) \). By Ito’s lemma, it is subject to the following dynamics:

\[
\begin{align*}
  dv(t, x_1, x_0, y, z) &= v_t dt + v_{x_1} dx_1 + \frac{1}{2} v_{x_1x_1} d < x_1 > + v_{x_0} dx_0 + v_y dy + v_z dz \\
  dP &= e^{-rt}(-rv + v_t + \frac{1}{2} v_{x_1x_1} B(t, X_1^t, X_1^t)^2 + v_{x_1} A^*(t, X_1^t, X_1^t) \\
  &\quad + x_1 v_{x_0} + b_0 x_1 v_y + b_1 x_0 v_y + y v_z) dt + v_{x_1} B(t, X_1^t, X_1^t) e^{-rt} dW_t
\end{align*}
\]

Using the martingale condition, the pricing P.D.E is:

\[
v_t + \frac{1}{2} v_{x_1x_1} B(t, X_1^t, X_1^t)^2 + v_{x_1} A^*(t, X_1^t, X_1^t) + x_1 v_{x_0} + b_0 x_1 v_y + b_1 x_0 v_y + y v_z = rv
\]

The derivative prices for higher order CARMA\((p, q)\) processes with \( p > q \) can be calculated analytically using this methodology except that differences in the group of S.D.Es lead to corresponding changes in the P.D.E terms, which is summarized in the following theorem.
Theorem 5.4.3 If the asset price process is subject to $Y_t$, where $Y_t = \int_0^t X_u du$ and $X_t$ is a CARMA($p, q$) process with $p > q$, the Asian type call option based on it can be priced via the following P.D.E:

$$
v_t + \frac{1}{2}v_{x_{p-1}x_{p-1}}B(t, X^0_t, \ldots, X^{p-1}_t)^2 + v_{x_{p-1}}A^*(t, X^0_t, \ldots, X^{p-1}_t)$$

$$+ \sum_{i=0}^{p-2} x_{i+1}x_i + \sum_{i=0}^{q} b_ix_{q-i}v_y + yv_z = rv$$

where $v = v(t, x_0, \ldots, x_{p-1}, y, z)$ with boundary conditions

$$\lim_{z \to -\infty} v(t, x_0, \ldots, x_{p-1}, y, z) = 0, 0 \leq t < T, x_0 \in R, \ldots, x_{p-1} \in R, y \in R$$

$$v(T, x_0, \ldots, x_{p-1}, y, z) = (z - K)^+, x_0 \geq 0, \ldots, x_{p-1} \in R, y \in R, z \in R$$

$$v(t, 0, \ldots, 0, 0, z) = e^{-r(T-t)}(z - K)^+, 0 \leq t < T, z \in R$$

proof:

Consider the following set of S.D.E

$$dZ_t = Y_t dt$$

$$dY_t = \sum_{i=0}^{q} b_i X^{q-i}_t dt$$

$$dX^0_t = X^0_t dt$$

$$\ldots$$

$$dX^{p-2}_t = X^{p-1}_t dt$$

$$dX^{p-1}_t = A^*(t, X^0_t, \ldots, X^{p-1}_t)dt + B(t, X^0_t, \ldots, X^{p-1}_t)dW_t$$
This group of differential equation defines a $p + 2$ dimensional Markovian process. The value of derivative contract will be $de^{-rt}v(t, x_{p-1}, \ldots, x_0, y, z)$. By Ito’s lemma, it is subject to the following dynamics:

$$
dv(t, x_{p-1}, \ldots, x_0, y, z) = v_t dt + v_{x_{p-1}} dx_{p-1} + \frac{1}{2} v_{x_{p-1} x_{p-1}} d < x_{p-1} > + v_{x_{p-2}} dx_{p-2} + \ldots + v_{x_0} dx_0 + v_y dy + v_z dz
$$

$$
dP = e^{-rt} \left( -rv + v_t + \frac{1}{2} v_{x_{p-1} x_{p-1}} B(t, X^0_t, \ldots, X^{p-1}_t) + v_{x_{p-1}} A^*(t, X_t^0, \ldots, X^{p-1}_t) + x_{p-1} v_{x_{p-2}} + \ldots + x_1 v_{x_0}ight)
+ \sum_{i=0}^{q} b_i x_{q-i} v_y + y v_z) dt + v_{x_{p-1}} B(t, X^0_t, \ldots, X^{p-1}_t) e^{-rt} dW_t
$$

Using the martingale condition, the pricing P.D.E is:

$$
v_t + \frac{1}{2} v_{x_{p-1} x_{p-1}} B(t, X^0_t, \ldots, X^{p-1}_t) + v_{x_{p-1}} A^*(t, X_t^0, \ldots, X^{p-1}_t)
+ \sum_{i=0}^{p-2} x_{i+1} v_x + \sum_{i=0}^{q} b_i x_{q-i} v_y + y v_z = rv
$$

For the boundary conditions, as $Y_t$ is obtained by removing trends and seasonal adjustments from the original travel time series, it can be negative. All the state variables corresponding to the CARMA($p, q$) process can be in $\mathbb{R}$.

If $Y(t)$ approaches $-\infty$, then the probability that the call expires in the money approaches zero and the option price approaches zero. This leads to the first boundary condition. The second boundary condition is just the payoff for the call option at time $T$. The third boundary condition follows the discounted payoff of Asian call option from $T$ to $t$.

Q.E.D
In order to numerically solve the equation in the theorems above, it would normally be necessary to also specify the behavior of $v$ as all variables approaches $+\infty$ or $-\infty$, which can be different to each case. Moreover, the prices of put options and other derivatives based on the travel time process can be computed in a similar fashion.

To connect this risk neutral representation with typical hedging strategies, the derivation in the previous section is applied to $Z_t$: Suppose the two derivatives which are both derived on $Z_t$ but have two different payoff functions: $F$ and $G$. A portfolio is defined $P = \alpha F + \beta G$ of $F$ and $G$ with $\alpha + \beta = 1$. If the portfolio is risk neutral, then its value should increase at the same rate as a risk-free rate. The following P.D.E can be then derived, defining $\theta_t$ as the market value of risk.

\[
v_t + \frac{1}{2}v_{x_{p-1}x_{p-1}}B(t, X^0_t, \cdots, X^{p-1}_t)^2 + v_{x_{p-1}}(A(t, X^0_t, \cdots, X^{p-1}_t) - \theta_t B(t, X^0_t, \cdots, X^{p-1}_t)) + \sum_{i=0}^{p-2} x_{i+1}v_x + \sum_{i=0}^{q} b_i x_{q-i}v_y + yv_z = rv
\]

This P.D.E incorporates more explicitly the market value of risk and corresponding hedging strategy, while it maintains similar theoretical properties as the general P.D.E in the theorem above. As discussed in the previous section, different hedging strategies may introduce different risk neutral measures in an incomplete market.

In summary, this chapter introduces travel time derivatives as an innovative value pricing scheme, an effective hedging tool against risk due to bad quality of traffic service and a new financial instrument to diversify portfolio risk. The market participants are mainly travelers and business whose businesses may be influenced by the traffic system and typical financial derivatives such as futures and options can be derived based on travel time. Ornstein-Uhlenbeck process and more generally, the continuous time auto regression moving average (CARMA) models are used to model travel time while risk neutral pricing principle under incomplete market conditions is used to price such products; both explicit P.D.E solutions
and Monte Carlo methods are used to obtain the numerical asset prices. The analysis of financial derivative based on travel time extends the literature of derivative pricing based on non-tradable assets to new disciplines and leads to enormous research opportunities.
Chapter 6

Conclusions and Future Work

This thesis has examined the methodology of estimating road travel time distributions employing copula methods. It has also introduced advanced value pricing schemes employing financial derivatives based on travel time. The study highlights the dependent structure of travel times among different links, and develops a promising methodology for estimating non-Gaussian travel time distributions. Finally, significant contributions are offered by exploring the connections between the transportation and financial systems. The major contributions of this thesis are listed below:

- First, this thesis proposes a new approach for generating the density of travel times for a focused link set given nearby observations made using copula methods. Travel time along a link is modeled as a stationary random variable and the dependence between the variables corresponding to different links is modeled with a suitable copula. The conditional density of observed travel times on the focused link is then calculated based on this copula, using observed travel time values on related links 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 link travel time performance changes with overall traffic conditions along the path. Based on this logic, the present research models the dependence with different Gaussian copulas for all possible scenarios of path travel conditions, which are termed path scenarios. Additionally, in a given path scenario, a Gaussian copula is used to describe the dependent structure between the constructing links where the link travel time distributions are modeled as stationary random variables. A conditional path travel time distribution is then generated by summing 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 overall mathematical form of the model is a GCMM. This model is estimated using a two-step maximum likelihood method if the path scenario is observable and can therefore be pre-defined. Alternatively, the Expectation Maximum algorithm is used to estimate the GCMM model based on unobservable path scenarios. Overall, GCMMs first add more flexibility to fit heavy tails on marginal distributions while remaining relatively robust against it; GCMMs further incorporate unsynchronized data into estimation, both of which improve the approximation to the complex dependence structures given limited number of components and limited GPS data.

Third, this thesis introduces the financial derivative based on travel time. This new financial derivative is based on road travel times in 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 Ornstein-Uhlenbeck process and more generally, the continuous time auto regression moving average (CARMA) models are used to model travel time as the underlying asset. Derivative prices are calculated using the
risk neutral pricing principle under incomplete market conditions.

In the following section, this thesis proposes recommendations for future research, organized according to the three chapters of the thesis. These topic areas can be explored and analyzed in order to better describe the behavior of the transportation system and the financial system.

• First, travel time distribution estimation is difficult to conduct given the strong dependence between travel times. By using copulas and scenario analysis, this complexity can be addressed with quantified errors. Other variations of copula models can be used to improve goodness-of-fit in further research, Hsu et al. (2008) and Aas et al. (2009).

• Second, the estimation algorithm for GCMM models can be further improved and corresponding convergence properties should be studied. This is particularly necessary for the varying-margin-optimization algorithm, which assembles the two-step maximum estimator for a copula, Trivedi & Zimmer (2007), with the expectation-maximum algorithm, Xu & Jordan (1996). This whole procedure may yield other interesting theoretical properties.

• Third, other statistical models can be used to describe the complex dependent structure between travel times. A good alternative is kernel regression, Fan & Yim (2004) and Fan et al. (1996). The general properties of the kernel-based method have been studied extensively in the statistical literature, while specific properties should be studied for transportation applications.

• Fourth, more complex stochastic models can be used to describe the dynamics of the travel time process. For example, stochastic volatility models, Barndorff-Nielsen (2002) can be introduced to describe the complex variance structure of travel times over time, Levy-driven CARMA processes, Brockwell & Lindner (2009) can also be used to model the jumps in travel time, especially when dense measurements of individual travel time
can be obtained. Correspondingly, various methods can be used to price travel time derivatives, such as indifference pricing methods and other alternative hedging strategies. These methods may further improve the performance of both the transportation and the financial systems.

- Fifth, the introduction and modeling for financial derivatives based on travel time represent an innovative concept in operations research and financial engineering. Due-equilibrium of the financial system and 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. Therefore, maintaining equilibrium in both the transportation and the financial systems after introducing travel time derivatives is both a challenging and a promising avenue for further work, Liu & Nagurney (2007) and Duffie (2010).
Bibliography


factor models of high dimension’.


Estimation for Multivariate Gaussian or Binary Data’, The Journal of Machine Learning
Research 9, 485–516.

Banerjee, O. & Natsoulis, G. (2006), Convex optimization techniques for fitting sparse gauss-
ian graphical models, in ‘Proceedings of the 23rd international conference on Machine


(Statistical Methodology) 64(2), 253–280.

processes of finite variation’, PROBABILITY AND MATHEMATICAL STATISTICS-
WROCLAW UNIVERSITY 27(1), 3.


217


Sklar, M. (1959), *Fonctions de répartition à n dimensions et leurs marges*, Université Paris 8.


