DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

A NEWAPPROXIMATION TO DEPENDENCE FUNCTION IN BIVARIATE EXTREME VALUE DISTRIBUTIONS

by
Alireza AHMADABADI

June, 2017
IZMIR

A NEWAPPROXIMATION TO DEPENDENCE FUNCTION IN BIVARIATE EXTREME VALUE DISTRIBUTIONS

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics

by

Alireza AHMADABADI

June, 2017

IZMIR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"A NEW APPROXIMATION TO DEPENDENCE FUNCTION IN BIVARIATE EXTREME VALUE DISTRIBUTIONS"** completed by **ALIREZA AHMADABADI** under supervision of **ASSOC. PROF. DR. BURCU HÜDAVERDI ÜÇER** and **ASSOC. PROF. DR. GHOLAMHOSSEIN GHOLAMI** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

.....

Assoc. Prof. Dr. Burcu HÜDAVERDI ÜÇER

Supervisor

.....

Prof. Dr. İsmihan BAYRAMOĞLU

Thesis Committee Member

.....

Prof. Dr. Özlem EGE ORUÇ

Thesis Committee Member

.....

Prof. Dr. Ali İhsan GENÇ

Examining Committee Member

.....

Prof. Dr. G. Yaşar TÜTÜNCÜ

Examining Committee Member

.....

Prof. Dr. Emine İlknur CÖCEN

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my eternal advisor, Dr. Burcu HÜDAVERDI ÜÇER, for her patience, motivation, immense knowledge and for the continuous support of my Ph.D. study and related research. I would like to thank my second advisor Dr. Gholamhossein GHOLAMI. His guidance helped me in all the time of research and writing of this thesis.

Besides my advisors, I would like to thank my thesis committee: Prof. İsmihan BAYRAMOĞLU, and Dr. Özlem EGE ORUÇ, for their insightful comments and professional guidance which incented me to widen my research from various perspectives. My sincere thanks also to Dr. Uwe LIGGES in technische universität Dortmund and Dr. Axel BÜCHER in Ruhr universität Bochum for their constructive suggestions to improve thesis.

A very special gratitude goes out to all my good friends in Turkey, especially in Izmir, and also Dokuz Eylül University lovely staff for their unfailing support and assistance.

I would also like to thank my parents, brothers and sister, whose love and guidance are with me in whatever I pursue. Most importantly, I wish to thank my loving and supportive wife, Nasrin, and my dear daughter, Fatemeh, who provide unending inspiration.

Alireza AHMADABADI

A NEW APPROXIMATION TO DEPENDENCE FUNCTION IN BIVARIATE EXTREME VALUE DISTRIBUTIONS

ABSTRACT

Dependence structures of joint extreme events can be modeled by extreme value copulas characterized by Pickings dependence function. In this thesis, two new approaches using Kernel Regression and Bayesian methods are proposed to estimate Pickands dependence function.

In the first approach, new data points obtained with Bernstein copula approximation which have flexible form can serve to estimate the unknown Pickands dependence function. Kernel Regression method is then used to derive an intrinsic estimator. The performance of the estimator is given by a simulation study. Test results mainly show that the estimator has a better performance than the conventional estimators. In the second approach, Initially, cubic B-spline regression is used to model the dependence function. Then, the estimator of Pickands dependence function is obtained by the Bayesian approach. Through the estimation process, the prior and the posterior distributions of the parameter vectors are provided. The posterior sampling algorithm is presented in order to approximate the posterior distribution. We give a simulation study to measure and compare the performance of the proposed Bayesian estimator of the Pickands dependence function. A real data example is also illustrated.

Keywords: Bivariate extreme value copula, Pickands dependence function, Bernstein copula, Kernel regression, Convexity, Bayesian method, Reversible jump, MCMC

İKİ DEĞİŞKENLİ ÜÇ DEĞER DAĞILIMLARINDA BAĞIMLILIK FONKSİYONUNA YENİ BİR YAKLAŞIM

ÖZ

Bileşik uç değer olaylarının bağımlılık yapıları, Pickands bağımlılık fonksiyonu tarafından karakterize edilen uç-değer kopulaları ile modellenebilir. Bu tezde, çekirdek regresyonu ve Bayes yöntemlerini kullanarak Pickands bağımlılık fonksiyonunu tahmin etmek için iki yeni yaklaşım önerilmiştir.

İlk yaklaşımda, esnek bir forma sahip olan Bernstein kopulalar ile elde edilen yeni veri noktaları bilinmeyen Pickands bağımlılık fonskiyonunu tahmin etmek üzere kullanılmışlardır. Daha sonra, çekirdek regresyon yöntemi, gerçek bir kestirici çıkarmak için önerilmiştir. Bu kestircinin peformansı bir benzetim çalışması ile verilmiştir. Test sonuçları kestiricinin geleneksel kestircilere göre daha iyi bir performans sergilediğini göstermiştir. İkinci yaklaşımda, öncelikle kübik B-spline regresyon yöntemi bağımlılığı modellemek için kullanılmıştır. Daha sonra, Pickands bağımlılık fonksiyonunun kestircisi Bayes yaklaşımı ile elde edilmiştir. Tahminleme boyunca, parametre vektörlerinin öncül ve artçıl dağılımları elde edilmiştir. Artçıl örnekleme algoritması, artçıl dağılıma yaklaşım için verilmiştir. Pickands bağımlılık fonksiyonu için önerilen Bayes kestiricisinin performansını ölçmek ve karşılaştırmak için bir benzetim çalışması sunulmuştur. Aynı zamanda çalışmaya, gerçek veri örneği de eklenmiştir.

Anahtar kelimeler: İki değişkenli üç değer kopulası, Pickands bağımlılık fonksiyonu, Bernstein kopula, Çekirdek regresyonu, Konvekslik, Bayes yöntemi, Ters sıçrama, MCMC

CONTENTS

Page

Ph.D. THESIS EXAMINATION RESULT FORM ii
ACKNOWLEDGEMENTS iii
ABSTRACT iv
ÖZ v
LIST OF FIGURES x
LIST OF TABLES xiv

CHAPTER ONE – INTRODUCTION 1

CHAPTER TWO – COPULA 3

2.1 Copula Definition 3
2.2 Sklar's Theorem 3
2.3 Properties of Copula 4
2.3.1 Fréchet-Hoeffding Bounds Inequality 4
2.3.2 Dependence 5
2.3.3 Tail Dependence 6
2.3.4 Invariance 7
2.3.6 Symmetry 8
2.3.7 Continuity 8

CHAPTER THREE – EXTREME VALUE COPULA 9

3.1 Extreme Value Theory 9

3.2 Bivariate Extreme Value Distribution and its Copula 10

3.3 Pickands Dependence Function 12

3.3.1 Properties of Pickands Dependence Function..... 13

3.3.2 Nonparametric Estimators of Pickands Dependence Function..... 13

3.4 Some Parametric Models 14

3.4.1 Logistic or Gumbel-Hougaard Model..... 15

3.4.2 Asymmetric Logistic Model 16

3.4.3 Negative Logistic or Galambos Model 17

3.4.4 Asymmetric Negative Logistic Model..... 17

3.4.5 Hüsler-Reiss Model 17

3.4.6 **t−EV** Model..... 18

CHAPTER FOUR – THE BERNSTEIN COPULA 19

4.1 Bernstein Polynomials 19

4.1.1 Some Properties of Polynomials in Bernstein Form..... 21

4.2 Bernstein Copula..... 21

4.2.1 Some Properties of the Bernstein Copula 22

4.3 Nonparametric Bernstein Copula..... 22

CHAPTER FIVE – NEW ESTIMATION OF PICKANDS DEPENDENCE FUNCTION: KERNEL REGRESSION APPROACH 24

5.1 Introduction..... 24

5.2 New Pointwise Estimation of the Pickands Dependence Function 25

5.2.1 Pointwise Estimation of Pickands Dependence Function Based on the Usual Empirical Copula 25

5.2.2 New Pointwise Estimation of Pickands Dependence Function Based on the Empirical Bernstein Copula 25

5.2.3 Performance of the New Estimated Pairs 26

5.2.4 Simulation Study 26

5.3 New Estimation of the Pickands Function: Kernel Regression Approach 51

5.3.1 Performance of the New Estimator 57

5.3.2 Simulation Study 58

5.3.3 Performance Testing 59

CHAPTER SIX – NEW ESTIMATION OF PICKANDS DEPENDENCE FUNCTION: BAYESIAN APPROACH 77

6.1 Technical Introduction 77

6.1.1 The Bayesian Approach to Statistical Inference 77

6.1.2 Prior Distributions 78

6.1.3 Point Estimation 81

6.1.4 Bayesian Computation 81

6.2 The Statistical Model 87

6.2.1 Cubic B-spline Regression Model 87

6.2.2 The Bayesian Model 88

6.2.3 Posterior Sampling 91

6.3 Simulation Study 95

CHAPTER SEVEN – CONCLUSIONS 101

REFERENCES..... 104

LIST OF FIGURES

Page

Figure 5.1 Plots based on sample size of 200 from the Logistic Copula …….. 30

Figure 5.2 Plots based on sample size of 200 from the Negative Logistic copula. 31

Figure 5.3 Plots based on sample size of 200 from the Hüsler-Reiss copula …… 32

Figure 5.4 Plots based on sample size of 200 from the Asymmetric Logistic copula ($\alpha = 0.5, \beta = 0.5$) ……. 33

Figure 5.5 Plots based on sample size of 200 from the Asymmetric Negative Logistic copula ($\alpha = 0.5, \beta = 0.5$) …… 34

Figure 5.6 Plots based on sample size of 200 from the Asymmetric Logistic copula ($\alpha = 0.35, \beta = 0.75$) ……. 35

Figure 5.7 Plots based on sample size of 200 from the Asymmetric Logistic copula ($\alpha = 0.75, \beta = 0.35$) ……. 36

Figure 5.8 Plots based on sample size of 200 from the Asymmetric Negative Logistic copula ($\alpha = 0.35, \beta = 0.75$) ……... 37

Figure 5.9 Plots based on sample size of 200 from the Asymmetric Negative Logistic Copula ($\alpha = 0.75, \beta = 0.35$) ……... 38

Figure 5.10 MSE of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Logistic copula……..……. 42

Figure 5.11 MSE of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Negative Logistic copula ……..… 43

Figure 5.12 MSE of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Hüsler-Reiss copula ……..…… 44

Figure 5.13 MSE of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Asymmetric Logistic copula ($\alpha = 0.50, \beta = 0.50$) …… 45

Figure 5.14 MSE of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Asymmetric Negative Logistic copula ($\alpha = 0.50, \beta = 0.50$) …….. 46

Figure 5.15 MSE of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Asymmetric Logistic copula ($\alpha = 0.35, \beta = 0.75$) …… 47

Figure 5.16 MSE of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Asymmetric Logistic copula ($\alpha = 0.75, \beta = 0.35$) …… 48

Figure 5.17 MSE of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Asymmetric Negative Logistic copula ($\alpha = 0.35, \beta = 0.75$) …….. 49

Figure 5.18 MSE of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Asymmetric Negative Logistic copula ($\alpha = 0.75, \beta = 0.35$) …….. 50

Figure 5.19 Plots based on sample size of 300 from the Logistic Copula ……. 60

Figure 5.20 Plots based on sample size of 300 from the Negative Logistic copula. 61

Figure 5.21 Plots based on sample size of 300 from the Hüsler-Reiss copula …… 62

Figure 5.22 Plots based on sample size of 300 from the Asymmetric Logistic copula (0.35,0.75) ……...…… 63

Figure 5.23 Plots based on sample size of 300 from the Asymmetric Logistic copula (0.75,0.35) ……..…. 64

Figure 5.24 Plots based on sample size of 300 from the Asymmetric Negative Logistic copula (0.35,0.75) ……. 65

Figure 5.25 Plots based on sample size of 300 from the Asymmetric Negative Logistic copula (0.75,0.35) ……...……... 66

Figure 5.26 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Logistic model …….. 69

Figure 5.27 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Negative Logistic model ..… 70

Figure 5.28 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Hüsler-Reiss model ……... 71

Figure 5.29 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Asymmetric Logistic model (α = 0.35, β = 0.75) ……...…….. 72

Figure 5.30 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Asymmetric Logistic model (α = 0.75, β = 0.35) ……..... 73

Figure 5.31 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Asymmetric Negative Logistic model (α = 0.35, β = 0.75) …….. 74

Figure 5.32 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Asymmetric Negative Logistic model (α = 0.75, β = 0.35) …….. 75

Figure 6.1 Plots with true Pickands dependence function A (red) and three estimates based on 1000 random samples of size 300 from the Asymmetric Logistic (α = 0.35, β = 0.75, τ = 0.15) (top left), Galambos (τ = 0.5) (top right), Logistic (τ = 0.5) (bottom left), Extreme-value t distributions (τ = 0.5) (bottom right) ……... 96

Figure 6.2 Sequence of m's of $\hat{A}_{Bayesian}$ and the corresponding approximated probability densities based on 1000 random samples of size 300 from Asymmetric Logistic (α = 0.35, β = 0.75, τ = 0.15) (top left), Galambos (τ = 0.5) (top right), Logistic (τ = 0.5) (bottom left), Extreme-value t distributions (τ = 0.5) (bottom right) ……..…. 97

Figure 6.3 Estimated Variance of $\hat{A}_{Bayesian}$ with running mean (red line) and corresponding histogram based on 1000 random samples of size 300 from Asymmetric Logistic (α = 0.35, β = 0.75, τ = 0.15) (top left), Galambos (τ = 0.5) (top right), Logistic (τ = 0.5) (bottom left), Extreme-value t distributions (τ = 0.5) (bottom right) ……..….. 98

Figure 6.4 Sequence of MSE's of $\hat{A}_{Bayesian}$ against iterations based on 1000 random samples of size 300 from Asymmetric Logistic ($\alpha = 0.35$, $\beta = 0.75$, $\tau = 0.15$) (top left), Galambos ($\tau = 0.5$) (top- right), Logistic ($\tau = 0.5$) (bottom left), Extreme-value t distributions ($\tau = 0.5$) (bottom right)… 99

Figure 6.5 Real data analysis: Plots with three estimates of Pickands dependence function A (red); $\hat{A}_{CFG}$ (green), $\hat{A}_{PIC}$ (pink), $\hat{A}_{Bayesian}$ (red dashed line) and gray area is a 95% credible interval for the estimated curve (top left)……. 100

LIST OF TABLES

Page

Table 5.1 Median of Mean Squared Error of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Logistic copula …….. 39

Table 5.2 Median of Mean Squared Error of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Asymmetric Logistic copula (0.5,0.5) …….. 39

Table 5.3 Median of Mean Squared Error of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Negative Logistic copula…… 40

Table 5.4 Median of Mean Squared Error of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Asymmetric Negative Logistic copula (0.5,0.5) ….. 40

Table 5.5 Median of Mean Squared Error of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Hüsler-Reiss copula …….. 41

Table 5.6 Median of Mean Squared Error of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Asymmetric Logistic copula (0.35,0.75) and (0.75,0.35) …….. 41

Table 5.7 Median of Mean Squared Error of our new pairs, (T_i,Q_i), and the estimated pair by usual empirical copula, (T_i,Z_i), based on 1000 sample size of 250 from Asymmetric Negative Logistic copula (0.35,0.75) and (0.75,0.35) …….. 51

Table 5.8 The probability of success $M^* < M$ in thousands of samples for Logistic, Asymmetric Logistic (0.5,0.5), Negative Logistic, Asymmetric Negative Logistic (0.5,0.5) and Hüsler-Reiss distributions ……. 52

Table 5.9 The statistically significant comparison of MSE's in level 0.001 based on 1000 random sample size of 250 from Logistic, Asymmetric Logistic

(0.5,0.5), Negative Logistic, Asymmetric Negative Logistic (0.5,0.5) and Hüsler-Reiss distributions …… 53

Table 5.10 The probability of success $M^* < M$ in thousands of samples for Asymmetric Logistic (0.35,0.75), Asymmetric Logistic (0.75,0.35), Asymmetric Negative Logistic (0.35,0.75) and Asymmetric Negative Logistic (0.75,0.35) distributions …… 54

Table 5.11 The statistically significant comparison of MSE's in level 0.001 based on 1000 random sample size of 250 from Asymmetric Logistic (0.35,0.75), Asymmetric Logistic (0.75,0.35), Asymmetric Negative Logistic (0.35,0.75) and Asymmetric Negative Logistic (0.75,0.35) distributions …… 54

Table 5.12 Median of Mean Squared Error of $\hat{A}$ from Logistic model for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level …… 59

Table 5.13 Median of Mean Squared Error of $\hat{A}$ from Negative Logistic model for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level …… 59

Table 5.14 Median of Mean Squared Error of $\hat{A}$ from Hüsler-Reiss model for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level …… 67

Table 5.15 Median of Mean Squared Error of $\hat{A}$ from Asymmetric Logistic model with parameters (0.35,0.75) for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level …… 67

Table 5.16 Median of Mean Squared Error of $\hat{A}$ from Asymmetric Logistic model with parameters (0.75,0.35) for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level …… 67

Table 5.17 Median of Mean Squared Error of $\hat{A}$ from Asymmetric Negative Logistic model with parameters (0.35,0.75) for convex Bernstein copula

estimator, Pickands estimator and CFG estimator of different values of dependence level ……. 68

Table 5.18 Median of Mean Squared Error of $\hat{A}$ from Asymmetric Negative Logistic model with parameters (0.75,0.35) for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level ……. 68

Table 5.19 Statistically significant comparisons in Median of Mean Squared Error of $\hat{A}$ from Logistic model for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level ……. 76

Table 5.20 Statistically significant comparisons in Median of Mean Squared Error of $\hat{A}$ from Asymmetric Logistic model (0.35,0.75) for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level …….. 76

Table 5.21 Statistically significant comparisons in Median of Mean Squared Error of $\hat{A}$ from Asymmetric Logistic Model (0.75,0.35) for Convex Bernstein Copula Estimator, Pickands Estimator and CFG estimator of different values of dependence level …….. 76

CHAPTER ONE
INTRODUCTION

There has been a growing interest for multivariate modeling of extreme values especially in the field of environmental and financial studies. In this context, inference methods for modeling the extreme-value dependence structure are getting improved. Pickands dependence function proposed by Pickands (1981) is used to model the extreme-value dependence structure through extreme-value copulas.

Estimation of the Pickands dependence function is the main step in modeling the multivariate extremes. Pickands (1981) and Deheuvels (1991) had early works in estimating the extreme-value dependence function. Capéraà et al. (1997), Hall and Tajvidi (2000), Segers (2007) suggested some alternative estimators for Pickands dependence function and they investigated some of its properties. Genest and Segers (2009) proposed a rank-based estimator of Pickands (1981) and Caperaà et al. (1997). Bücher et al. (2011) considered minimum-distance principle for the estimator and provided an infinite class estimators. Zhang et al. (2008) and Gudendorf and Segers (2011) studied multivariate case assuming knowledge of marginal distributions and also they discussed the asymptotic properties of the estimator. Berghaus et al (2012) constructed an alternative class of estimator for Pickands dependence function for generalized case. They used the minimum distance approach. Marcon et al. (2014) proposed a nonparametric estimator by using Bernstein polynomial approximation. They discussed the properties of the proposed estimation method. Cormier et al. (2014) proposed a visual tool for detecting the presence of extreme-value dependence and they used B-splines method for their estimator.

The main aim in this study is to estimate the Pickands dependence function of some bivariate extreme value copulas using Bernstein copula approximation with Kernel Regression method and Bayesian method. These are new approaches for estimating Pickands dependence function verifying all the constraints defined on the family of the Pickands dependence functions.

This thesis is organized as follows. The copula function and its important properties is introduced in Chapter two. Chapter three is concerned with extreme value theory. This

chapter describes bivariate extreme value copula and its related Pickands dependence function which is the main object of our study. We discuss the properties and some nonparametric estimators of the Pickands dependence function. Some popular extreme value models are introduced at the end of chapter three. In Chapter four, we discuss the Bernstein copula and its properties. We define empirical Bernstein copula method which utilizes to derive our new estimator of the Pickands dependence function via Kernel Regression method.

In Chapter five, a new approch called the pointwise estimation is introduced to obtain the pairs which can be used to estimate the dependence function. The properties of new pairs are investigated with a simulation study. The test results of the new estimated pairs are given. The new flexible non-parametric estimation method is applied to some well-known bivariate extreme value models and the dependence structure with changing parameters is observed. Then by utilizing Kernel regression on new estimated pairs, we derive an intrinsic estimator and convexify it by using convex hull procedure. Aguilera et al. (2011) proved that after applying the convexification, the uniform rate of convergence of estimator is preserved. The new flexible nonparametric estimation method is applied to some well-known bivariate extreme value models as Logistic (Gumbel-Hougaard) , Asymmetric Logistic, Negative Logistic (Galambos), Asymmetric Negative Logistic and Hüsler-Reiss distributions and the dependence structure with changing parameters is observed through a simulation study.

In Chapter six, a Bayesian approach which is based on specifying a probability model for the data is suggested to estimate Pickands dependence function. We define cubic B-spline regression function for the Pickands dependence function and the Bayesian model which consist of prior and the posterior distributions for the unknown vector of parameters. We present posterior sampling algorithm using reversible jump MCMC method. A simulation study is carried out in order to compare the estimators in terms of accuracy and efficiency with some well-known extreme-value copula as Logistic, Asymmetric Logistic, Negative Logistic and Extreme Value t-student with four degrees of freedom and real-data example is also illustrated. Finally, Chapter seven presents conclusions.

CHAPTER TWO
COPULA

2.1 Copula Definition

Let (X,Y) be a pair of continuous random variables with cumulative distribution functions $F(x) = Pr(X \leq x)$ and $G(y) = P(Y \leq y)$, respectively, and let $H(x,y) = Pr(X \leq x, Y \leq y)$ be a joint cumulative distribution function of (X,Y). Thus for each pair of numbers (x,y) in $[-\infty,\infty]^2$, there are three numbers $F(x)$, $G(y)$ and $H(x,y)$ so that each of these values lies in the interval $[0,1]$. To further understand, suppose that each point (x,y) can be written as $(F(x),G(y))$ in the unit square $[0,1]\times[0,1]$ by using the values of marginal distribution functions $F(x)$ and $G(y)$. Consider a function C that establishes a correspondence between values of joint cumulative distribution function $H(x,y)$ and ordered pair $(F(x),G(y))$. This function $C : [0,1]^2 \to [0,1]$ is called *Copula*. In other words, copulas are functions that join or couple joint distribution function to its univariate marginal distribution functions.

Definition 2.1. A two dimensional copula $C : [0,1]^2 \to [0,1]$ is a bivariate cumulative distribution function with following properties:

1. $C(u,0) = C(0,v) = 0$
2. $C(u,1) = u$ and $C(1,v) = v$
3. C is 2-increasing. For every hyper rectangle $B = [u_1,u_2]\times[v_1,v_2]$, the $C-Volume$ of B is non-negative, i.e.

$$V_C(B) = C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$$

2.2 Sklar's Theorem

To achieve a better understanding of the importance and the role of the copula in dependence modelling, addressing the following theorem dates back to Sklar (1959) is necessary.

Theorem 2.1. Let H be a joint distribution function with margins F and G. Then, there exists a copula $C : [0,1]^2 \to [0,1]$ such that for all x and y in $R = [-\infty,\infty]$

$$H(x,y) = C\left(F(x),G(y)\right). \tag{2.1}$$

If margins are continuous, then C is unique; otherwise C is uniquely determined on $RanF \times RanG$, the Cartesian product of the ranges of the F and G, conversely, if C is a copula and F and G are univariate distribution functions, then the function H defined in (2.1) is a joint distribution function with margins F and G. Indeed, Sklar's theorem states that every multivariate cumulative distribution function can be expressed in terms of its margins and copula C. Sklar's theorem was first provided by Abe Sklar in 1959. Sklar introduced the name and the notion of copula and also obtained very important theoretical foundation for the application of copula.

The importance of the Sklar's theorem in the concept of dependency is not negligible. It is clear that the joint distribution function $H(x,y)$ contains all information about the dependence of two variables X and Y. The joint distribution function can be divided into margins and into a copula. Because the margins F and G are not able to explain the dependency between X and Y, thus it is completely done by means of copula C.

Corollary 2.1.1. Let X and Y be continuous random variables with copula C, joint distribution function $H(x,y)$ and margins $F(x)$ and $G(y)$, and let F^{-1} and G^{-1} be quasi-inverses of F and G respectively. Then for any $(u,v) \in [0,1]^2$

$$C(u,v) = H(F^{-1}(u), G^{-1}(v)).$$

2.3 Properties of Copula

2.3.1 Fréchet-Hoeffding Bounds Inequality

Let $C(u,v)$ be a two dimensional copula then for every (u,v) in $[0,1]^2$,

$$W(u,v) = max(u+v-1,0) \leq C(u,v) \leq min(u,v) = M(u,v)$$

where $W(u,v)$ and $M(u,v)$ are the Fréchet-Hoeffding lower and upper bounds respectively.

The Fréchet-Hoeffding upper bound $M(u,v)$ corresponds to perfect positive dependence and lower bound $W(u,v)$ corresponds to perfect negative dependence. Random variables which their copula is equal to $M(u,v)$ are called *Comonotonic*, and also random variables which their copula is equal to $W(u,v)$ are called *Counter Monotonic*. Note that there is a third case which represents independence between two

variables. If margins F and G are continuous, then X and Y are independent if and only if $C(u,v) = uv = \Pi(u,v)$. Π is called the independence copula or product copula.

2.3.2 Dependence

Copulas play an important role in the study of dependence. Two most widely known measures of association are Kendall's tau (τ) and Spearman's rho (ρ) that are defined in terms of copulas. Before any discussing about these dependency measures, it is better to define the term *Concordance*.

Definition 2.2. Let (X,Y) be a pair of continuous random variables with joint distribution function $H(x,y)$, and let (x_1,y_2) and (x_2,y_2) be two observations from (X,Y). These observations are *Concordant* if $(x_1-x_2)(y_1-y_2) > 0$ and *Discordant* if $(x_1-x_2)(y_1-y_2) < 0$.

Theorem 2.2. Let (X_1,Y_1) and (X_2,Y_2) be two pairs of independent continuous random variables with a common joint distribution function $H(x,y)$ and copula $C(u,v)$, and let τ be equal to the probability of concordance minus the probability of discordance. i.e.,

$$\tau = Pr\{(X_1-X_2)(Y_1-Y_2) > 0\} - Pr\{(X_1-X_2)(Y_1-Y_2) < 0\},$$

then the Kendall's tau measure is given by

$$\tau = 4 \iint_{(\mathbf{u},\mathbf{v})\in\mathbf{I}^2} C(u,v)dC(u,v) - 1 = 4\mathbb{E}[C(u,v)] - 1. \tag{2.2}$$

Theorem 2.3. Let (X_1,Y_1), (X_2,Y_2) and (X_3,Y_3) be three pairs of independent continuous random variables with a common joint distribution function $H(x,y)$ and copula $C(u,v)$, and let ρ be equal to the probability of concordance minus the probability of discordance for the pairs (X_1,Y_1) and(X_2,Y_3). i.e.

$$\rho = 3\, Pr\{(X_1-X_2)(Y_1-Y_3) > 0\} - Pr\{(X_1-X_2)(Y_1-Y_3) < 0\}$$

then the Spearman's rho measure is given by

$$\rho_{XY} = 12 \iint_{(\mathbf{u},\mathbf{v})\in\mathbf{I}^2} uvdC(u,v) - 3 = 12 \iint_{(\mathbf{u},\mathbf{v})\in\mathbf{I}^2} C(u,v)dudv - 3. \tag{2.3}$$

See Nelsen (2006).

Theorem 2.4. Let τ and ρ be Kendall's tau and Spearman's rho of continuous random variables respectively, then

$$-1 \leq 3\tau - 2\rho \leq +1$$

See Nelsen (2006).

2.3.3 Tail Dependence

Tail dependence measures the dependence between two random variables in the tails (upper-right and lower-left quadrant of $\mathbf{I^2}$) of the distributions.

Definition 2.3. Let C be a copula of random variables X and Y with marginals F and G respectively. The *lower tail* (λ_L) and the *upper tail* (λ_U) dependence are defined as

$$\begin{aligned} \lambda_L &= \lim_{u \downarrow 0} Pr\{F(X) < u, \mid G(Y) < u\} && (2.4) \\ &= \lim_{u \downarrow 0} \frac{C(u,u)}{u} && (2.5) \\ \lambda_U &= \lim_{u \uparrow 1} Pr\{F(X) > u, \mid G(Y) > u\} && (2.6) \\ &= \lim_{u \uparrow 1} \frac{1-2u+C(u,u)}{1-u} && (2.7) \end{aligned}$$

if there exists limits in (2.5) and (2.7), that is, $\lambda_L \neq 0$ and $\lambda_U \neq 0$.

If $\lambda_L = 0$ ($\lambda_U = 0$), then the copula C has no lower (upper) tail dependence. And if λ_L (λ_U) is in $(0,1]$, then the copula C has lower (upper) tail dependence.

Example 2.1. Consider following bivariate copula

$$C(u,v) = \exp\left[-\left[(-\log u)^{\frac{1}{\alpha}} + (-\log v)^{\frac{1}{\alpha}}\right]^{\alpha}\right]$$

for (u,v) in $\mathbf{I^2}$ and $0 < \alpha \leq 1$. The tail dependence coefficients λ_L and λ_U are evaluated as follows:

$$\begin{aligned} \lambda_L &= \lim_{u \downarrow 0} \frac{C(u,u)}{u} \\ &= \lim_{u \downarrow 0} \frac{u^{2^\alpha}}{u} \\ &= \lim_{u \downarrow 0} u^{(2^\alpha - 1)} \\ &= 0 \end{aligned}$$

then this copula has no lower tail dependence, and

$$\begin{aligned}\lambda_U &= \lim_{u\uparrow 1}\frac{1-2u+C(u,u)}{1-u}\\ &= \lim_{u\uparrow 1} 2-2^{\alpha}u^{(2^{\alpha}-1)}\\ &= 2-2^{\alpha}\end{aligned}$$

as can be seen for $0<\alpha\leq 1$ the copula has the upper tail dependence $2-2^{\alpha}$.

2.3.4 Invariance

Let (X,Y) be a pair of continuous random variables with joint distribution function $H(x,y)$, margins $F(x)$ and $G(y)$ and copula C_{XY}. Let α and β be strictly increasing on the ranges of X and Y respectively, and let αX and βY have distributions F^* and G^* respectively. Because α and β are strictly increasing on their ranges,

$$F^*(x) = Pr\{\alpha X \leq x\} = Pr\{X \leq \alpha^{-1}x\} = F(\alpha^{-1}x)$$

and

$$G^*(y) = Pr\{\beta Y \leq y\} = Pr\{Y \leq \beta^{-1}y\} = G(\beta^{-1}y).$$

Thus, for any $(x,y)\in\mathbb{R}^2$

$$\begin{aligned}C_{\alpha X\ \beta Y}(F^*(x),G^*(y)) &= Pr\{\alpha X\leq x,\beta Y\leq y\}\\ &= Pr\{X\leq \alpha^{-1}x, Y\leq \beta^{-1}y\}\\ &= C_{XY}(F(\alpha^{-1}x),G(\beta^{-1}y))\\ &= C_{XY}(F^*(x),G^*(y)).\end{aligned}$$

Now, by continuity of X and Y, $C_{\alpha X\ \beta Y}=C_{XY}$ on $\mathbf{I}^2$, C_{XY} is *invariant* under strictly increasing transformations of X and Y. Similarly, one can obtain results for the other forms of α and β. If α is strictly increasing and β is strictly decreasing, then

$$C_{\alpha X\ \beta Y}(u,v) = u - C_{XY}(u,1-v),$$

if α is strictly decreasing and β is strictly increasing, then

$$C_{\alpha X\ \beta Y}(u,v) = v - C_{XY}(1-u,v),$$

and if α and β are strictly decreasing, then

$$C_{\alpha X\ \beta Y}(u,v) = u+v-1+C_{XY}(1-u,1-v).$$

2.3.5 Symmetry

Let X and Y be continuous random variables with copula C, joint distribution function $H(x,y)$ and marginals $F(x)$ and $G(y)$. X and Y are exchangeable if and only if for all (u,v) in $[0,1]^2$, $F = G$ and $C(u,v) = C(v,u)$. When $C(u,v) = C(v,u)$, C is symmetric.

2.3.6 Continuity

Theorem 2.5. Let C be a copula of random variables X and Y. Then for every u_1, u_2, v_1 and v_2 in domain C,

$$|C(u_2,v_2) - C(u_1,v_1)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

Hence, C is uniformly continuous on its domain.

CHAPTER THREE
EXTREME VALUE COPULA

3.1 Extreme Value Theory

Extreme velue theory is one of the most important fields of the statistics that is widely used in the other sciences such as finance, insurance, environmental and engineering. Suppose we want to estimate the events outside the range of the data, for example, annual maximum sea levels, then extreme value theory can be used as a strong theoretical framework for modelling such kind of data and phenomenia.

Extreme value distribution is the limiting distribution for the minimum or the maximum of a large collection of random observations. There are two main approaches to identifying the extremes: *Block Maxima* and *Peak Over Threshold*. The block maxima method is based on modelling the maximum of a sample and the peak of threshold is the method which considers the excess values of a sample over a threshold. In this research, we consider the block maxima method.

Let $\{X_i\}_{i=1}^n$ be a sequence of identically independent random variables with cumulative distribution function F and let $Z_n = max\{X_i\}_{i=1}^n$. It is clear that the exact distribution of Z_n can be derived as follows,

$$\begin{aligned} Pr(Z_n \leq z) &= Pr\{X_1 \leq z, X_2 \leq z, \cdots, X_n \leq z\} \\ &= Pr\{X_1 \leq z\} \times \cdots \times Pr\{X_n \leq z\} \\ &= [F(z)]^n. \end{aligned}$$

In practice, the distributon function F is often unknown. One can approximate F via observed data, but for any z less than upper end point of F, $F^n(z) \to 0$ as $n \to \infty$. This problem can be solved by choosing the appropriate sequence of constants a_n and $b_n > 0$, such that $\lim_{n\to\infty} F^n(a_n + b_n z) = G(z)$, where $G(z)$ is a non-degenerate distribution function. The *Fisher-Tippet- Gnedenko* theorem provides the following asymptotic result.

Theorem 3.1 (Fisher-Tippet-Gnedenko)**.** Let $\{X_i\}_{i=1}^n$ be a sequence of identically independent random variables and let $Z_n = max\{X_i\}_{i=1}^n$. If there exist norming constants

$b_n > 0$ and a_n such that

$$Pr\left(\frac{Z_n - a_n}{b_n} \leq z\right) \longrightarrow G(z)$$

as $n \to \infty$, where G is a non-degenerate distribution function, then G belongs to type of one of the following standard extreme value distributions: *Gumbel*, *Fréchet* and *Weibull*, respectively, as

$$\begin{aligned}
G_1(z) &= \exp\left\{-\exp\left[-\left(\frac{z-a}{b}\right)\right]\right\} \quad -\infty < z < \infty \\
G_2(z) &= \begin{cases} 0 & z \leq a \\ \exp\left[-\left(\frac{z-a}{b}\right)^{-\alpha}\right] & z > a \end{cases} \\
G_3(z) &= \begin{cases} \exp\left\{-\left[-\left(\frac{z-a}{b}\right)\right]^{\alpha}\right\} & z < a \\ 1 & z \geq a \end{cases}
\end{aligned}$$

where in both Frechet and Weibull distributions, $\alpha > 0$.

These three distributions given in Theorem 3.1 can be combined into a distribution called *Generalized Extreme Value Distribution*. The distribution function of generalized extreme value is as follows:

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\gamma}{\delta}\right)\right]^{-\frac{1}{\xi}}\right\}, \tag{3.1}$$

where $1 + \xi\left(\frac{z-\gamma}{\delta}\right) > 0$, $\delta > 0$ and $-\infty < \gamma, \xi < \infty$.

The parameter ξ is the shape parameter which determines the rate of the tail decay. One can obtain the *heavy-tailed* Fréchet distribution for $\xi > 0$, *light-tailed* Gumbel distribution for $\xi = 0$ and *short-tailed* Weibull distribution for $\xi < 0$. For more details, see Coles (2001).

3.2 Bivariate Extreme Value Distribution and its Copula

Let $\{X_i, Y_i\}_{i=1}^n$ be identically independent pairs of random variables with joint distribution function $H(x,y)$, copula $C(u,v)$ and margins F and G. Let $Z_{1,n} = \max\{X_i\}_{i=1}^n$ and $Z_{2,n} = \max\{Y_i\}_{i=1}^n$. In theory, the distribution functions of $Z_{1,n}$ and $Z_{2,n}$ are $F^n(x)$ and

$G^n(y)$, respectively.

$$\begin{aligned} Pr\{Z_{1,n} \le z_1, Z_{2,n} \le z_2\} &= Pr\left\{ \frac{Z_{1,n} - a_{1,n}}{b_{1,n}} \le x,\ \frac{Z_{2,n} - a_{2,n}}{b_{2,n}} \le y \right\} \\ &= Pr\{Z_{1,n} \le a_{1,n} + b_{1,n}x,\ Z_{2,n} \le a_{2,n} + b_{2,n}y\} \\ &= F^n(a_{1,n} + b_{1,n}x,\ a_{2,n} + b_{2,n}y) \end{aligned}$$

If there are suitable $a_{i,n}$ and $b_{i,n} > 0$ for $i = 1, 2$, then

$$\lim_{n\to\infty} F^n(a_{1,n} + b_{1,n}x,\ a_{2,n} + b_{2,n}y) = H(x,y)$$

where the limit is in the sense of the weak convergence and $H(x,y)$ is a non-degenerate bivariate distribution function. In accordance to the Sklar's theorem

$$H(x,y) = C_H(F(x), G(y)) \tag{3.2}$$

where C_H is a bivariate extreme value copula. In the other words, assuming that F^n and G^n are the distribution functions of $Z_{1,n}$ and $Z_{2,n}$, respectively, Sklar's theorem states that,

$$\begin{aligned} Pr\{Z_{1,n} \le x, Z_{2,n} \le y\} &= H_n(x,y) \\ &= C_n\left(F^n(x), G^n(y)\right) \\ &= (H(x,y))^n \\ &= (C(F(x), G(y)))^n \end{aligned}$$

for enough large n, by transformation $F^n(x) = u$, $G^n(y) = v$, that is, $F(x) = u^{\frac{1}{n}}$, $G(y) = v^{\frac{1}{n}}$ and by considering above relationship,

$$C_n(u,v) = \left(C\left(u^{\frac{1}{n}}, v^{\frac{1}{n}}\right)\right)^n. \tag{3.3}$$

Note that the marginals of $H_n(x,y)$ are $F^n(x)$ and $G^n(y)$ and the copula of $H_n(x,y)$ is $C_n(u,v)$. $C(u,v)$ is copula of $H(x,y)$ whose marginals are $F(x)$ and $G(y)$.

Definition 3.1. In the bivariate case, a copula C is a extreme value copula if there exists a copula C_* such that

$$C_*^n(u^{\frac{1}{n}},\ v^{\frac{1}{n}}) \longrightarrow C(u,\ v) \quad u, v \in \mathbf{I^2}$$

as n tends to infinity.

Remark that the copula C_* is in the *domain of attraction* of the extreme value copula C.

3.3 Pickands Dependence Function

Before addressing the issue of the *Pickands dependence function*, it is necessary to mention the following two definitions.

Definition 3.2. A copula C is *Max-stable* if for every positive real number r and (u,v) in $\mathbf{I}$,

$$C(u,v) = C^r\left(u^{\frac{1}{r}}, v^{\frac{1}{r}}\right).$$

Definition 3.3. Let Z be a continuous random variable with distribution function ζ. The *Survival function* is the probability that variable Z takes on a value greater than $z > 0$, that is,

$$S(z) = Pr(Z > z) = 1 - \zeta(z).$$

Dependence Function[**Pickands 1981**]: Suppose that X and Y are standard exponential random variables which their survival functions are $u = e^{-x}$ and $v = e^{-y}$, respectively. The joint survival function is given by

$$\begin{aligned}
\overline{H}(x,y) &= Pr(X > x,\, Y > y) \\
&= 1 - F(x) - G(y) + H(x,y) \\
&= 1 - (1 - F(x)) - (1 - G(y)) + C(1 - (1 - F(x)), 1 - (1 - G(y))) \\
&= \hat{C}\left(\overline{F}(x), \overline{G}(y)\right).
\end{aligned}$$

Letting $\overline{F}(x) = t$ and $\overline{G}(y) = s$, then

$$\hat{C}(t,s) = 1 - t - s + C(1-t, 1-s).$$

Now let $\overline{H}(x,y) = \hat{C}(e^{-x}, e^{-y})$ and for $t \in [0,1]$ and $s > 0$, let

$$\left\{ \begin{array}{l} x = s(1-t) \\ y = st \end{array} \right. \Longrightarrow \left\{ \begin{array}{l} s = x + y \\ t = \frac{y}{x+y} \end{array} \right. \tag{3.4}$$

then,

$$\begin{aligned}
\hat{C}(e^{-x},e^{-y}) &= \hat{C}\left(e^{-s(1-t)},e^{-st}\right)\\
&= \left(\hat{C}\left(e^{-(1-t)},e^{-t}\right)\right)^{s}\\
&= \exp\log\left(\hat{C}\left(e^{-(1-t)},e^{-t}\right)\right)^{s}\\
&= \exp s\log\hat{C}\left(e^{-(1-t)},e^{-t}\right)\\
&= \exp\left(-sA(t)\right)\\
&= \exp\left(-(x+y)A\left(\frac{y}{x+y}\right)\right)
\end{aligned}$$

where

$$A(t) = -\log\hat{C}\left(e^{-(1-t)},\ e^{-t}\right). \tag{3.5}$$

So by assuming $u = e^{-x}$ and $v = e^{-y}$,

$$\hat{C}(u,v) = \exp\left[\log(uv)A\left(\frac{\log v}{\log(uv)}\right)\right]. \tag{3.6}$$

In the equation (3.6), the function A is called Pickands dependence function.

3.3.1 Properties of Pickands Dependence Function

1. A is a function with domain $[0,1]$ and range $[\frac{1}{2},1]$.

2. $A(0) = A(1) = 1$.

3. A is convex.

4. For all $t \in [0,1]$, $M = \max(t,1-t) \leq A(t) \leq 1 = \Pi$.

Note that if $A(t) = M$, the copula is the Frechet-Hoeffding upper bound, and if $A(t) = \Pi$, the copula is the independence copula.

Theorem 3.2. A bivariate copula C is an extreme value copula if and only if

$$C(u,v) = (uv)^{A\left(\frac{\log v}{\log(uv)}\right)}, \quad (u,v) \in (0,1]^2 \setminus \{(1,1)\} \tag{3.7}$$

where $A : [0,1] \to [\frac{1}{2},1]$ is convex and satisfies $max(t,(1-t)) \leq A(t) \leq 1$ for all $t \in [0,1]$.

3.3.2 Nonparametric Estimators of Pickands Dependence Function

Let $\{X_i,Y_i\}_{i=1}^{n}$ be identically independent pairs of random variables with joint distribution function $H(x,y)$, copula $C(u,v)$ and margins F and G. Put $U_i = F(X_i)$ and

$V_i = G(Y_i)$ and also put $S_i = -\log U_i$ and $T_i = -\log V_i$. We know that the variables S_i and T_i have standard exponential distribution. For every t in $[0,1]$, define

$$\xi_i(t) = min\left(\frac{S_i}{1-t}, \frac{T_i}{t}\right).$$

The original Pickands estimator dating back to Pickands (1981) is as follows

$$\frac{1}{\hat{A}_P(t)} = \frac{1}{n}\sum_{i=1}^{n}\xi_i(t). \tag{3.8}$$

This estimator does not satisfy the conditions of the Pickands dependence function $A(t)$.

Deheuvels (1991) proposed an estimator that only satisfies the endpoint condition, that is, $A(0) = A(1) = 1$. The Deheuvels estimator is given by

$$\frac{1}{\hat{A}_D(t)} = \frac{1}{n}\sum_{i=1}^{n}\xi_i(t) - t\frac{1}{n}\sum_{i=1}^{n}\xi_i(1) - (1-t)\sum_{i=1}^{n}\xi_i(0) + 1. \tag{3.9}$$

Hall and Tajvidi (2000) offered a method to improve the properties of the original Pickands estimator. For every t in $[0,1]$, define

$$\overline{\xi}_i(t) = min\left(\frac{\overline{S}_i}{1-t}, \frac{\overline{T}_i}{t}\right)$$

where $\overline{S}_i = \frac{S_i}{\frac{1}{n}\sum_{i=1}^{n} S_i}$ and $\overline{T}_i = \frac{T_i}{\frac{1}{n}\sum_{i=1}^{n} T_i}$. The Hall and Tajvidi estimator is as follows

$$\frac{1}{\hat{A}_{HT}(t)} = \frac{1}{n}\sum_{i=1}^{n}\overline{\xi}_i(t). \tag{3.10}$$

Note that this estimator satisfies only on two conditions of the Pickands dependence function, endpoints and $\hat{A}_{HT}(t) \geq max(t, 1-t)$.

Capéraà, Fougéres and Genest (1997) proposed an estimator called CFG-estimator. This estimator is consistent and asymptotically normal. For every t in $[0,1]$, the CFG-estimator is given by

$$\log\hat{A}_{CFG}(t) = -\frac{1}{n}\sum_{i=1}^{n}\log\xi_i(t) - (1-t)\sum_{i=1}^{n}\log\xi_i(0) - t\sum_{i=1}^{n}\log\xi_i(1). \tag{3.11}$$

3.4 Some Parametric Models

Constructing a Pickands dependence function under the conditions mentioned in the Theorem 3.2 is one of the popular methods to obtain an extreme value copula. In the following subsections, these copulas are introduced.

Recall that any bivariate extreme value copula can be written as

$$C(F(x),G(y)) = \exp\left[\log(F(x)G(y))A\left(\frac{\log G(y)}{\log(F(x)G(y))}\right)\right] \tag{3.12}$$

where $F(x)$ and $G(y)$ are the univariate marginal cumulative distribution functions and A is the Pickands dependence function. It should be noted that the marginal distribution functions which used in this thesis are the generelized extreme value distribution functions.

Let $H(x,y)$ be a bivariate extreme value distribution function with margins $F(x)$ and $G(y)$ which are given by

$$F(x) = \exp\left[-\left[1+\Lambda_1\left(\frac{x-\gamma_1}{\delta_1}\right)\right]^{-\frac{1}{\xi_1}}\right] \tag{3.13}$$

$$G(y) = \exp\left[-\left[1+\Lambda_2\left(\frac{y-\gamma_2}{\delta_2}\right)\right]^{-\frac{1}{\xi_2}}\right] \tag{3.14}$$

Taking $\omega_1(x) = \left[1+\Lambda_1\left(\frac{x-\gamma_1}{\delta_1}\right)\right]^{-\frac{1}{\Lambda_1}}$ and $\omega_2(y) = \left[1+\Lambda_2\left(\frac{y-\gamma_2(y)}{\delta_2}\right)\right]^{-\frac{1}{\Lambda_2}}$, and replacing $F(x)$ and $G(y)$ by $\exp(-\omega_1(x))$ and $\exp(-\omega_2(y))$ respectively in form 3.12, thus

$$\begin{aligned} H(x,y) &= C(e^{-\omega_1(x)},\, e^{-\omega_2(y)}) && (3.15)\\ &= \exp\left[\log(e^{-\omega_1(x)}e^{-\omega_2(y)})A\left(\frac{\log e^{-\omega_2(y)}}{\log(e^{-\omega_1(x)}e^{-\omega_2(y)})}\right)\right] && (3.16)\\ &= \exp\left[-(\omega_1(x)+\omega_2(y))A\left(\frac{\omega_2(y)}{\omega_1(x)+\omega_2(y)}\right)\right] && (3.17) \end{aligned}$$

3.4.1 Logistic or Gumbel-Hougaard Model

Logistic or Gumbel-Hougaard copula dating back to Gumbel (1960) and Hougaard (1986) can be considered as one of the oldest bivariate extreme value models. Logistic copula is the only copula which is at the same time extreme value and Archimedean copula. This copula has asymmetric shape with more weight in the right tail. The Pickands dependence function of Logistic model with dependence parameter r is given by

$$A(t) = \left[t^{\frac{1}{r}} + (1-t)^{\frac{1}{r}}\right]^r \quad 0 < r \leq 1 \tag{3.18}$$

where $t = \frac{\omega_2(y)}{\omega_1(x)+\omega_2(y)}$. Replacing function A in the form (3.17) by (3.18) yields

$$\begin{aligned} H(x,y) &= \exp\left[-(\omega_1(x)+\omega_2(y))A\left(\frac{\omega_2(y)}{\omega_1(x)+\omega_2(y)}\right)\right] \\ &= \exp\left[-(\omega_1(x)+\omega_2(y))\left[\left(\frac{\omega_2(y)}{\omega_1(x)+\omega_2(y)}\right)^{\frac{1}{r}}+\left(1-\frac{\omega_2(y)}{\omega_1(x)+\omega_2(y)}\right)^{\frac{1}{r}}\right]^r\right] \\ &= \exp\left[-(\omega_1(x)+\omega_2(y))\left[\frac{\omega_1(x)^{\frac{1}{r}}+\omega_2(y)^{\frac{1}{r}}}{(\omega_1(x)+\omega_2(y))^{\frac{1}{r}}}\right]^r\right] \end{aligned}$$

then the bivariate Gumbel-Hougaard distribution is obtained as follows

$$H(x,y) = \exp\left[-\left(\omega_1(x)^{\frac{1}{r}}+\omega_2(y)^{\frac{1}{r}}\right)^r\right] \tag{3.19}$$

and also by letting $\omega_1(x) = -\log u$ and $\omega_2(y) = -\log v$ in relationship (3.19), we can obtain the bivariate Gumbel-Hougaard copula as follows,

$$C(u,v) = \exp\left[-\left((-\log u)^{\frac{1}{r}}+(-\log v)^{\frac{1}{r}}\right)^r\right]. \tag{3.20}$$

Note that the Kendall's tau coefficient is $1-r$, where $0 < r \leq 1$. See Genest and Mackay (1986). The complete dependence and the independence are obtained when $r \to 0$ and $r = 1$ respectively.

In the other models, we only present the bivariate distribution function and associated Pickands dependence function, one can replace the corresponding Pickands dependence function in relationship (3.17) and obtain the copula.

3.4.2 Asymmetric Logistic Model

The bivariate Asymmetric Logistic distribution function with the parameters $0 < r \leq 1$, α and β, that is presented by Tawn (1988) is as follows

$$H(x,y) = \exp\left[-(1-\alpha)\omega_1(x)-(1-\beta)\omega_2(y)-\left[(\alpha\omega_1(x))^{\frac{1}{r}}+(\beta\omega_2(y))^{\frac{1}{r}}\right]^r\right] \tag{3.21}$$

where $0 \leq \alpha, \beta \leq 1$. The Asymmetric Logistic copula adds further flexibility to the Gumbel-Hougaard copula. Note that by taking $\alpha = \beta = 1$ in relationship (3.21), we can get the Logistic model, and also by letting $\alpha = \beta$, the Asymmetric Logistic copula is symmetric. The Pickands dependence function of Asymmetric Logistic copula is as follows

$$A(t) = (1-\alpha)(1-t)+(1-\beta)t+\left[(\alpha t)^{\frac{1}{r}}+(\beta(1-t))^{\frac{1}{r}}\right]^r. \tag{3.22}$$

The complete dependence is obtained when $\alpha = \beta = 1$ and $r \to 0$. And also the independence is obtained when $r = 1$ and $\alpha = 0$ or $\beta = 0$.

3.4.3 Negative Logistic or Galambos Model

The bivariate distribution function of Negative Logistic model dating back to Galambos (1975) is given by

$$H(x,y) = \exp\left[-\omega_1(x) - \omega_2(y) + \left[\omega_1(x)^{-r} + \omega_2(y)^{-r}\right]^{-\frac{1}{r}}\right], \quad r > 0 \tag{3.23}$$

and the associated Pickands dependence function is given by

$$A(t) = 1 - \left[t^{-r} + (1-t)^{-r}\right]^{-\frac{1}{r}}. \tag{3.24}$$

Independence is obtained as $r \to 0$ and complete dependence is obtained when $r \to \infty$.

3.4.4 Asymmetric Negative Logistic Model

The bivariate Asymmetric Negative Logistic copula dating back to Joe (1990) is an extension of Negative Logistic copula. The Joe copula has two parameters α and β which allow the model to be asymmetric. The distribution function of Asymmetric Negative Logistic model is given by

$$H(x,y) = \exp\left[-\omega_1(x) - \omega_2(y) + \left[(\alpha\omega_1(x))^{-r} + (\beta\omega_2(y))^{-r}\right]^{-\frac{1}{r}}\right], \quad r > 0 \tag{3.25}$$

where $0 \leq \alpha, \beta \leq 1$. The associated Pickands dependence function is given by

$$A(t) = 1 - \left[(\alpha t)^{-r} + (\beta(1-t))^{-r}\right]^{-\frac{1}{r}}. \tag{3.26}$$

Note that if $\alpha = \beta = 1$, we get the Negative Logistic copula. If $\alpha = \beta$, then the Asymmetric Negative Logistic copula is symmetric. Independence is obtained as $\alpha = \beta = 0$ or $r \to 0$ and complete dependence is obtained when $\alpha = \beta = 1$ and $r \to \infty$.

3.4.5 Hüsler-Reiss Model

The bivariate Hüsler-Reiss distribution function with parameter $r > 0$ is

$$H(x,y) = \exp\left[-\omega_1(x)\Phi\left[\frac{1}{r} + \frac{r}{2}\left(\log\left(\frac{\omega_1(x)}{\omega_2(y)}\right)\right)\right] - \omega_2(y)\Phi\left[\frac{1}{r} + \frac{r}{2}\left(\log\left(\frac{\omega_2(y)}{\omega_1(x)}\right)\right)\right]\right], \tag{3.27}$$

where $\Phi(.)$ is the standard normal distribution function. The Pickands dependence function is given by

$$A(t) = (1-t)\Phi(Z_{1-t}) + t\Phi(Z_t) \tag{3.28}$$

where $Z_t = \left(\frac{1}{r} + \frac{r}{2}\log\frac{t}{1-t}\right)$. Independence is obtained as $r \to 0$ and complete dependence is obtained when $r \to \infty$. For more details, see Hüsler and Reiss (1989).

3.4.6 t − EV *Model*

The Extreme Value **t** copula dating back to Demarta and McNeil (2005) is given by

$$C(u,v) = \int_{-\infty}^{T_n^{-1}(u)} \int_{-\infty}^{T_n^{-1}(v)} \frac{1}{\pi n |P|^{0.5}} \frac{\Gamma(\frac{n}{2}+1)}{\Gamma(\frac{n}{2})} \left(1 + \frac{x' P^{-1} x}{n}\right)^{-\frac{n}{2}+1} dx, \tag{3.29}$$

where T_n represents the univariate t distribution with n degrees of freedom and P represents the 2×2 correlation matrix. The associated Pickands dependence function is

$$A(t) = (1-t)T_{n+1}(Z_{1-t}) + tT_{n+1}(Z_t), \tag{3.30}$$

where $Z_t = (1+n)^{\frac{1}{2}} \left[\left(\frac{t}{1-t}\right)^{\frac{1}{n}} - \rho \right] (1-\rho^2)^{-\frac{1}{2}}$ for $n > 0$ and $\rho \in (-1,1)$. For more details, see Nikoloulopoulos et al. (2009).

CHAPTER FOUR
THE BERNSTEIN COPULA

Bernstein polynomials play an important and an effective role in the mathematics. These polynomials are frequently used in the approximation theory. They were first used by Sergei Natanovich Bernstein (1912) to prove the *Weierstrass* approximation theorem. It can be said that the application of Bernstein polynomials on the approximation of the copula has been studied for the first time by Li et al. (1997).

4.1 Bernstein Polynomials

According to the following theorem that is known as Weierstrass theorem, any continuous function defined on closed interval $[a,b]$ can be uniformly approximated by a polynomial.

Theorem 4.1. Let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then, for any $\varepsilon > 0$, there exists a polynomial $p(x)$, such that

$$|f(x) - p(x)| < \varepsilon \quad for\ all\ x \in [a,b].$$

Definition 4.1. Consider a continuous function $f(x)$ on $[0,1]$, the Bernstein polynomial of degree n is defined as follows,

$$B_n(f;x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) b_k^n(x) \tag{4.1}$$

where n is positive integer and $b_k^n(x)$ is called Bernstein basis of degree n and is defined by

$$b_k^n(x) = \binom{n}{k} x^k (1-x)^{n-k} \quad k = 0,1,\cdots,n$$

and

$$\sum_{k=0}^{n} b_k^n(x) = 1.$$

One of the important features of Bernstein polynomials is that they preserve shape properties of the function f. For example, if f is convex function, then $B_n(f,x)$ is also convex. Besides, the Bernstein polynomial of function $f(x)$ on $[0,1]$ interpolates $f(x)$ at the both endpoints of that interval. That is, $B_n(f;0) = f(0)$ and $B_n(f;1) = f(1)$.

Noting that replacing $f\left(\frac{k}{n}\right)$ by arbitrary coefficients β_k, $k = 0, 1, \cdots, n$ in formula 4.1, we obtain a polynomial in Bernstein form as follows

$$P_n(x) = \sum_{k=0}^{n} \beta_k \, b_k^n(x) \quad x \in [0,1].$$

Polynomial $P_n(x)$ is called *Bernstein-Bezier* polynomial or Bezier curve with control points $\beta_1, \beta_2, \cdots, \beta_n$. Mathematical basis for Bezier curve is Bernstein polynomials. Control points play a fundamental role in the Bezier curve, one can obtain control polygon by connecting the control points which can help to analyze the curve shape.

Definition 4.2 (Bivariate Case)**.** Let $f : \mathbf{I^2} \to \mathbb{R}$ be a continuous function, and let n, k_1 and k_2 be positive integers so that $0 \leq k_1, k_2 \leq n$, then the two dimensional Bernstein polynomial is as follows,

$$B_n(f;x,y) = \sum_{k_1=0}^{n} \sum_{k_2=0}^{n} f\left(\frac{k_1}{n}, \frac{k_2}{n}\right) b_{k_1}^n(x) b_{k_2}^n(y) \tag{4.2}$$

where

$$b_{k_1}^n(x) = \binom{n}{k_1} x^{k_1} (1-x)^{n-k_1}$$

and

$$b_{k_2}^n(y) = \binom{n}{k_2} y^{k_2} (1-y)^{n-k_2}.$$

Theorem 4.2. Let $f : \mathbf{I} \to \mathbb{R}$ be a continuous function. For any $\varepsilon > 0$, there exists an integer N such that

$$|f(x) - B_n(f;x)| < \varepsilon \quad x \in [0,1]$$

for all $n \geq N$.

See Phillips (2003).

Theorem 4.3. Let $f : \mathbf{I} \to \mathbb{R}$ be a continuous function, then the Bernstein polynomial $B_n(f)$ converges uniformly to f as $n \to \infty$. And if f is continuously differentiable on $\mathbf{I}$, then $B_n^{'}(f)$ converges uniformly to $f^{'}$ as $n \to \infty$.

See Phillips (2003).

4.1.1 Some Properties of Polynomials in Bernstein Form

For each x in $[0,1]$;

1. Non-negativity: $b_k^n(x) \geq 0$.

2. Symmetry: $b_k^n(x) = b_{n-k}^n(1-x)$.

3. Unimodality: $b_k^n(x)$ has a single extremum at $x = \frac{k}{n}$.

4. Endpoint interpolation: both β_0 and β_n are the endpoints of the curves, that is, $P_n(0) = \beta_0$ and $P_n(1) = \beta_n$.

5. Derivatives: $\frac{d}{dx} b_k^n(x) = n\left[b_{k-1}^{n-1}(x) - b_k^{n-1}(x)\right]$, where $b_{-1}^{n-1}(x) = b_n^{n-1}(x) = 0$. And $P_n^{'}(x) = \sum_{k=0}^{n} n\left[\beta_{k+1} - \beta_k\right] b_k^{n-1}(x)$.

6. Integration: $\int_0^1 b_k^n(x)dx = \frac{1}{n+1}$ and $\int_0^1 P_n(x)dx = \frac{1}{n+1}\sum_{k=0}^{n}\beta_k$.

7. Convex hull property: $P_n(x)$ always lies in the convex hull of the control points. Note that we say domain A is convex hull if for any two points in the A such as a and b, the segment $\overline{ab}$ is completely contained in the A.

8. Invariant property: $P_n(x)$ is invariant under affine transformation and affine parameter transformation.

4.2 Bernstein Copula

The Bernstein copula is fairly new concept in the literature, however it has a growing interest because of the flexible form of Bernstein polynomials. Also, Bernstein copulas have nonparametric description of dependence structure even in higher dimensions and they approximate any given copula arbitrarily well. Sancetta and Satchell (2004) introduced a family of copulas defined as Bernstein polynomials. Bernstein polynomials are closed under differentiation and with some simple restrictions on the coefficients, they always provide a proper copula. Bernstein copula leads to a general approach in estimation of many operations when a parametric form of copula is given. Also, the Bernstein copula can be used in place of original copula where the parametric copula may be available but in a very complicated form. The Bernstein polynomials have the best rate of convergence within the class of all operators which preserving the same shape compared to the other polynomial approximations. Note that in the following definitions and discussions, we use the symbol C_{BC} for the Bernstein copula.

Definition 4.3. Let $b_{k_1}^m(u) = \binom{m}{k_1} u^{k_1}(1-u)^{m-k_1}$ and $b_{k_2}^m(v) = \binom{m}{k_2} v^{k_2}(1-v)^{m-k_2}$, where

$\alpha(\frac{k_1}{m}, \frac{k_2}{m})$ is a real-valued constant indexed by $(k_1, k_2) \in N$ such that $0 \leq k_1, k_2 \leq m$. In terms of 2-dimensional Bernstein polynomial,

$$C_{BC}(u,v) = \sum_{k_1=0}^{m} \sum_{k_2=0}^{m} \alpha\left(\frac{k_1}{m}, \frac{k_2}{m}\right) b_{k_1}^m(u) b_{k_2}^m(v), \tag{4.3}$$

if $C_B : \mathbf{I^2} \to \mathbf{I}$ satisfies the properties of the copula function, then C_{BC} is a Bernstein copula for any $m \geq 1$.

The parameter m defined in 4.3 is the order of polynomial or the *grid-size* of the Bernstein copula. Indeed it can be said that the Bernstein copula is evaluated at each vertex of m grid by the constant $\alpha\left(\frac{k_1}{m}, \frac{k_2}{m}\right)$. The constant α is smoothed by the product of $b_{k_1}^m(u)$ and $b_{k_2}^m(v)$.

4.2.1 Some Properties of the Bernstein Copula

In this subsection, some of the important properties of the Bernstein copula are provided as follows

1. C_{BC} is increasing in all its arguments.

2. C_{BC} satisfies the Fréchet Bounds, that is, $\max(0, u+v-1) \leq C_{BC} \leq \min(u,v)$.

3. uv is a copula for independent random variables.

4. C_{BC} is Libschitz, i.e. $|C_{BC}(x_1,...,x_k) - C_{BC}(y_1,...,y_k)| \leq \sum_{j=1}^{k} \left|x_j - y_j\right|$.

For more detailed theory, see Sancetta and Satchell (2004), Baker (2008) and Janssen et al. (2012).

4.3 Nonparametric Bernstein Copula

In the bivariate case, the nonparametric Bernstein copula can be represented as follows

$$\hat{C}_{BC}(u,v) = \sum_{k_1=0}^{m} \sum_{k_2=0}^{m} C_n\left(\frac{k_1}{m}, \frac{k_2}{m}\right) b_{k_1}^m(u) b_{k_2}^m(v) \tag{4.4}$$

where $C_n\left(\frac{k_1}{m}, \frac{k_2}{m}\right)$ is the empirical copula at the points $\left(\frac{k_1}{m}, \frac{k_2}{m}\right)$ that is given by

$$C_n\left(\frac{k_1}{m}, \frac{k_2}{m}\right) = \frac{1}{n} \sum_{i=1}^{n} I\left(\mathscr{R}_1^i \leq \frac{k_1}{m}, \mathscr{R}_2^i \leq \frac{k_2}{m}\right)$$

in which $I(.)$ is the indicator function, $\mathscr{R}_1^i = \frac{rank(X_i)}{n+1}$ and $\mathscr{R}_2^i = \frac{rank(Y_i)}{n+1}$.

Bernstein polynomial estimators have been utilized as smooth estimators for density and distribution functions. The idea of using them for copula estimation has been given in Sancetta and Satchell (2004). Recently, Janssen et al. (2012) have shown that the empirical Bernstein copula has better performance than usual empirical copula. They study about the asymptotic properties of this estimator: almost sure consistency rates and asymptotic normality. For more details and theorems, see Janssen et al. (2012).

CHAPTER FIVE
NEW ESTIMATION OF PICKANDS DEPENDENCE FUNCTION: KERNEL REGRESSION APPROACH

5.1 Introduction

In this chapter, we propose a new nonparametric approach using Bernstein approximation to estimate the new points which has flexible form. These new pairs can serve to estimate the unknown Pickands dependence function. Kernel regression method is then used to derive an estimator. Cormier et al. (2014) proposed this procedure for the first time. They used usual empirical copula to estimate points and utilized the Quadratic constrained B-spline smoothing to derive an estimator of Pickands dependence function.

Consider a random pair (X,Y) with joint distribution function H and continuous marginal distributions F and G, respectively, then following Pickands (1981), the pair (X,Y) has extreme value dependence if and only if its copula C can be expressed in the form

$$C(u,v) = P(F(X) \leq u, G(Y) \leq v) = \exp\left(\ln(uv)A\left[\frac{\ln v}{\ln(uv)}\right]\right) \tag{5.1}$$

for all $u, v \in \mathbf{I}$ where $A : [0,1] \to [\frac{1}{2}, 1]$ is convex. Let $t = \frac{\ln v}{\ln(uv)}$ and then $A(t) = \frac{\ln C(u,v)}{\ln(uv)}$. The transformation from(u,v) to $(t, A(t))$ reduce to a convex curve if and only if, the copula C is the form by (5.1). For more details, see Cormier et al. (2014).

As discussed in the previous chapter, Bernstein approximation method has more advantages such as shape preserving and flexibility than the other approximation methods. Since the Bernstein polynomials have flexible forms, the points obtained by our method are closer to the true Pickands dependence function. The plan for this chapter is as follows. In Section 5.2 we introduce the new approch called the pointwise estimation to obtain the pairs which can be used to estimate the dependence function. The performance of the new pointwise estimation is presented in Subsection 5.2.3. In Subsection 5.2.4 a simulation study is carried out to investigate the sampling properties of the new estimated pairs. In Section 5.3 we utilize the Kernel regression approach to derive a new estimator of the Pickands dependence function. The performance and sampling properties of this new estimator are presented in Subsections 5.3.1 and 5.3.2,

respectively.

5.2 New Pointwise Estimation of the Pickands Dependence Function

5.2.1 Pointwise Estimation of Pickands Dependence Function Based on the Usual Empirical Copula

Let $\{(X_i,Y_i)\}_{i=1}^n$ be a random sample of size n from an extreme value copula C and marginals F and G. Consider $\hat{F}_n$ and $\hat{G}_n$ be the empirical distributions of F and G which are defined respectively as

$$\hat{F}_n(x_i) = \frac{1}{n}\sum_{j=1}^{n} I(X_j \leq x_i)$$

$$\hat{G}_n(y_i) = \frac{1}{n}\sum_{j=1}^{n} I(Y_j \leq y_i).$$

Let $\hat{U}_i = \hat{F}_n(X_i)$ and $\hat{V}_i = \hat{G}_n(Y_i)$ for $i = 1,...,n$. The empirical copula of $\{(\hat{U}_i,\hat{V}_i)\}_{i=1}^n$ is as follows

$$\hat{C}_n(\hat{u}_i,\hat{v}_i) = \frac{1}{n}\sum_{j=1}^{n} I(\hat{U}_j \leq \hat{u}_i, \hat{V}_j \leq \hat{v}_i).$$

Define $T_i = \frac{\ln(\hat{V}_i)}{\ln(\hat{U}_i\hat{V}_i)}$ and $Z_i = \frac{\ln\hat{C}_n(\hat{U}_i,\hat{V}_i)}{\ln(\hat{U}_i\hat{V}_i)}$. Then, the pair

$$(T_i,Z_i) = \left\{\left(\frac{\ln(\hat{V}_i)}{\ln(\hat{U}_i\hat{V}_i)}, \frac{\ln\hat{C}_n(\hat{U}_i,\hat{V}_i)}{\ln(\hat{U}_i\hat{V}_i)}\right)\right\}_{i=1}^n$$

is the estimation of pairs $(t_i,A(t_i))$ or pointwise estimation of the Pickands dependence function based on usual empirical copula.

5.2.2 New Pointwise Estimation of Pickands Dependence Function Based on the Empirical Bernstein Copula

Let $\{(X_i,Y_i)\}_{i=1}^n$ be i.i.d. random vectors with extreme value copula C and marginals F and G. Let $\hat{F}_n$ and $\hat{G}_n$ be the empirical distributions of the samples X_i and Y_i, respectively Let $\hat{U}_i = \hat{F}_n(X_i)$ and $\hat{V}_i = \hat{G}_n(X_i)$ for $i = 1,...,n$, then the empirical Bernstein copula of $\{(\hat{U}_i,\hat{V}_i)\}_{i=1}^n$ is

$$\hat{C}_{BC}(\hat{u}_i,\hat{v}_i) = \sum_{k_1=0}^{m}\sum_{k_2=0}^{m} C_n\left(\frac{k_1}{m},\frac{k_2}{m}\right) b_{k_1,m}(\hat{u}_i) b_{k_2,m}(\hat{v}_i).$$

For more details, see section 3.3. Define $T_i = \frac{\ln(\hat{V}_i)}{\ln(\hat{U}_i\hat{V}_i)}$ and $Q_i = \frac{\ln\hat{C}_{BC}(\hat{U}_i,\hat{V}_i)}{\ln(\hat{U}_i\hat{V}_i)}$. Then, the pair

$$(T_i,Q_i) = \left\{\left(\frac{\ln(\hat{V}_i)}{\ln(\hat{U}_i\hat{V}_i)}, \frac{\ln\hat{C}_{BC}(\hat{U}_i,\hat{V}_i)}{\ln(\hat{U}_i\hat{V}_i)}\right)\right\}_{i=1}^n$$

is the estimation of pairs $(t_i, A(t_i))$ or new pointwise estimation of the Pickands dependence function based on empirical Bernstein copula.

5.2.3 Performance of the New Estimated Pairs

The new pairs of (T_i, Q_i) are obtained as $\left\{\left(\frac{\ln(\hat{V}_i)}{\ln(\hat{U}_i\hat{V}_i)}, \frac{\ln(\hat{C}_{BC}(\hat{U}_i,\hat{V}_i))}{\ln(\hat{U}_i\hat{V}_i)}\right)\right\}_{i=1}^{n}$ based on empirical Bernstein copula which are closer to the true values of Pickands dependence function. They are compared with the pairs calculated by the usual empirical copula. The behaviour of the estimated pairs using empirical Bernstein copula for the various orders of Bernstein polynomials, m $(7, 15, 25$ and $50)$ and usual empirical copula based on a sample of size 200 from five known extreme value distributions: Logistic (or Gumbel), Asymmetric Logistic with $(\alpha, \beta) = (0.5, 0.5)$, Negative Logistic (or Galambos), Asymmetric Negative Logistic with $(\alpha, \beta) = (0.5, 0.5)$ and Hüsler-Reiss models for four different levels of Kendall's tau coefficients, $\tau = 0.1, 0.25, 0.5$ and 0.75 are shown in the figures (5.1)-(5.9). Note that in order to study the effect of asymmetry, Asymmetric Logistic distribution with $(\alpha, \beta) = (0.35, 0.75)$ and $(\alpha, \beta) = (0.75, 0.35)$, and Asymmetric Negative Logistic distribution with $(\alpha, \beta) = (0.35, 0.75)$ and $(\alpha, \beta) = (0.75, 0.35)$ for three different levels of Kendall's correlation coefficients, $\tau = 0.1, 0.15$ and 0.25 are considered. The figures show whether an extreme-value model is appropriate for the data or not.

5.2.4 Simulation Study

A simulation study is designed to measure the performance and investigate the behavior of the new estimated pairs through extreme value distributions having generalized extreme value marginals. First, we generate $\{(X_i, Y_i)\}_{i=1}^{n}$ random vector of size 250 from five extreme value distributions; Symmetric Logistic, Asymmetric Logistic, Negative Logistic, Asymmetric Negative Logistic and Hüsler-Reiss distributions. Note that we consider four dependence levels as measured by Kendall's tau coefficient $\tau = 0.1, 0.25, 0.5$ and 0.75 to generate random samples, also in order to study the effect of asymmetry in both asymmetric logistic and asymmetric negative logistic families, $(\alpha, \beta) = (0.35, 0.75)$ and $(\alpha, \beta) = (0.75, 0.35)$ are considered in three different dependence levels as $\tau = 0.1, 0.15, 0.25$. See Tawn (1988).

In each case, we draw 1000 samples of size 250. For each sample, we approximate the points of Pickands dependence function using usual empirical copula as Cormier, Genest and Neslehova(2014) and also using empirical Bernstein copula for $m = 7, 15, 25, 50$, then we calculate true Pickands function for all extreme value distributions. To compare the efficiency of these two methods, we compute the MSEs of the estimated points via Pickands function. The main results of this simulation study are summarized in the following figures and tables.

Generally, Figures (5.1)-(5.9) show that the estimated data have best fit for small values of τ and low orders of Bernstein polynomials for the corresponding extreme value distributions. When order of the Bernstein polynomials gets smaller, the estimated data fits better. Also, all the fits are better than the usual emprical Pickands estimator. It is seen that all the constraints of Pickands dependence function are satisfied.

Tables (5.1)-(5.7) give median of Mean Squared Error (MSE) of Pickands dependence function estimated by empirical Bernstein copula and usual Pickands esimator for five known extreme value distributions. From the tables, it can be concluded that as the polynomial order decreases and the dependence level decreases, the proposed estimator has a better performance than the usual one. If there is high level of dependence, for a good performance, the order of the polynomials should be increased in order to have a better fit. Also if there is low level of dependence, then for a smaller MSE, the order of the polynomials should be decreased. In order to measure the asymmetry effect, Tables (5.6) and (5.7) are constructed for asymmetric logistic and asymmetric negative logistic extreme value distributions. For different values of asymmetry measure, it captures the tail behaviour. The results presented in Tables (5.1)-(5.7) are visually shown in Figures (5.10)-(5.18).

5.2.4.1 Performance Testing

The comparisons between the Pickands estimators are provided in terms of the MSEs. The significant comparisons are obtained using the sign test where the observed MSEs are based on the same data. Let M be the MSE of the estimated Pickands points using usual empirical copula and M^* be the estimated points using empirical Bernstein copula based on random sample of size 250 from five extreme value distributions. In order to have

statistically significant comparisons of MSE's in level 0.001 based on 1000 paired sample, we test $H_0 : P(M^* < M) = 0.5$. If $M^* < M$, then it means that the Pickands estimator using Bernstein copula is preferable than the Pickands estimator using usual empirical copula when H_0 is rejected at $\alpha = 0.001$ significance level with a one-sided test. The results are summarized in the tables (5.8)-(5.11) . Tables 8 and 10 show the probability of success in thousands of samples and Tables (5.9) and (5.11) show the significant comparisons of MSE's in level 0.001. The algorithm is given below. Let M be the MSE of the estimated points using usual empirical copula and M^* be the MSE of the estimated points using empirical Bernstein copula. We test the hypothesis

$$\begin{aligned} H_0 &: Median(M_i^* - M_i) = 0 \\ H_1 &: Median(M_i^* - M_i) < 0 \end{aligned}$$

1. Compute $D_i = M_i^* - M_i$ for each sample, that is, for $i = 1, \cdots, 1000$.

2. Let n^+ be the number of positive differences and n^- be the number of negative differences, i.e. $n^+ = \sum_{i=1}^{1000} I(D_i > 0)$ and $n^- = \sum_{i=1}^{1000} I(D_i < 0)$. Note that we ignore the equal cases.

3. Let $N = n^+ + n^-$, then

   - If $N = 0$ then the one-sided exact probability is $P = 0.5$.
   - If $0 < N \leq 25$, then $p - value = P_1 \wedge P_2$ where $P_1 = \sum_{i=1}^{N} \binom{N}{i} (0.5)^N$ and $P_2 = 1 - \sum_{i=1}^{N-1} \binom{N}{i} (0.5)^N$.
   - If $N > 25$, then the test statistic is

     $$Z^* = \frac{(n^+ \vee n^-) - 0.5N - 0.5}{0.5\sqrt{N}}.$$

     The one-sided $p - value$ is $Pr(Z \geq |Z^*|)$.

4. If $p - value$ is less than α, we reject the null hypothesis in favour of the alternative hypothesis where $n^+ < n^-$.

Tables (5.8) and (5.9) are constructed to make a significant comparison between MSE's of empirical Bernstein copula estimator and usual empirical estimator. The results show that the proposed estimator has smaller MSE, especially in low level of dependence for

Table 5.1 Median of Mean Squared Error of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Logistic copula.

		MSE					
Logistic		(T_i, Q_i)	(T_i, Z_i)			(T_i, Q_i)	(T_i, Z_i)
$\tau = 0.10$	$m = 7$	0.0005181	0.0031064	$\tau = 0.25$	$m = 7$	0.0018183	0.0032204
	$m = 15$	0.0006943	0.0031064		$m = 15$	0.0008410	0.0032204
	$m = 25$	0.0009127	0.0031064		$m = 25$	0.0008562	0.0032204
	$m = 50$	0.0012381	0.0031064		$m = 50$	0.0009523	0.0032204
$\tau = 0.50$	$m = 7$	0.0077470	0.0031959	$\tau = 0.75$	$m = 7$	0.020441	0.0025829
	$m = 15$	0.0028830	0.0031959		$m = 15$	0.0082820	0.0025829
	$m = 25$	0.0020943	0.0031959		$m = 25$	0.0051520	0.0025829
	$m = 50$	0.0012671	0.0031959		$m = 50$	0.0024172	0.0025829

all orders of Bernstein polynomials. For high level of dependence, the proposed estimator performance is better for higher order polynomials. In Tables (5.10) and (5.11), for the asymmetry case, all the levels of dependence and orders, the proposed estimator has a good performance.

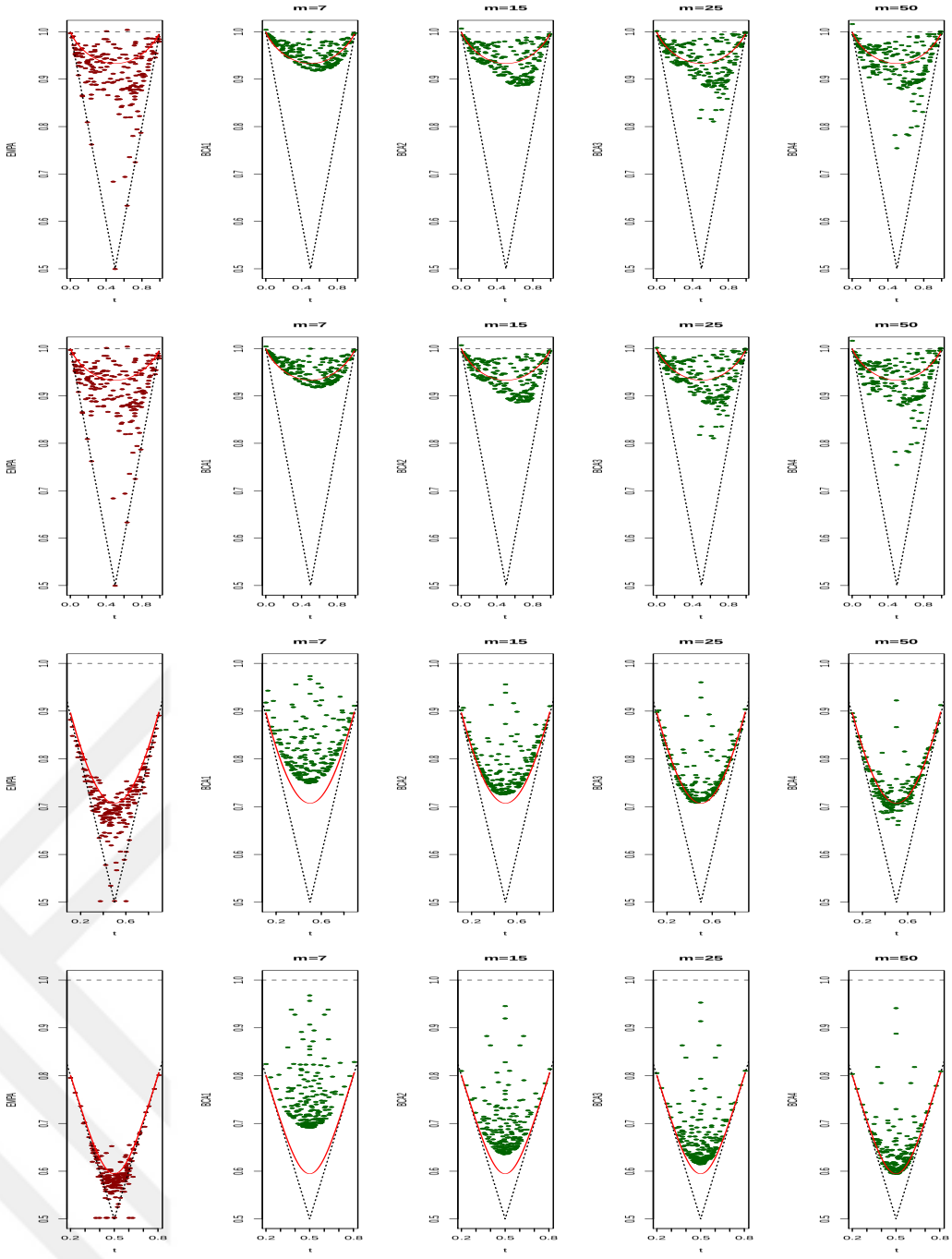


Figure 5.1 Plots based on sample size of 200 from the Logistic copula with $\tau = 0.1$ (top row), $\tau = 0.25$ (second row), $\tau = 0.5$ (third row) and $\tau = 0.75$ (bottom row). In each row, the left columns correspond to the estimated pairs by the usual empirical copula and the next four columns are correspond to the estimated pairs by the empirical Bernstein copula for $m = 7$ (second column), $m = 15$ (third column), $m = 25$ (fourth column) and $m = 50$ (fifth column).

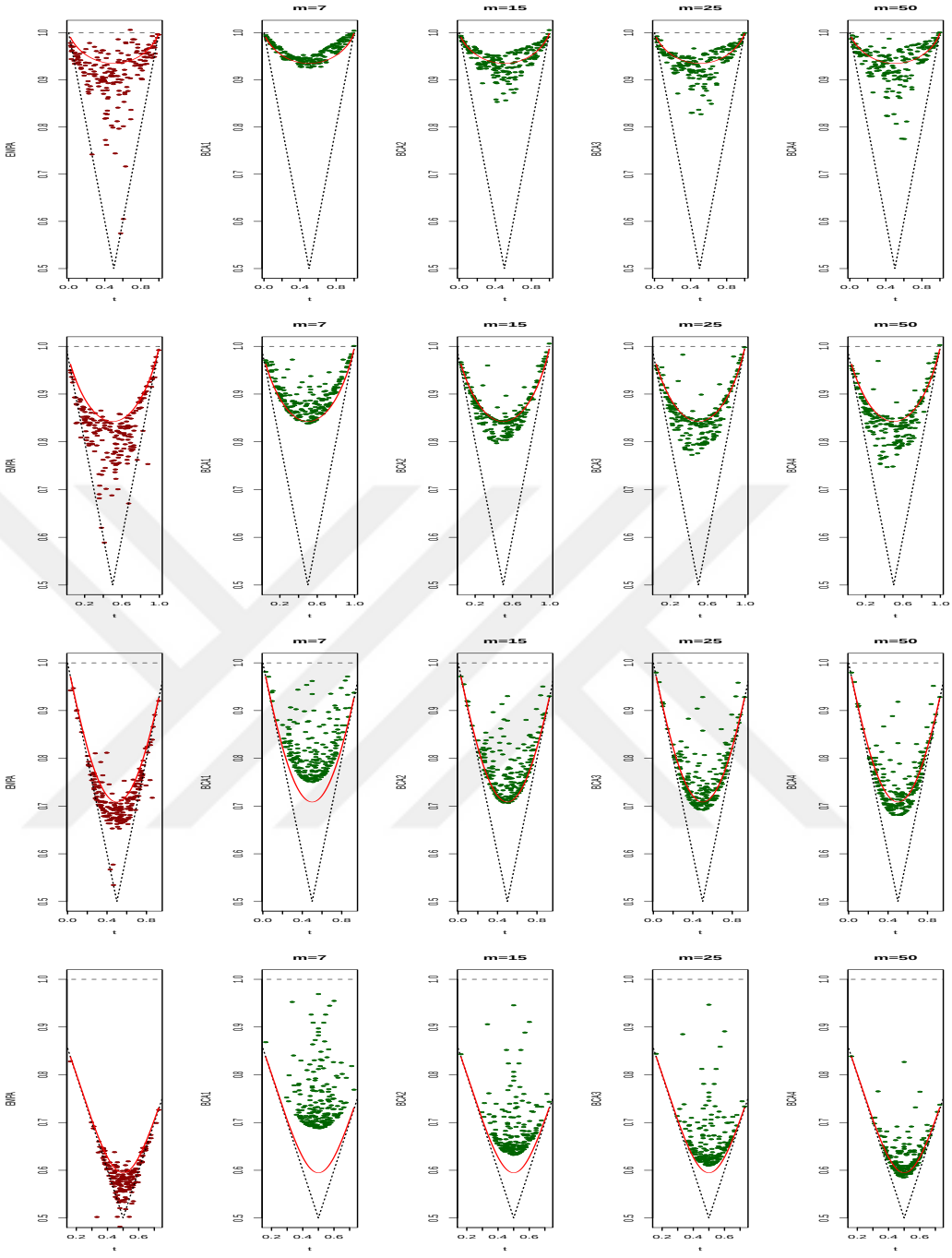


Figure 5.2 Plots based on sample size of 200 from the Negative Logistic copula with $\tau = 0.1$ (top row), $\tau = 0.25$ (second row), $\tau = 0.5$ (third row) and $\tau = 0.75$ (bottom row). In each row, the left columns correspond to the estimated pairs by the usual empirical copula and the next four columns are correspond to the estimated pairs by the empirical Bernstein copula for $m = 7$ (second column), $m = 15$ (third column), $m = 25$ (fourth column) and $m = 50$ (fifth column).

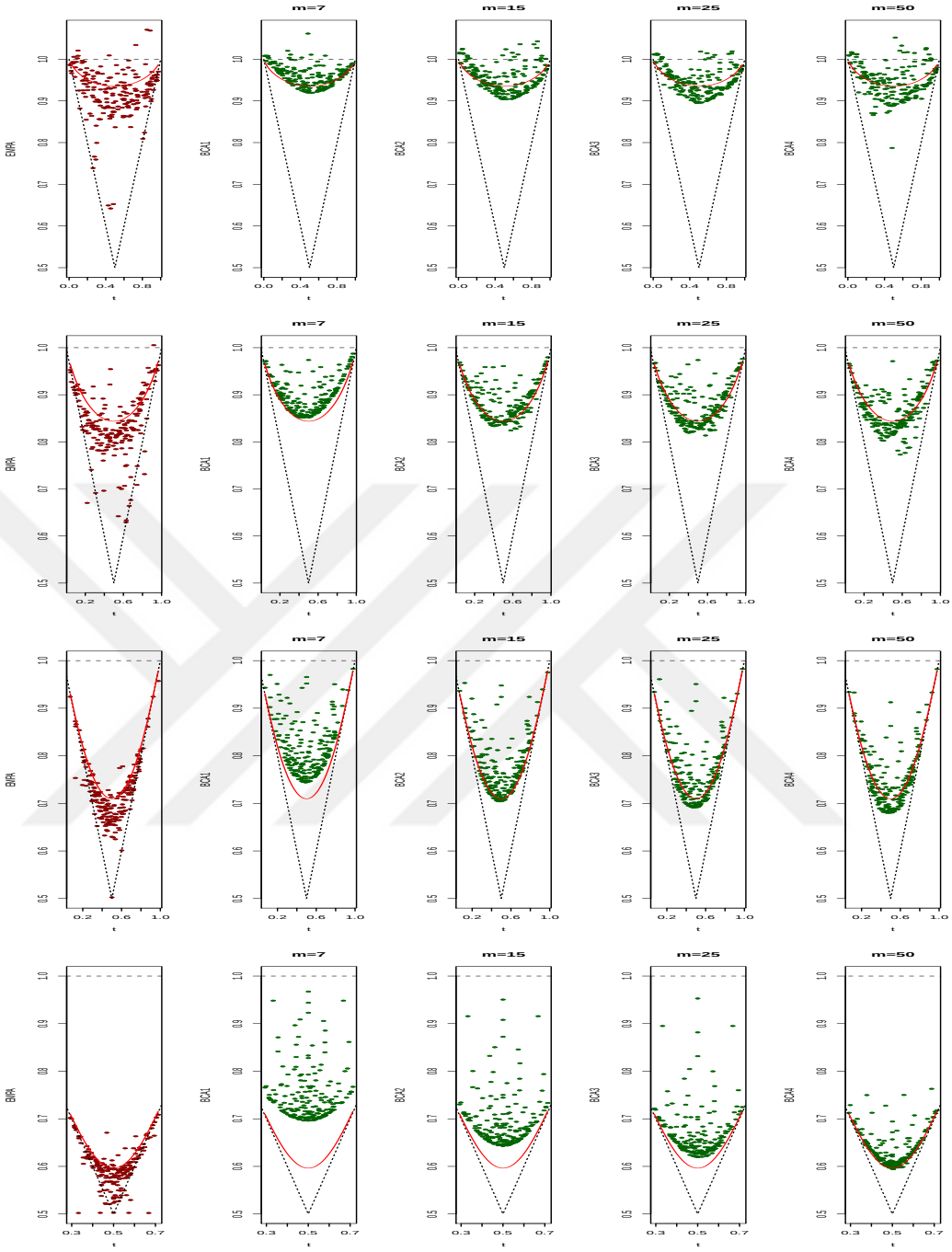


Figure 5.3 Plots based on sample size of 200 from the Hüsler-Reiss copula with $\tau = 0.1$ (top row), $\tau = 0.25$ (second row), $\tau = 0.5$ (third row) and $\tau = 0.75$ (bottom row). In each row, the left columns correspond to the estimated pairs by the usual empirical copula and the next four columns are correspond to the estimated pairs by the empirical Bernstein copula for $m = 7$ (second column), $m = 15$ (third column), $m = 25$ (fourth column) and $m = 50$ (fifth column).

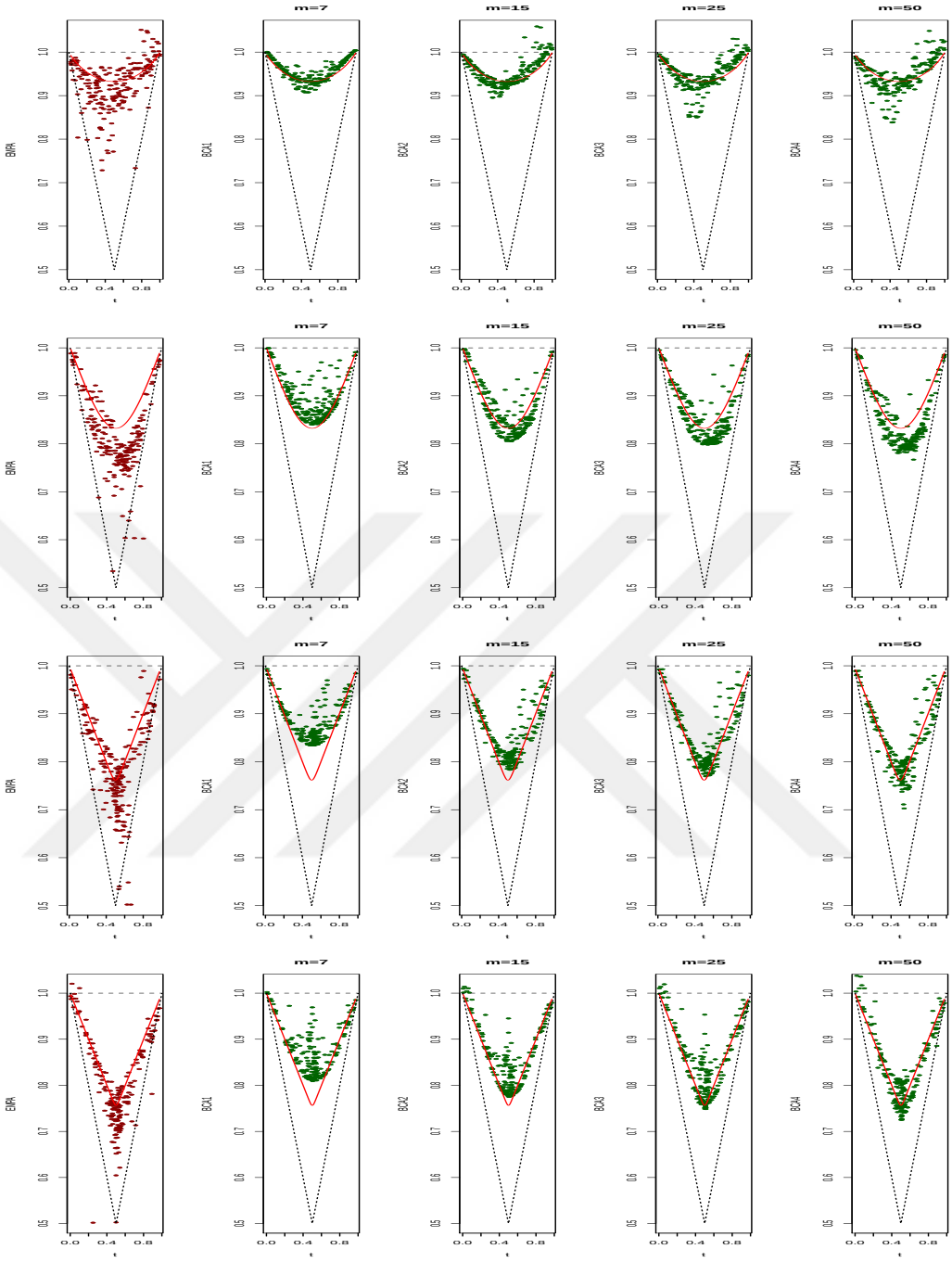


Figure 5.4 Plots based on sample size of 200 from the Asymmetric Logistic copula ($\alpha = 0.5$, $\beta = 0.5$) with $\tau = 0.1$ (top row), $\tau = 0.25$ (second row), $\tau = 0.5$ (third row) and $\tau = 0.75$ (bottom row). In each row, the left columns correspond to the estimated pairs by the usual empirical copula and the next four columns are correspond to the estimated pairs by the empirical Bernstein copula for $m = 7$ (second column), $m = 15$ (third column), $m = 25$ (fourth column) and $m = 50$ (fifth column).

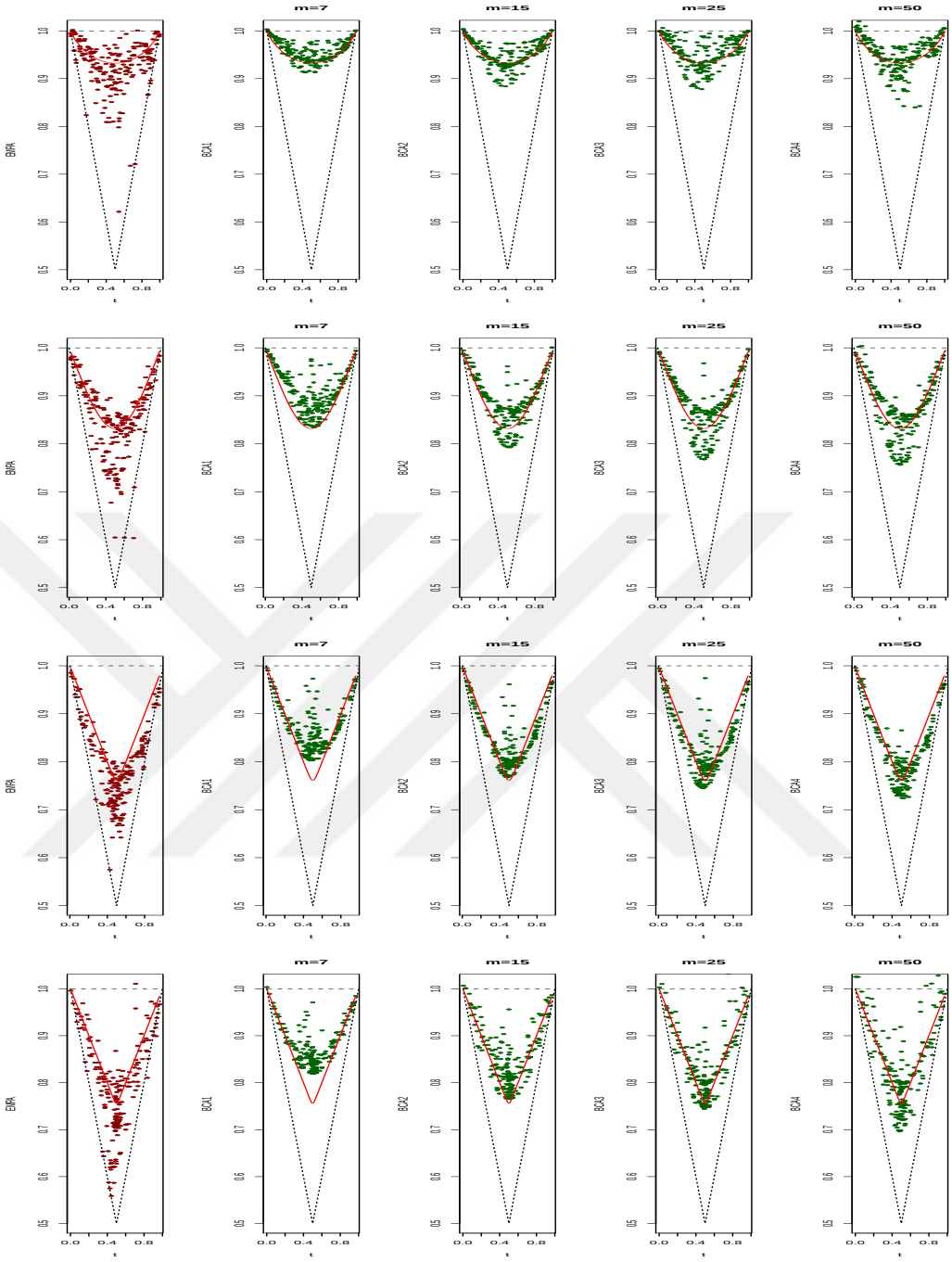


Figure 5.5 Plots based on sample size of 200 from the Asymmetric Negative Logistic copula ($\alpha = 0.5$, $\beta = 0.5$) with $\tau = 0.1$ (top row), $\tau = 0.25$ (second row), $\tau = 0.5$ (third row) and $\tau = 0.75$ (bottom row). In each row, the left columns correspond to the estimated pairs by the usual empirical copula and the next four columns are correspond to the estimated pairs by the empirical Bernstein copula for $m = 7$ (second column), $m = 15$ (third column), $m = 25$ (fourth column) and $m = 50$ (fifth column).

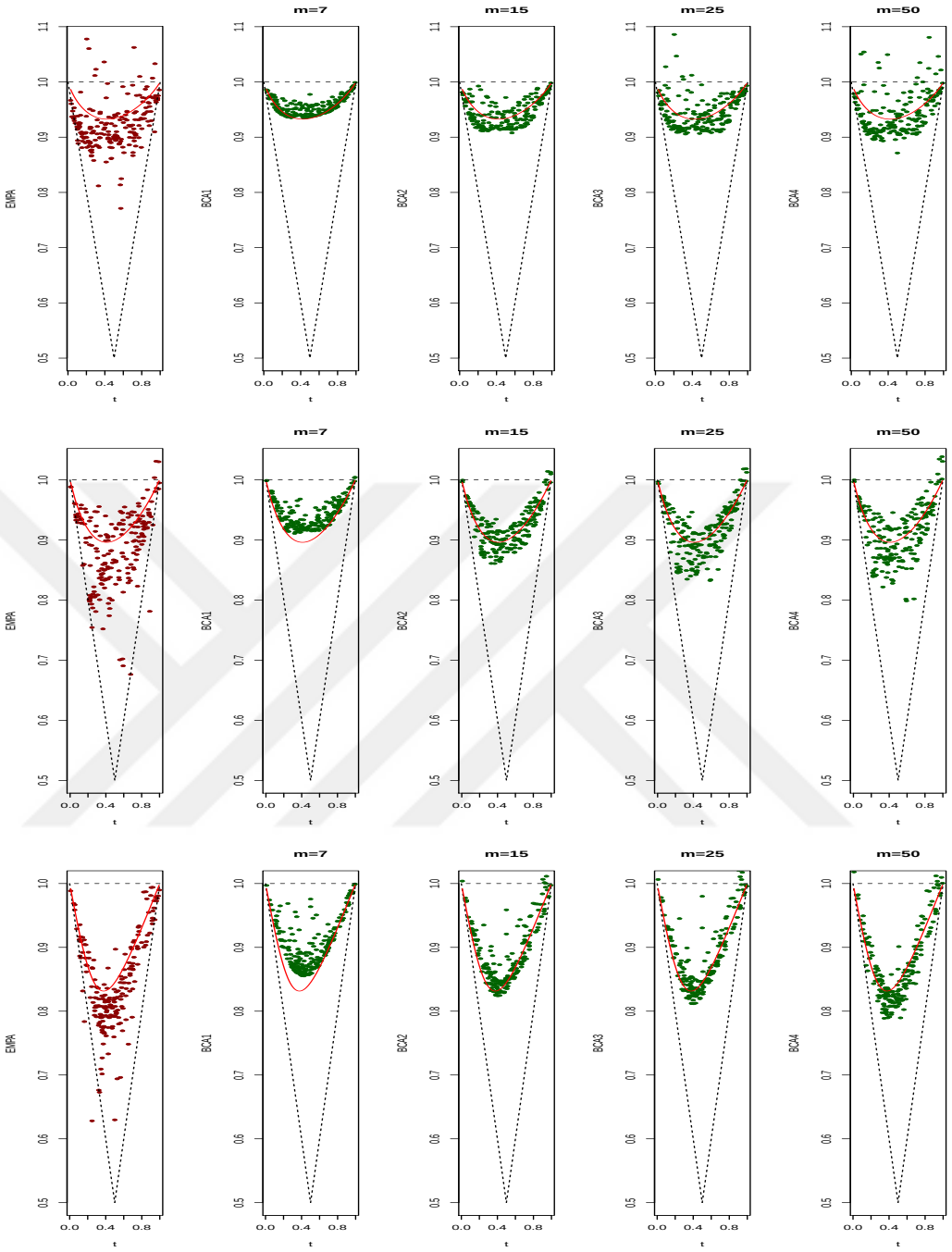


Figure 5.6 Plots based on sample size of 200 from the Asymmetric Logistic copula ($\alpha = 0.35$, $\beta = 0.75$) with $\tau = 0.1$ (top row), $\tau = 0.15$ (second row) and $\tau = 0.25$(bottom row). In each row, the left columns correspond to the estimated pairs by the usual empirical copula and the next four columns are correspond to the estimated pairs by the empirical Bernstein copula for $m = 7$ (second column), $m = 15$ (third column), $m = 25$ (fourth column) and $m = 50$ (fifth column).

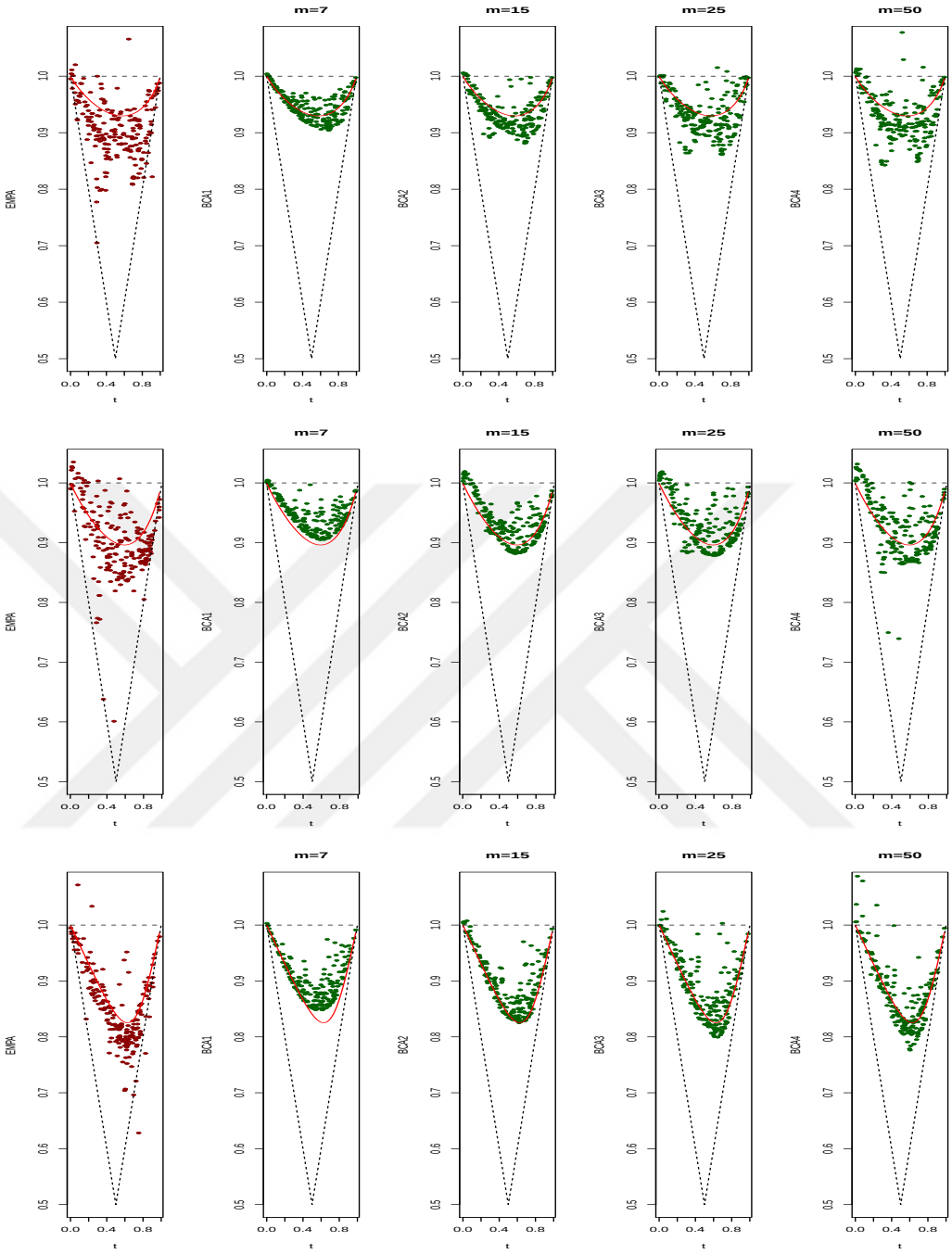


Figure 5.7 Plots based on sample size of 200 from the Asymmetric Logistic copula ($\alpha = 0.75$, $\beta = 0.35$) with $\tau = 0.1$ (top row), $\tau = 0.15$ (second row), and $\tau = 0.25$ (bottom row). In each row, the left columns correspond to the estimated pairs by the usual empirical copula and the next four columns are correspond to the estimated pairs by the empirical Bernstein copula for $m = 7$ (second column), $m = 15$ (third column), $m = 25$ (fourth column) and $m = 50$ (fifth column).

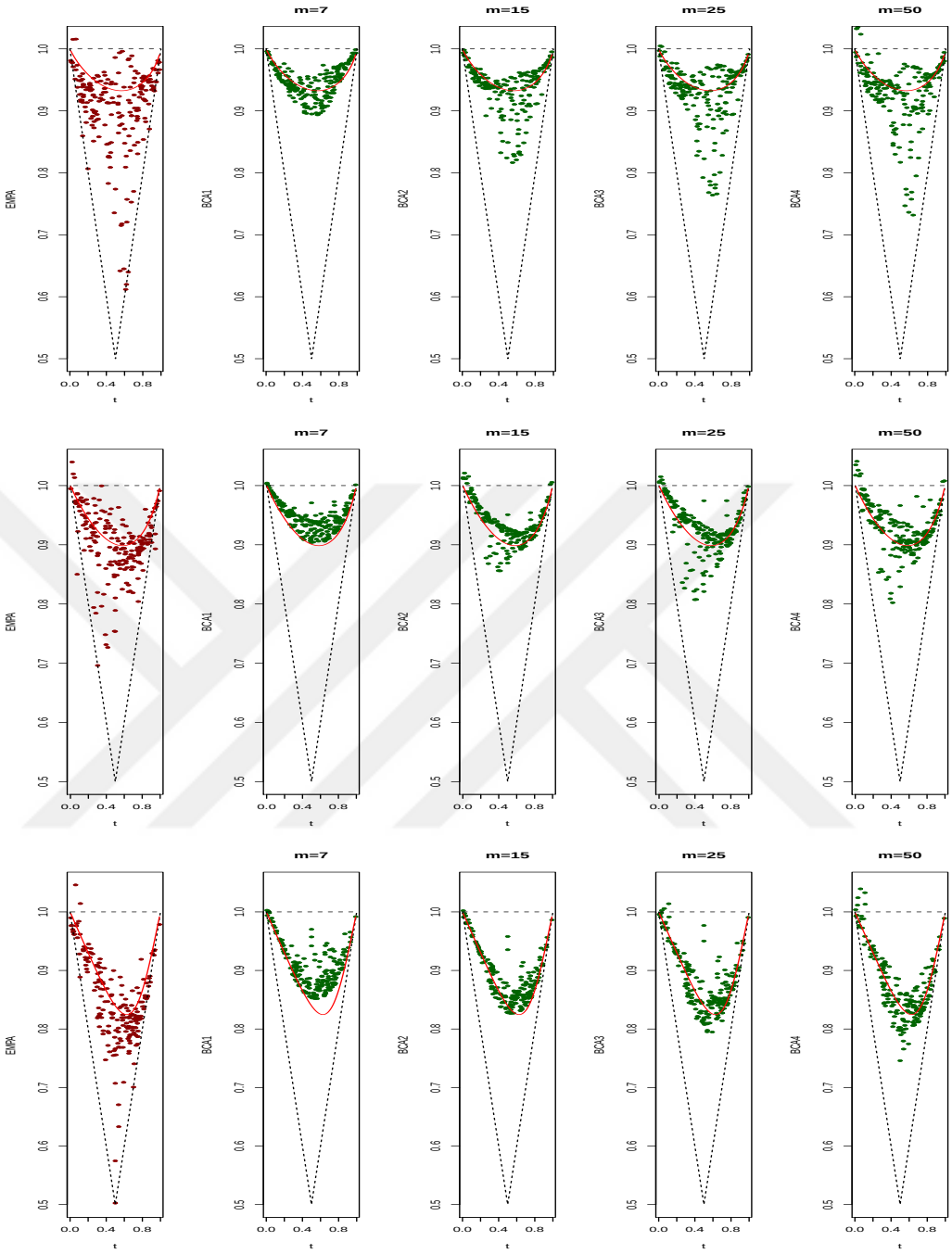


Figure 5.8 Plots based on sample size of 200 from the Asymmetric Negative Logistic copula ($\alpha = 0.35$, $\beta = 0.75$) with $\tau = 0.1$ (top row), $\tau = 0.15$ (second row) and $\tau = 0.25$ (bottom row). In each row, the left columns correspond to the estimated pairs by the usual empirical copula and the next four columns are correspond to the estimated pairs by the empirical Bernstein copula for $m = 7$ (second column), $m = 15$ (third column), $m = 25$ (fourth column) and $m = 50$ (fifth column).

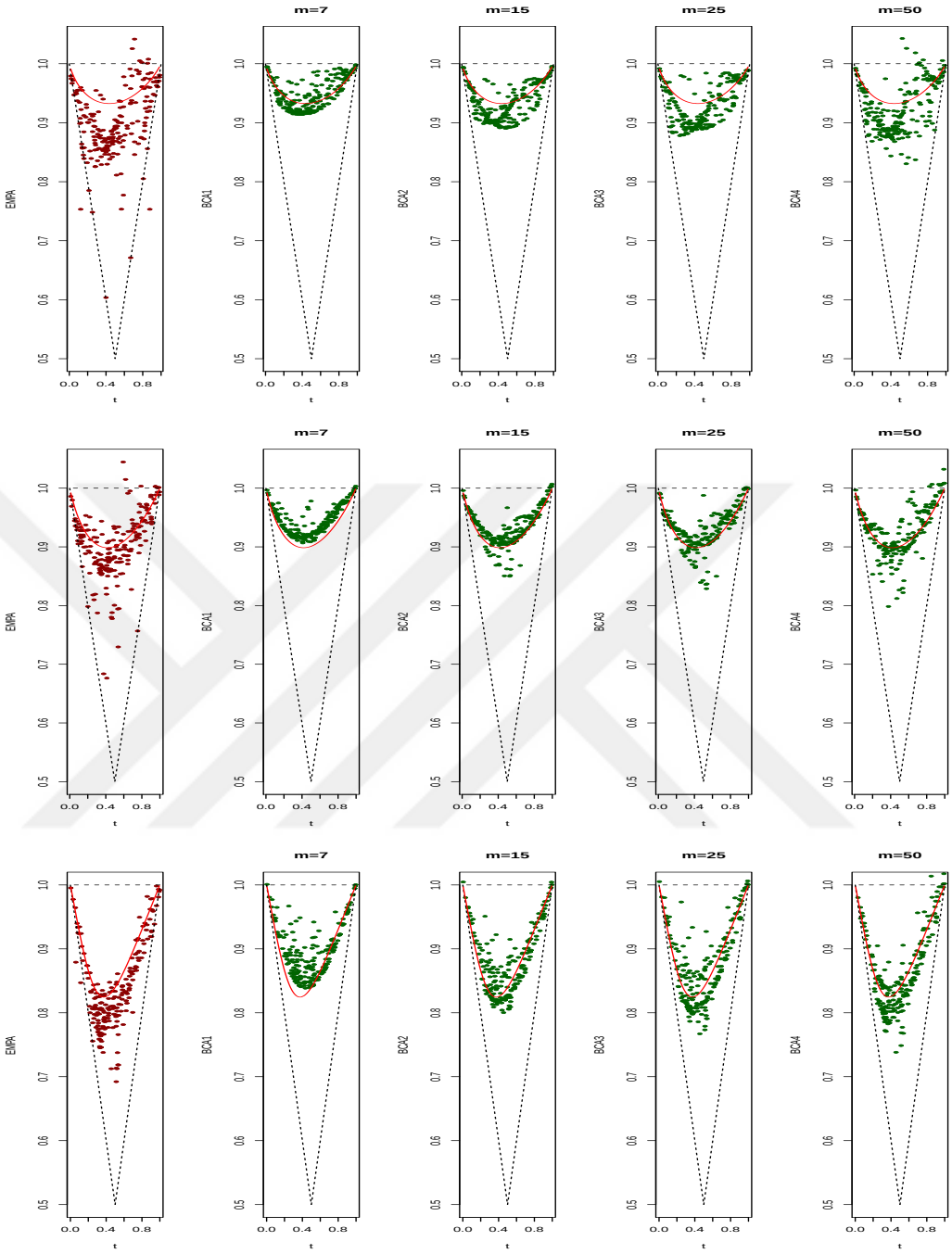


Figure 5.9 Plots based on sample size of 200 from the Asymmetric Negative Logistic copula ($\alpha = 0.75$, $\beta = 0.35$) with $\tau = 0.1$ (top row), $\tau = 0.15$ (second row), and $\tau = 0.25$ (bottom row). In each row, the left columns correspond to the estimated pairs by the usual empirical copula and the next four columns are correspond to the estimated pairs by the empirical Bernstein copula for $m = 7$ (second column), $m = 15$ (third column), $m = 25$ (fourth column) and $m = 50$ (fifth column).

Table 5.2 Median of Mean Squared Error of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Asymmetric Logistic copula $(0.5, 0.5)$.

		MSE					
Asy. Log.		(T_i, Q_i)	(T_i, Z_i)			(T_i, Q_i)	(T_i, Z_i)
$\tau = 0.10$	$m = 7$	0.0004569	0.0030520	$\tau = 0.25$	$m = 7$	0.0020214	0.0030715
	$m = 15$	0.0006794	0.0030520		$m = 15$	0.0009685	0.0030715
	$m = 25$	0.0009157	0.0030520		$m = 25$	0.0009943	0.0030715
	$m = 50$	0.0012484	0.0030520		$m = 50$	0.0010400	0.0030715
$\tau = 0.50$	$m = 7$	0.0049280	0.0033201	$\tau = 0.75$	$m = 7$	0.0052290	0.0035151
	$m = 15$	0.0021839	0.0033201		$m = 15$	0.0023254	0.0035151
	$m = 25$	0.0017074	0.0033201		$m = 25$	0.0018098	0.0035151
	$m = 50$	0.0013055	0.0033201		$m = 50$	0.0013422	0.0035151

Table 5.3 Median of Mean Squared Error of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Negative Logistic copula.

		MSE					
Neg. Log.		(T_i, Q_i)	(T_i, Z_i)			(T_i, Q_i)	(T_i, Z_i)
$\tau = 0.10$	$m = 7$	0.0005034	0.0030474	$\tau = 0.25$	$m = 7$	0.0018759	0.0031863
	$m = 15$	0.0006820	0.0030474		$m = 15$	0.0009158	0.0031863
	$m = 25$	0.0008898	0.0030474		$m = 25$	0.0009302	0.0031863
	$m = 50$	0.0011929	0.0030474		$m = 50$	0.0010022	0.0031863
$\tau = 0.50$	$m = 7$	0.0078540	0.0032004	$\tau = 0.75$	$m = 7$	0.0204900	0.0026030
	$m = 15$	0.0029790	0.0032004		$m = 15$	0.0083540	0.0026030
	$m = 25$	0.0020229	0.0032004		$m = 25$	0.0052170	0.0026030
	$m = 50$	0.0011975	0.0032004		$m = 50$	0.0015901	0.0026030

Table 5.4 Median of Mean Squared Error of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Asymmetric Negative Logistic copula $(0.5, 0.5)$.

		MSE					
Asy. Neg.		(T_i, Q_i)	(T_i, Z_i)			(T_i, Q_i)	(T_i, Z_i)
$\tau = 0.10$	$m = 7$	0.0005030	0.0030243	$\tau = 0.25$	$m = 7$	0.0020622	0.0031201
	$m = 15$	0.0006856	0.0030243		$m = 15$	0.0009991	0.0031201
	$m = 25$	0.0008902	0.0030243		$m = 25$	0.0009937	0.0031201
	$m = 50$	0.0011993	0.0030243		$m = 50$	0.0010351	0.0031201
$\tau = 0.50$	$m = 7$	0.0050850	0.0033347	$\tau = 0.75$	$m = 7$	0.0053621	0.0033664
	$m = 15$	0.0022871	0.0033347		$m = 15$	0.0024591	0.0033664
	$m = 25$	0.0017673	0.0033347		$m = 25$	0.0018716	0.0033664
	$m = 50$	0.0013034	0.0033347		$m = 50$	0.0013597	0.0033664

Table 5.5 Median of Mean Squared Error of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Hüsler-Reiss copula.

		MSE					
Hüsler-Reiss		(T_i, Q_i)	(T_i, Z_i)			(T_i, Q_i)	(T_i, Z_i)
$\tau = 0.10$	$m = 7$	0.0005047	0.0030510	$\tau = 0.25$	$m = 7$	0.0018800	0.0031868
	$m = 15$	0.0006834	0.0030510		$m = 15$	0.0009127	0.0031868
	$m = 25$	0.0008826	0.0030510		$m = 25$	0.0009343	0.0031868
	$m = 50$	0.0011929	0.0030510		$m = 50$	0.0009958	0.0031868
$\tau = 0.50$	$m = 7$	0.0077851	0.0031977	$\tau = 0.75$	$m = 7$	0.0202411	0.0025887
	$m = 15$	0.0029550	0.0031977		$m = 15$	0.0082120	0.0025887
	$m = 25$	0.0020017	0.0031977		$m = 25$	0.0051215	0.0025887
	$m = 50$	0.0011695	0.0031977		$m = 50$	0.0015498	0.0025887

Table 5.6 Median of Mean Squared Error of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Asymmetric Logistic copula $(0.35, 0.75)$ and $(0.75, 0.35)$.

		MSE				MSE	
$(0.35, 0.75)$		(T_i, Q_i)	(T_i, Z_i)	$(0.75, 0.35)$		(T_i, Q_i)	(T_i, Z_i)
$\tau = 0.10$	$m = 7$	0.000453	0.003082	$\tau = 0.10$	$m = 7$	0.000521	0.003024
	$m = 15$	0.000638	0.003082		$m = 15$	0.000673	0.003024
	$m = 25$	0.000863	0.003082		$m = 25$	0.000890	0.003024
	$m = 50$	0.001191	0.003082		$m = 50$	0.001194	0.003024
$\tau = 0.15$	$m = 7$	0.0007835	0.003160	$\tau = 0.15$	$m = 7$	0.000783	0.003064
	$m = 15$	0.000676	0.003160		$m = 15$	0.000708	0.003064
	$m = 25$	0.000883	0.003160		$m = 25$	0.000854	0.003064
	$m = 50$	0.001119	0.003160		$m = 50$	0.001117	0.003064
$\tau = 0.25$	$m = 7$	0.002129	0.003064	$\tau = 0.25$	$m = 7$	0.002421	0.003036
	$m = 15$	0.001053	0.003064		$m = 15$	0.001149	0.003036
	$m = 25$	0.001055	0.003064		$m = 25$	0.001073	0.003036
	$m = 50$	0.001101	0.003064		$m = 50$	0.001074	0.003036

Table 5.7 Median of Mean Squared Error of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Asymmetric Negative Logistic copula $(0.35, 0.75)$ and $(0.75, 0.35)$.

		MSE				MSE	
$(0.35, 0.75)$		(T_i, Q_i)	(T_i, Z_i)	$(0.75, 0.35)$		(T_i, Q_i)	(T_i, Z_i)
$\tau = 0.10$	$m = 7$	0.000519	0.003027	$\tau = 0.10$	$m = 7$	0.000515	0.002999
	$m = 15$	0.000693	0.003027		$m = 15$	0.000689	0.002999
	$m = 25$	0.000895	0.003027		$m = 25$	0.000877	0.002999
	$m = 50$	0.001198	0.003027		$m = 50$	0.001200	0.002999
$\tau = 0.15$	$m = 7$	0.000821	0.003065	$\tau = 0.15$	$m = 7$	0.000806	0.003052
	$m = 15$	0.000713	0.003065		$m = 15$	0.000699	0.003052
	$m = 25$	0.000889	0.003065		$m = 25$	0.000861	0.003052
	$m = 50$	0.001114	0.003065		$m = 50$	0.001084	0.003052
$\tau = 0.25$	$m = 7$	0.002436	0.003055	$\tau = 0.25$	$m = 7$	0.002468	0.003059
	$m = 15$	0.001154	0.003055		$m = 15$	0.001113	0.003059
	$m = 25$	0.001105	0.003055		$m = 25$	0.001085	0.003059
	$m = 50$	0.001076	0.003055		$m = 50$	0.001077	0.003059

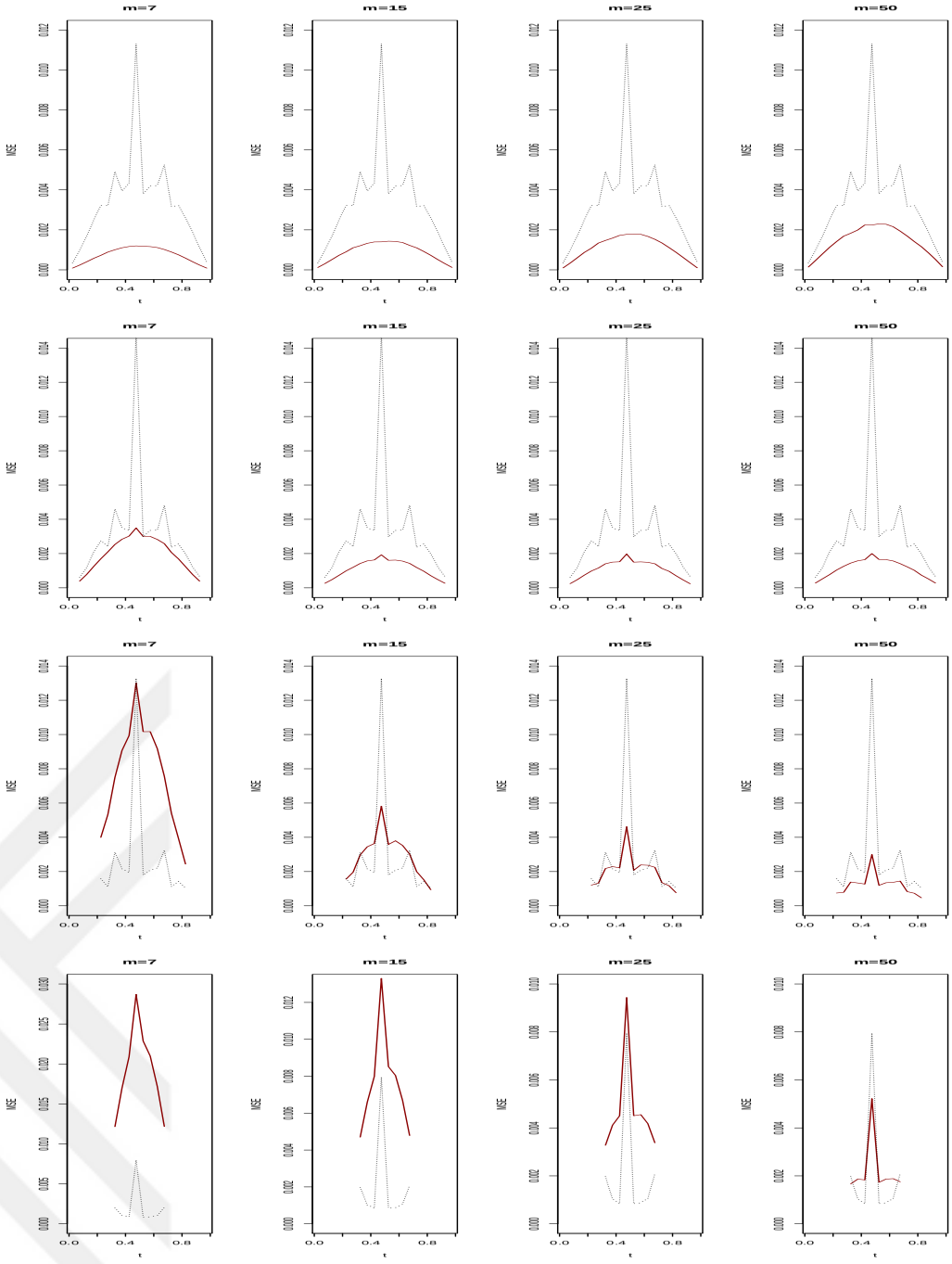


Figure 5.10 MSE of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Logistic copula when $\tau = 0.1$ (top row), $\tau = 0.25$ (second row), $\tau = 0.50$ (third row) and $\tau = 0.75$ (bottom row). Columns are correspond to the order of Bernstein polynomials $m = 7$ (first column), $m = 15$ (second column), $m = 25$ (third column) and $m = 50$ (fourth column).

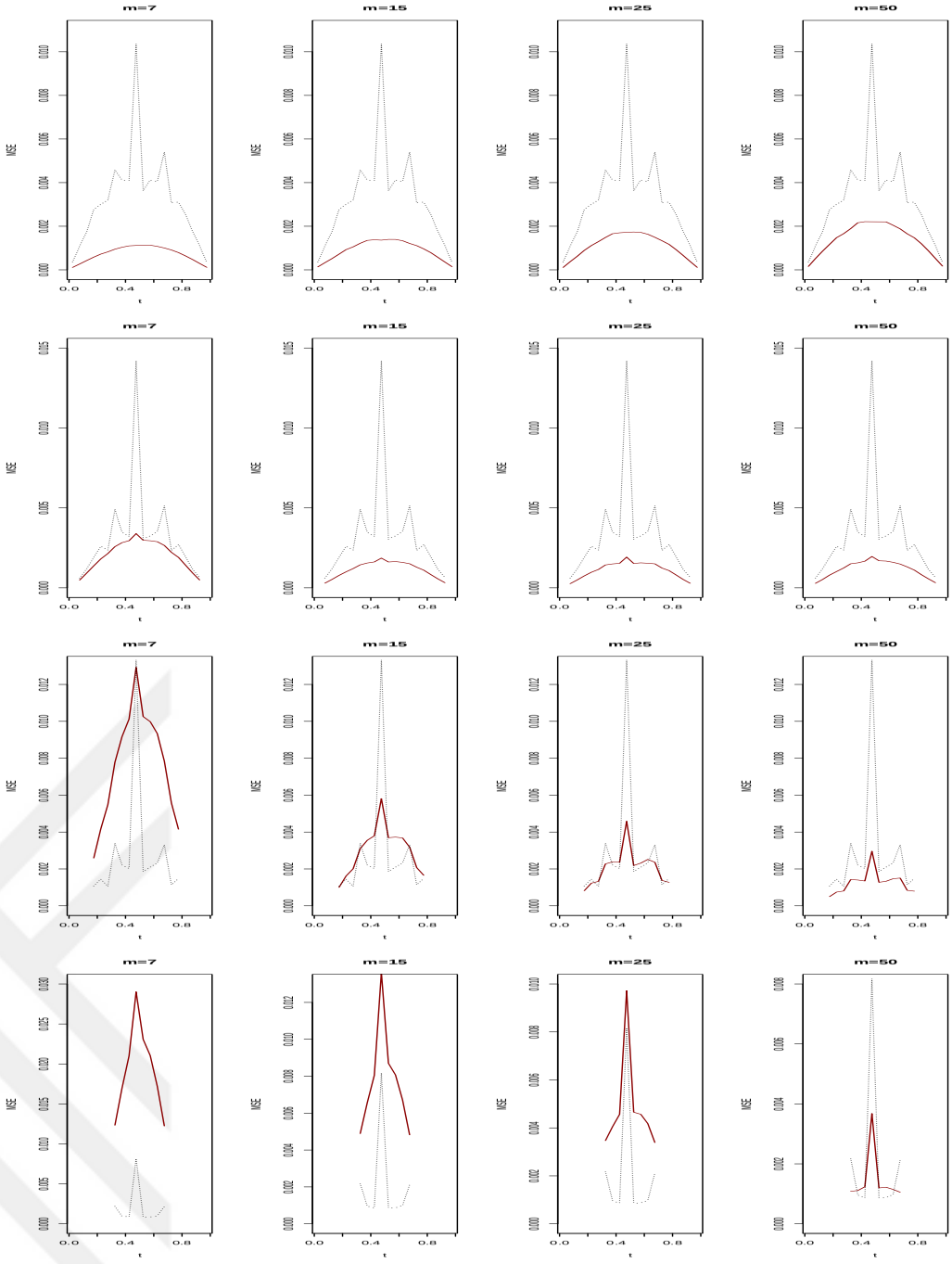


Figure 5.11 MSE of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Negative Logistic copula when $\tau = 0.1$ (top row), $\tau = 0.25$ (second row), $\tau = 0.50$ (third row) and $\tau = 0.75$ (bottom row). Columns are correspond to the order of Bernstein polynomials $m = 7$ (first column), $m = 15$ (second column), $m = 25$ (third column) and $m = 50$ (fourth column).

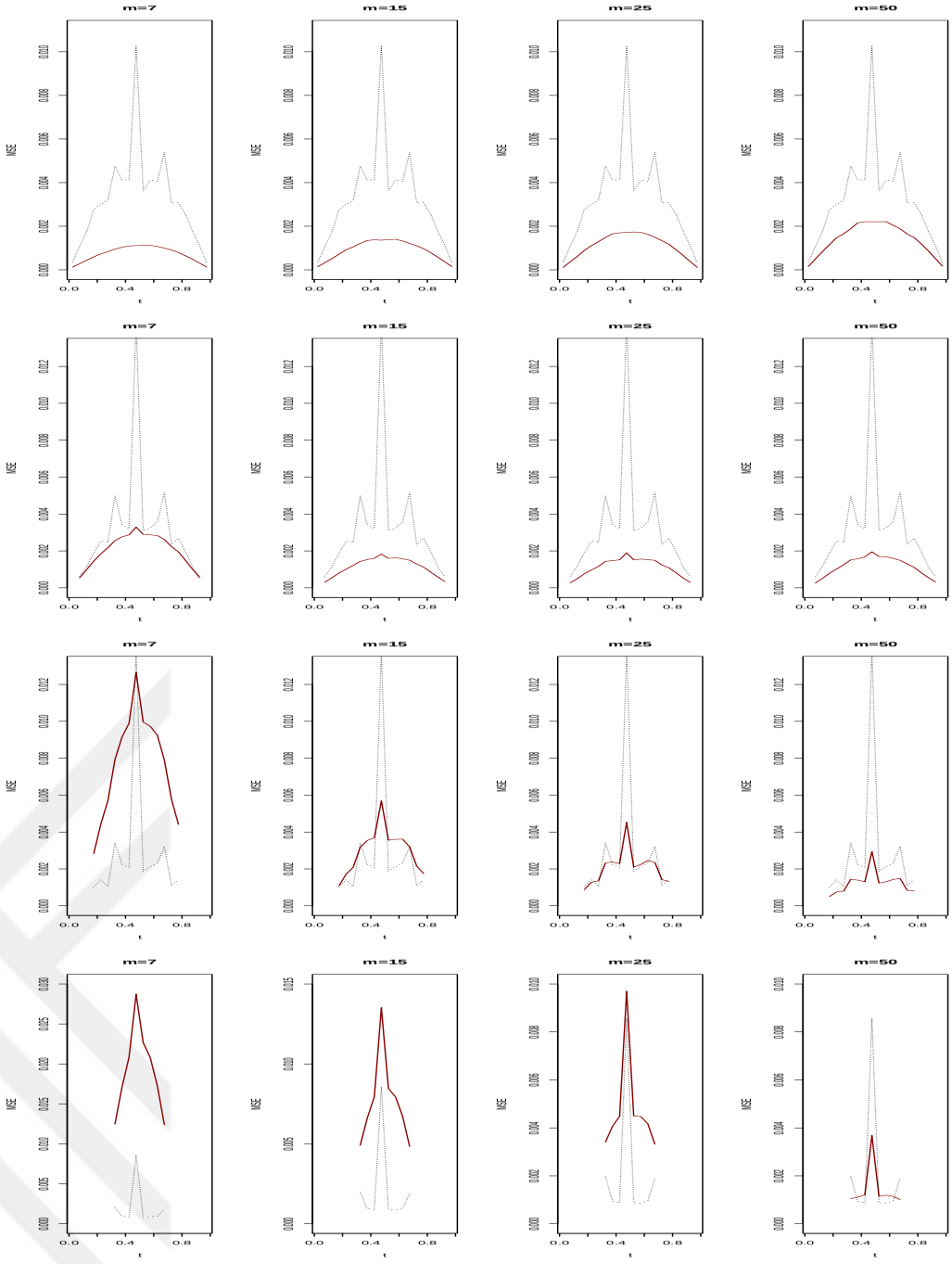


Figure 5.12 MSE of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Hüsler-Reiss copula when $\tau = 0.1$ (top row), $\tau = 0.25$ (second row), $\tau = 0.50$ (third row) and $\tau = 0.75$ (bottom row). Columns are correspond to the order of Bernstein polynomials $m = 7$ (first column), $m = 15$ (second column), $m = 25$ (third column) and $m = 50$ (fourth column).

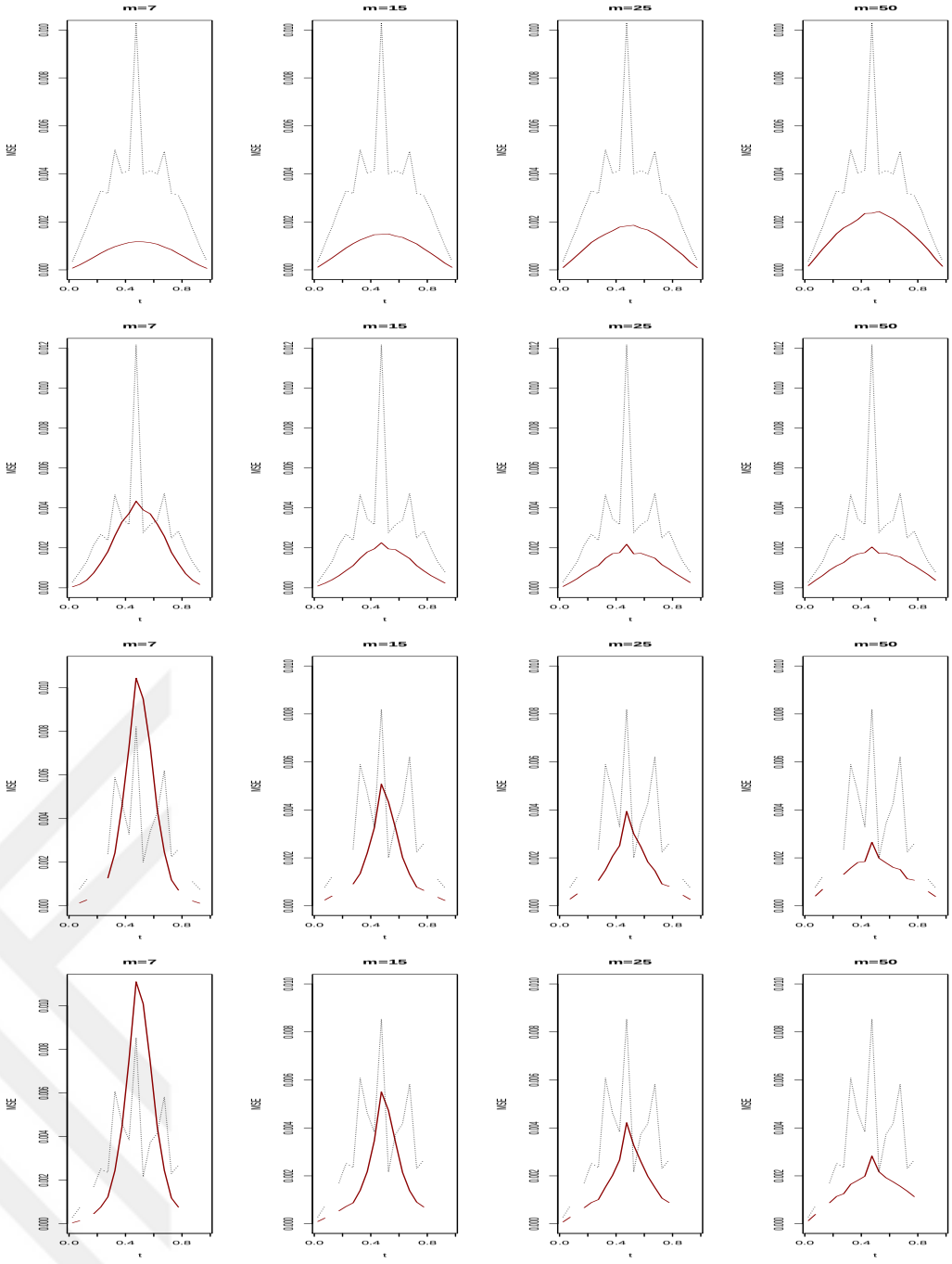


Figure 5.13 MSE of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Asymmetric Logistic copula ($\alpha = 0.50$, $\beta = 0.50$) when $\tau = 0.1$ (top row), $\tau = 0.25$ (second row), $\tau = 0.50$ (third row) and $\tau = 0.75$ (bottom row). Columns are correspond to the order of Bernstein polynomials $m = 7$ (first column), $m = 15$ (second column), $m = 25$ (third column) and $m = 50$ (fourth column).

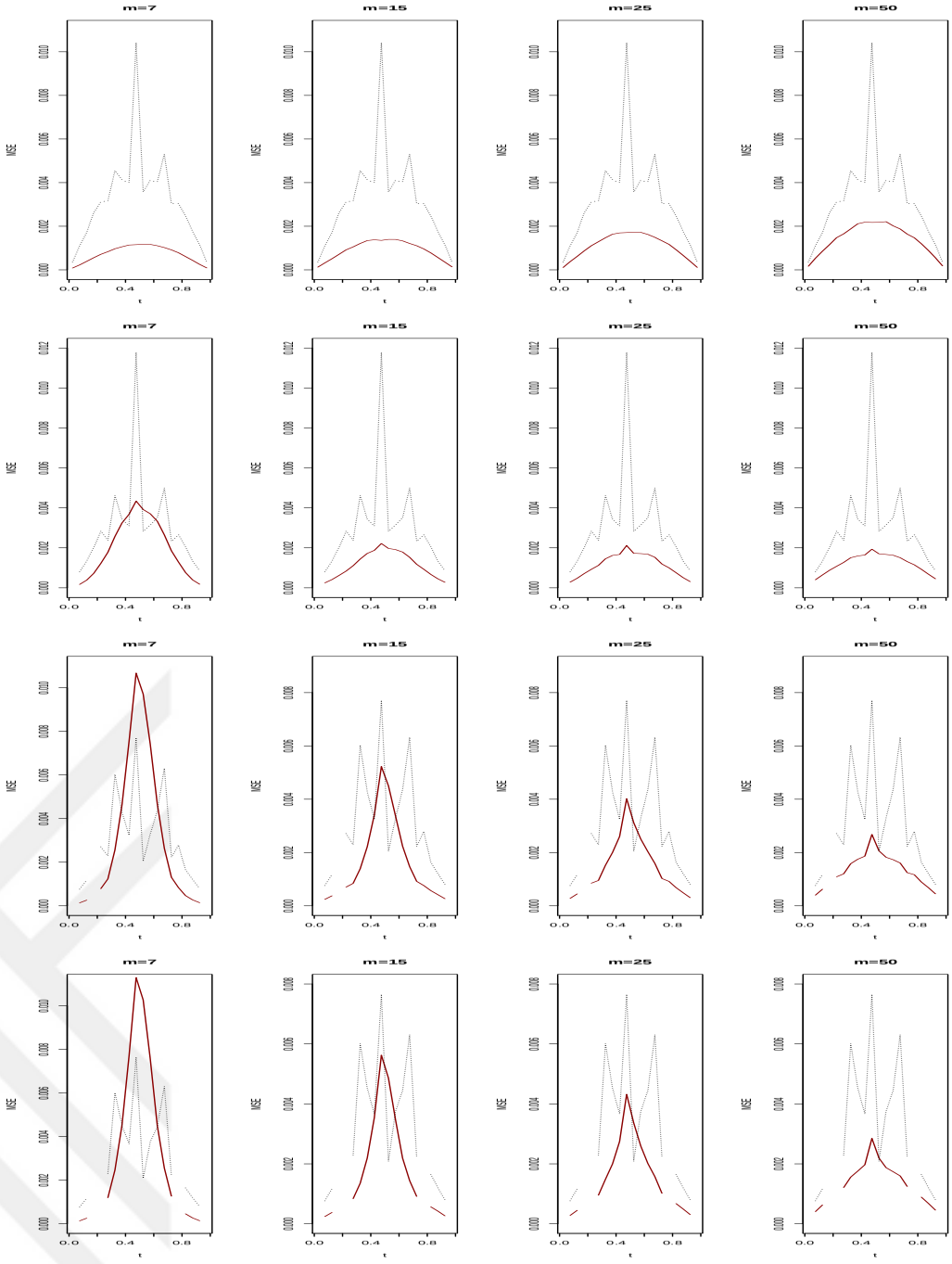


Figure 5.14 MSE of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Asymmetric Negative Logistic copula ($\alpha = 0.50$, $\beta = 0.50$) when $\tau = 0.1$ (top row), $\tau = 0.25$ (second row), $\tau = 0.50$ (third row) and $\tau = 0.75$ (bottom row). Columns are correspond to the order of Bernstein polynomials $m = 7$ (first column), $m = 15$ (second column), $m = 25$ (third column) and $m = 50$ (fourth column).

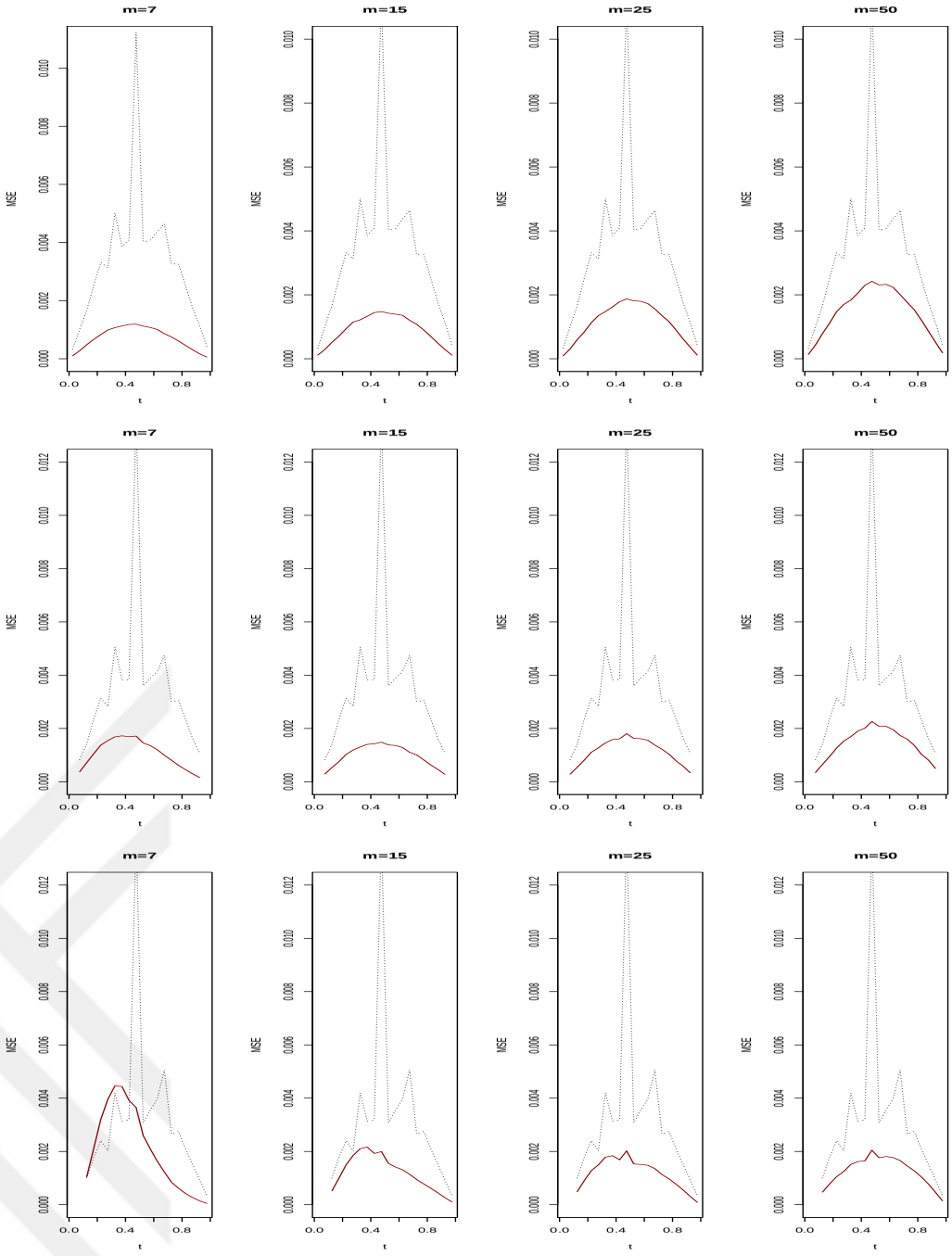


Figure 5.15 MSE of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Asymmetric Logistic copula ($\alpha = 0.35$, $\beta = 0.75$) when $\tau = 0.1$ (top row), $\tau = 0.15$ (second row), and $\tau = 0.25$ (bottom row). Columns are correspond to the order of Bernstein polynomials $m = 7$ (first column), $m = 15$ (second column), $m = 25$ (third column) and $m = 50$ (fourth column).

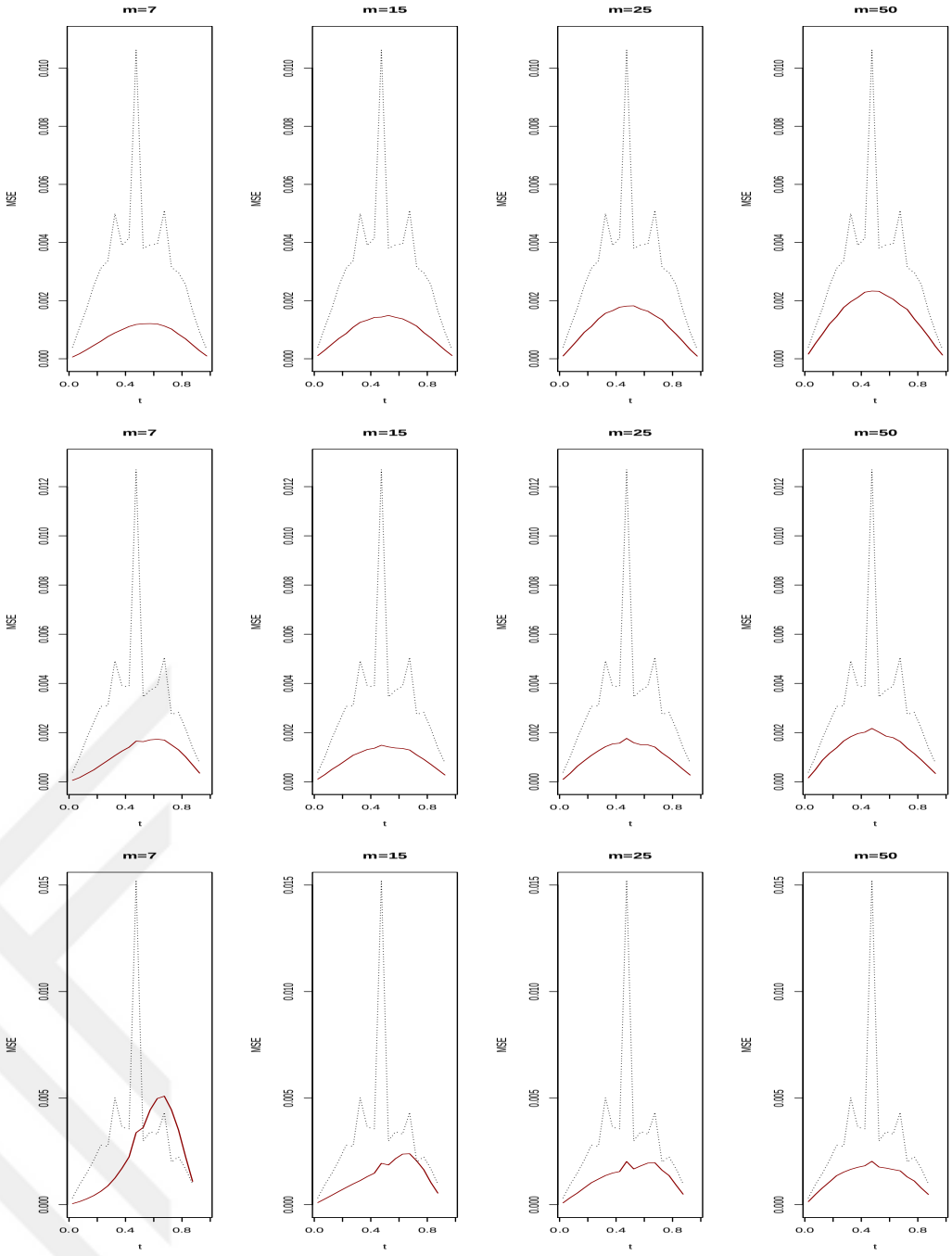


Figure 5.16 MSE of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Asymmetric Logistic copula ($\alpha = 0.75$, $\beta = 0.35$) when $\tau = 0.1$ (top row), $\tau = 0.15$ (second row), and $\tau = 0.25$ (bottom row). Columns are correspond to the order of Bernstein polynomials $m = 7$ (first column), $m = 15$ (second column), $m = 25$ (third column) and $m = 50$ (fourth column).

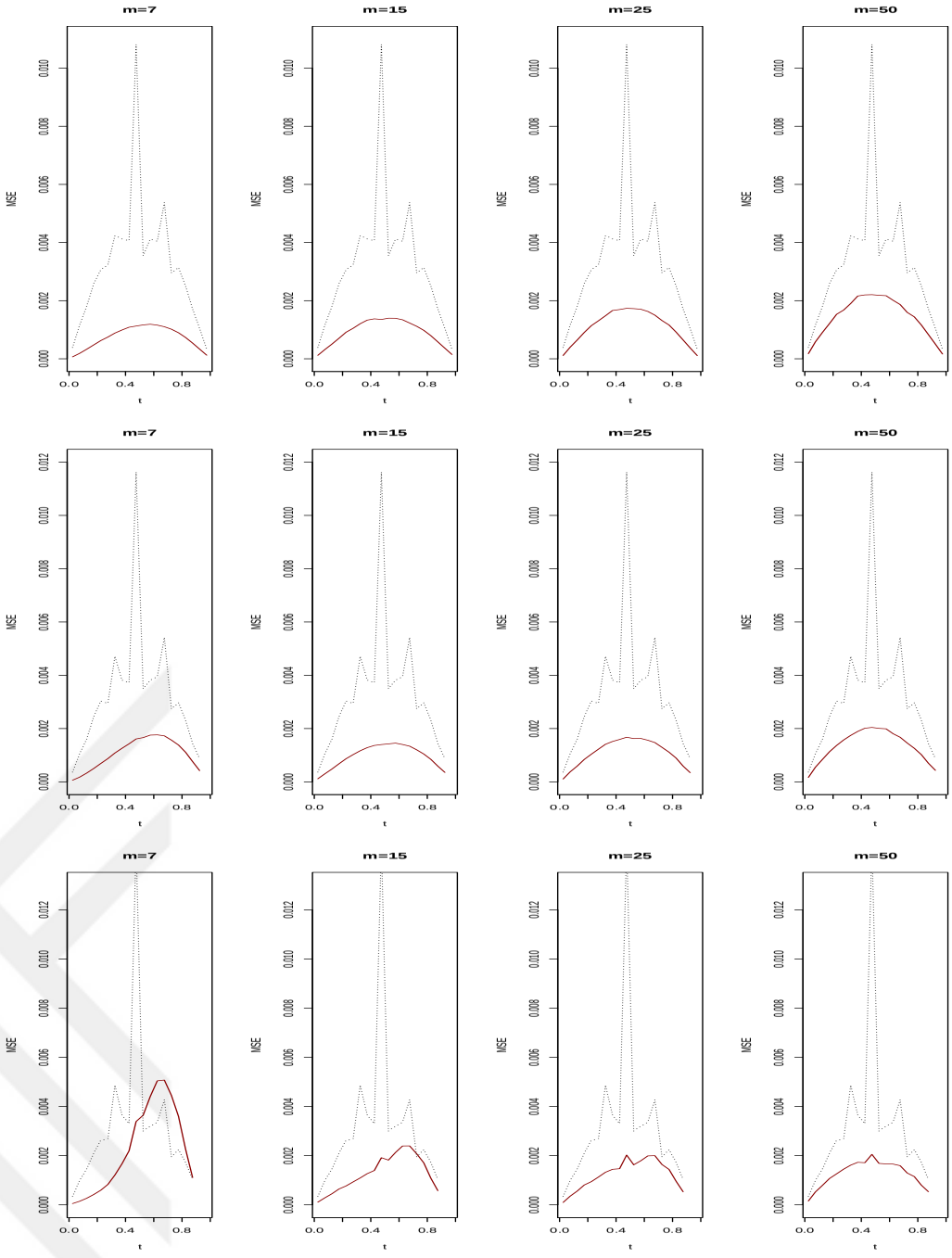


Figure 5.17 MSE of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Asymmetric Negative Logistic copula ($\alpha = 0.35$, $\beta = 0.75$) when $\tau = 0.1$ (top row), $\tau = 0.15$ (second row), and $\tau = 0.25$ (bottom row). Columns are correspond to the order of Bernstein polynomials $m = 7$ (first column), $m = 15$ (second column), $m = 25$ (third column) and $m = 50$ (fourth column).

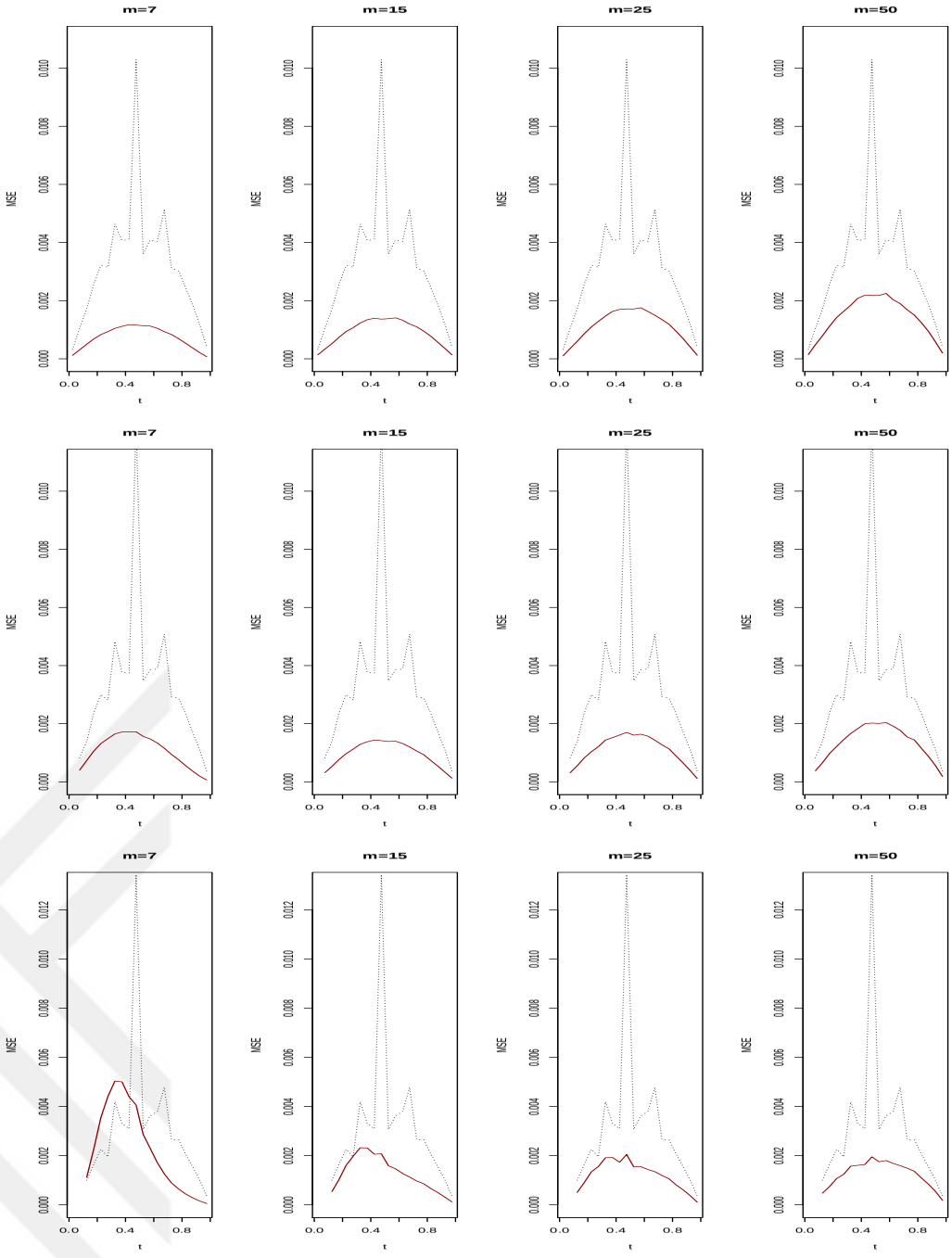


Figure 5.18 MSE of our new pairs, (T_i, Q_i), and the estimated pairs by usual empirical copula, (T_i, Z_i), based on 1000 sample size of 250 from Asymmetric Negative Logistic copula ($\alpha = 0.75$, $\beta = 0.35$) when $\tau = 0.1$ (top row), $\tau = 0.15$ (second row), and $\tau = 0.25$ (bottom row). Columns are correspond to the order of Bernstein polynomials $m = 7$ (first column), $m = 15$ (second column), $m = 25$ (third column) and $m = 50$ (fourth column).

Table 5.8 The probability of success $M^* < M$ in thousands of samples for Logistic, Asymmetric Logistic $(0.5, 0.5)$, Negative Logistic, Asymmetric Negative Logistic $(0.5, 0.5)$ and Hüsler-Reiss distributions.

		Logistic	Asy.Logistic	Neg.Logistic	Asy.Neg.Log.	Hüsler-Reiss
$\tau = 0.10$	$m = 7$	0.983	0.992	0.993	0.992	0.994
	$m = 15$	0.983	0.991	0.993	0.994	0.993
	$m = 25$	0.984	0.980	0.987	0.988	0.989
	$m = 50$	0.961	0.949	0.954	0.956	0.957
$\tau = 0.25$	$m = 7$	0.747	0.681	0.732	0.692	0.725
	$m = 15$	0.944	0.898	0.925	0.893	0.930
	$m = 25$	0.967	0.927	0.955	0.928	0.962
	$m = 50$	0.970	0.922	0.953	0.938	0.958
$\tau = 0.50$	$m = 7$	0.000	0.261	0.000	0.228	0.000
	$m = 15$	0.531	0.665	0.506	0.647	0.514
	$m = 25$	0.764	0.768	0.767	0.750	0.778
	$m = 50$	0.937	0.846	0.932	0.843	0.939
$\tau = 0.75$	$m = 7$	0.000	0.240	0.000	0.203	0.000
	$m = 15$	0.000	0.653	0.000	0.621	0.000
	$m = 25$	0.000	0.754	0.000	0.730	0.000
	$m = 50$	0.555	0.843	0.576	0.826	0.569

5.3 New Estimation of the Pickands Dependence Function: Kernel Regression Approach

Let $\{(x_i, y_i)\}_{i=1}^n$ be n observations. The regression model can be given by

$$y_i = m(x_i) + \varepsilon_i, \quad i = 1, \cdots, n \tag{5.2}$$

where m is an unknown regression function and ε_i are the errors with mean zero and variance σ^2. By reducing the observational errors can be focused on the main details of the mean of the variable Y on X. This method of approximation is called *Smoothing*. Smoothing of $\{(x_i, y_i)\}_{i=1}^n$ involves the approximation of m in the form (5.2). A approximation of $m(x)$ in the form (5.2) can be any delegate point near to the center of the group of y_i's (*Local Average*). i.e.

$$\hat{m}(x) = n^{-1} \sum_{i=1}^{n} \omega_{hi}(x) y_i \tag{5.3}$$

where $\omega_{ni}(x)$ is a sequence of weights depend on $\{x_i\}_{i=1}^n$. Every smoothing method such as for (5.3) is called *Smoother*. For more details, see Härdle (1990). Smoothers are a

Table 5.9 The statistically significant comparison of MSE's in level 0.001 based on 1000 random sample size of 250 from Logistic, Asymmetric Logistic $(0.5, 0.5)$, Negative Logistic, Asymmetric Negative Logistic $(0.5, 0.5)$ and Hüsler-Reiss distributions.

		Logistic	Asy.Logistic	Neg.Logistic	Asy.Neg.Log.	Hüsler-Reiss
$\tau = 0.10$	$m = 7$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 15$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 25$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 50$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
$\tau = 0.25$	$m = 7$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 15$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 25$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 50$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
$\tau = 0.50$	$m = 7$	$M^* > M$	$M^* > M$	$M^* > M$	$M^* > M$	$M^* > M$
	$m = 15$	$M^* > M$	$M^* < M$	$M^* > M$	$M^* < M$	$M^* > M$
	$m = 25$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 50$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
$\tau = 0.75$	$m = 7$	$M^* > M$	$M^* > M$	$M^* > M$	$M^* > M$	$M^* > M$
	$m = 15$	$M^* > M$	$M^* < M$	$M^* > M$	$M^* < M$	$M^* > M$
	$m = 25$	$M^* > M$	$M^* < M$	$M^* > M$	$M^* < M$	$M^* > M$
	$m = 50$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$

broad class of techniques that describe the relationship between two variables. There are different kinds of smoothers. In this study we use Kernel regression. Nadaraya and Watson (1964) proposed the local mean estimator $\hat{m}(x)$ using a kernel as a weighting function. The Nadaraya-Watson estimator is as follows:

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} K(x - x_i; h) y_i}{\sum_{i=1}^{n} K(x - x_i; h)} \tag{5.4}$$

where the kernel weight is

$$\omega_{hi} = \frac{K(x - x_i; h)}{\sum_{i=1}^{n} K(x - x_i; h)}.$$

The kernel function $K(x_i - x; h)$ is a smooth positive function which peaks at zero and decreases monotonically as $(x_i - x; h)$. It is a continuous, symmetric and bounded real function which $\int K(u)du = 1$. The kernel functions that are commonly used include: Gaussian, Epanechnikov, Rectangular, Triangular, Biweight, Cosine, Optcosine, Uniform,

Table 5.10 The probability of success $M^* < M$ in thousands of samples for Asymmetric Logistic $(0.35, 0.75)$, Asymmetric Logistic $(0.75, 0.35)$, Asymmetric Negative Logistic $(0.35, 0.75)$ and Asymmetric Negative Logistic $(0.75, 0.35)$ distributions.

		$A.L._(0.35, 0.75)$	$A.L._(0.75, 0.35)$	$A.N.L._(0.35, 0.75)$	$A.N.L._(0.75, 0.35)$
$\tau = 0.10$	$m = 7$	0.992	0.990	0.990	0.991
	$m = 15$	0.981	0.988	0.994	0.987
	$m = 25$	0.975	0.984	0.984	0.988
	$m = 50$	0.944	0.951	0.955	0.960
$\tau = 0.15$	$m = 7$	0.944	0.951	0.924	0.940
	$m = 15$	0.970	0.979	0.969	0.973
	$m = 25$	0.977	0.980	0.977	0.979
	$m = 50$	0.960	0.967	0.959	0.967
$\tau = 0.25$	$m = 7$	0.656	0.601	0.603	0.587
	$m = 15$	0.890	0.871	0.867	0.876
	$m = 25$	0.932	0.916	0.920	0.916
	$m = 50$	0.934	0.929	0.934	0.934

Table 5.11 The statistically significant comparison of MSE's in level 0.001 based on 1000 random sample size of 250 from Asymmetric Logistic $(0.35, 0.75)$, Asymmetric Logistic $(0.75, 0.35)$, Asymmetric Negative Logistic $(0.35, 0.75)$ and Asymmetric Negative Logistic $(0.75, 0.35)$ distributions.

		$A.L._(0.35, 0.75)$	$A.L._(0.75, 0.35)$	$A.N.L._(0.35, 0.75)$	$A.N.L._(0.75, 0.35)$
$\tau = 0.10$	$m = 7$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 15$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 25$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 50$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
$\tau = 0.15$	$m = 7$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 15$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 25$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 50$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
$\tau = 0.25$	$m = 7$	$M^* > M$	$M^* > M$	$M^* > M$	$M^* > M$
	$m = 15$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 25$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$
	$m = 50$	$M^* < M$	$M^* < M$	$M^* < M$	$M^* < M$

and others. The popular choice of kernel function is the Gaussian kernel.

$$K(x-x_i;h) = (2\pi)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\frac{x-x_i}{h})^2\}$$

The smoothing parameter h (or bandwidth) controls the width of the kernel function. In fact, the amount of smoothing depends on the kernel bandwidth and varying the bandwidth of the kernel estimator controls the smoothness of the estimated regression function: Larger bandwidths produce smoother results. There are two common approaches for choosing an appropriate bandwidth: Cross-Validation and Normal reference rule.

Proposition 5.1 (Consistency)**.** Consider the regression model $y_i = m(X_i) + \varepsilon_i$ for $i = 1, \cdots, n$, and let $\hat{m}(x)$ be the Nadaraya-Watson kernel estimator and ω_{hi} be the kernel weights. Assume

1. $\int |K(u)|du < \infty$,
2. $\lim_{|u|\to\infty} uK(u) = 0$,
3. $EY^2 < \infty$'
4. $n \to \infty$, $h \to 0$ and $nh \to \infty$.

Then, at every point of continuity of $m(x)$, $f(x)$ and $\sigma^2(x)$, with $f(x) > 0$,

$$n^{-1}\sum_{i=1}^{n} \omega_{hi}(x) y_i \xrightarrow{P} m(x),$$

that is, the kernel estimator, $\hat{m}(x)$, converges in probability to the true curve $m(x)$.

Proof. By definition of Nadaraya-Watson kernel estimator,

$$\hat{m}(x) = \frac{n^{-1}\sum_{i=1}^{n} K(x-x_i;h) y_i}{n^{-1}\sum_{i=1}^{n} K(x-x_i;h)} = \frac{\hat{r}_h(x)}{\hat{f}_h(x)},$$

it can be shown that

$$\hat{r}_h(x) = n^{-1}\sum_{i=1}^{n} K(x-x_i;h) Y_i \xrightarrow{P} m(x) f(x) = r(x) \tag{5.5}$$

and

$$\hat{f}_h(x) = n^{-1} \sum_{i=1}^{n} K(x - x_i; h) \xrightarrow{P} f(x). \tag{5.6}$$

Finally, from (5.5) and (5.6) and by using of Slutzky's Theorem,

$$\hat{m}(x) = \frac{\hat{r}_h(x)}{\hat{f}_h(x)} \xrightarrow{P} \frac{m(x)f(x)}{f(x)} = m(x).$$

For more details, see Härdle (1990). □

Note that the estimator $\hat{r}_h(x)$ is asymptotically *unbiased* estimator of $r(x)$. See Härdle (1990).

Definition 5.1. Let $\{(T_i, Q_i)\}_{i=1}^{n}$ be the estimated points of Pickands dependence function using Bernstein copula approach and $K(t - T_i; h)$ be the gaussian kernel function. Then from Nadaraya-Watson estimator, the Pickands dependence function estimator $\hat{A}(t)$ is defined as,

$$\hat{A}_{BK}(t) = \frac{\sum_{i=1}^{n} K(t - T_i; h) Q_i}{\sum_{i=1}^{n} K(t - T_i; h)}. \tag{5.7}$$

Due to the use of the Bernstein polynomials to approximate the points (T, Q), and using these points to estimate the Pickands dependence function, $A(t)$, the new estimator $\hat{A}_{BK}(t)$ satisfies all the properties of $A(t)$, but is not necessarily convex. Thus after constructing the estimator $\hat{A}_{BK}(t)$, we use *Convhull* function in R which is based on the *QHull* algorithm (Barber et al. 1996) to reconstruct a bounded convex estimator. Note that this algorithm has been used with a slight change so that we consider $A(0) = 1$ as the start point and $A(1) = 1$ as the end point of the algorithm. The algorithm attributed to Barber et al. (1996) is as follows:

```
Create a simplex of d + 1 points
for each facet F
    for each unassigned point p
        if p is above F
            assign p to F's outside set
for each facet F with a non-empty outside set
    select the furthest point p of F's outside set
    initialize the visible set V to F
    for all unvisited neighbors N of facets in V
        if p is above N
            add N to V
    the boundary of V is the set of horizon ridge H
    for each ridge R in H
        create a new facet from R and p
        link the new facet to its neighbors
    for each new facet F'
        for each unassigned point q in a outside set of a facet in V
            if q is above F'
                assign q to F''s outside set
    delete the facets in V.
```

Definition 5.2. Let $\mathscr{D} \subseteq \mathbb{R}^d$ be a bounded closed convex set, and let $\mathscr{A}$ be a set of convex functions defined on $\mathscr{D}$. Assume that $A \in \mathscr{A}$ and $\hat{A}_{BK} \in \mathscr{A}_n$ such that $\mathscr{A}_n$ is a set of convex functions below $\hat{A}_{BK}$ on $\mathscr{T}$, where $\mathscr{T}$ is a subset of $\mathscr{D}$ such that the convex hull of $\mathscr{T}$ is $\mathscr{D}$, and

$$\mathscr{A}_n = \{\eta \in \mathscr{A} : \eta(t) \leq \hat{A}_{BK}(t), \quad for all t \in \mathscr{T}\}$$

then the convex estimator, $\hat{A}_{CBK}$, is defined as,

$$\hat{A}_{CBK} = \sup\{\eta : \eta \in \mathscr{A}_n\}.$$

Because $\mathscr{T}$ contains all vertices of $\mathscr{D}$, we can see that $\hat{A}_{CBK} \in \mathscr{A}$ and also $\hat{A}_{CBK} \in \mathscr{A}_n$. For more details see Aguilera et al. (2011). Furthermore, since $\hat{A}_{CBK}$ is the lower part of the convex hull of set $\{(t, \hat{A}_{BK}(t)) : t \in \mathscr{T}\}$, we can use algorithm *Qhull* (Barber et al. 1996) to obtain convex hull of any finite set of points. In fact, we construct the convex

estimator, $\hat{A}_{CBK}$, by choosing $\delta > 0$ and also $\mathscr{T} \in \mathscr{D}$ so that every t in the bounded convex set $\mathscr{D}$ is the convex combination of points in $\mathscr{T}$ whose distance to t is not more than δ. Suppose that A is a continuous convex function on $\mathscr{D}$ with $||A||_{Lip} = L < \infty$ where

$$||A||_{Lip} = \sup\left[\frac{|A(t_1) - A(t_2)|}{|t_1 - t_2|} : t_1, t_2 \in \mathscr{D},\ and,\ t_1 \neq t_2\right]$$

then

$$-\varepsilon \leq \hat{A}_{CBK}(t) - A(t) \leq \varepsilon + L\delta$$

or

$$||\hat{A}_{CBK} - A|| \leq ||\hat{A}_{BK} - A|| + L\delta$$

that is, $\hat{A}_{CBK}$ holds the properties of $\hat{A}$ which is obtained by the Kernel regression approach, and also the rate of convergence is preserved or even better if δ is small enough.

5.3.1 Performance of the New Estimator

The new estimator $\hat{A}_{CBK}$ is obtained based on new pairs (T, Q) and kernel regression approach. It is compared with the Pickands and CFG estimators. In the figures (5.19)-(5.25), the behavior of the new estimator based on sample size of 300 from five known extreme value distributions: Logistic (or Gumbel), Asymmetric Logistic with $(\alpha, \beta) = (0.35, 0.75)$ and $(\alpha, \beta) = (0.75, 0.35)$, Negative Logistic (or Galambos), Asymmetric Negative Logistic with $(\alpha, \beta) = (0.35, 0.75)$ and $(\alpha, \beta) = (0.75, 0.35)$ and Hüsler-Reiss models for different levels of Kendall's tau coefficients are presented. The curves displayed in figures are true Pickands dependence function A and four estimates $\hat{A}_{BC}$ (black), $\hat{A}_{CBK}$ (green) $\hat{A}_{CFG}$ (blue) and $\hat{A}_{PIC}$ (pink). We can conclude that the proposed estimator $\hat{A}_{CBK}$ is closer to the true Pickands function A for low level of dependence and low level of polynomial degree m.

5.3.2 Simulation Study

A simulation study is carried out to investigate the behavior of the new Pickands estimator and compare its performance with the main competitors. The aim of this analysis is to measure the convenience of the new estimator and the effect of fitting the true models. We generate $\{(X_i, Y_i)\}_{i=1}^{n}$ random vector of size 300 from five known extreme value distributions: Logistic (or Gumbel), Asymmetric Logistic with $(\alpha, \beta) = (0.35, 0.75)$ and $(\alpha, \beta) = (0.75, 0.35)$, Negative Logistic (or Galambos),

Asymmetric Negative Logistic with $(\alpha,\beta) = (0.35, 0.75)$ and $(\alpha,\beta) = (0.75, 0.35)$ and Hüsler-Reiss. Note that we consider three dependence levels as measured by Kendall's correlation coefficient $\tau = 0.10, 0.25, 0.50$ for Symmetric Logistic, Negative Logistic and Hüsler-Reiss and $\tau = 0.10, 0.15, 0.25$ for Asymmetric Logistic and Asymmetric Negative Logistic to generate random samples.

In each case, we draw 1000 samples size of 300. For each sample, we approximate the points of Pickands dependence function using empirical Bernstein copula for different Bernstein polynomial degrees $m = 10, 20, 30, 50$. Then we use these estimated points to create kernel regression function and transform this function to the convex function by using convex hull algorithm. To compare the efficiency of the new method, we compute the MSE's of convex Bernstein copula estimator (CBK) and also two well-known, Pickands (PIC) and Caperaa-Fougéres-Genest (CFG) estimators toward the true Pickands function. The main results of this simulation study are summarized in the following tables and graphs. From Tables (5.12)-(5.18), we can also conclude that the new estimator $\hat{A}_{CBK}$ is more efficient than $\hat{A}_{PIC}$ and $\hat{A}_{CFG}$, especially for low level of dependence and low level of polynomial degree m. Fortunately, if there is high level of dependence, then we can increase the polynomial degree m to have lower variance. The results presented in Tables (5.12)-(5.18) are visually shown in Figures (5.26)-(5.32). We can see that $\hat{A}_{CBK}$ is an efficient estimator. Empirical Bernstein copula is closer to the true Pickands dependence function when the level of dependence and Bernstein polynomial degree are low. We can also obtain better result with increasing Bernstein polynomial degree when the level of dependence is high. So, the new convex Bernstein Pickands estimator provides us a flexible way to have a better fit.

5.3.3 Performance Testing

The significant comparisons between three Pickands estimators are obtained by using MSE. These comparisons are made by using sign test which the observed MSE's are based on the same data. Let $\hat{A}$ and $\hat{A}'$ denote the MSE of any two of the Pickands estimators which are obtained from a random sample from Logistic distribution and Asymmetric Logistic distribution, then we test $H_0 : P(\hat{A} < \hat{A}^{'}) = 0.50$. We can conclude that $\hat{A}$ is preferable to $\hat{A}'$ at level $\alpha = 0.001$ for a one-sided test. When H_0 cannot be rejected with a two-sided test at level $\alpha = 0.001$, we cannot make any choice between the

Table 5.12 Median of Mean Squared Error of $\hat{A}$ from Logistic model for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level.

	Logistic Copula					
	CBK				PIC	CFG
MSE	$m = 10$	$m = 20$	$m = 30$	$m = 50$		
$\tau = 0.10$	0.0002663	0.0002790	0.0002981	0.0003250	0.0006850	0.0014210
$\tau = 0.25$	0.0008806	0.0003285	0.0002360	0.0002324	0.0006033	0.0008375
$\tau = 0.50$	0.0036733	0.0012454	0.0004474	0.0001800	0.0003210	0.0002670

Table 5.13 Median of Mean Squared Error of $\hat{A}$ from Negative Logistic model for convex Bernstein copula estimator, Pickands Estimator and CFG estimator of different values of dependence level.

	Negative Logistic Copula					
	CBK				PIC	CFG
MSE	$m = 10$	$m = 20$	$m = 30$	$m = 50$		
$\tau = 0.10$	0.0002758	0.0002755	0.0002933	0.0003169	0.0006942	0.0014580
$\tau = 0.25$	0.0009402	0.0003909	0.0003016	0.0002656	0.0005819	0.0009364
$\tau = 0.50$	0.0038190	0.0013640	0.0007345	0.0003607	0.0002850	0.0003232

estimators. Based on random 1000 sample of size 300 from logistic and asymmetric logistic distributions, statistically significant comparisons of MSE's are given in Tables (5.19)-(5.21) for three intrinsic estimators of Pickands dependence functions.

Table 5.14 Median of Mean Squared Error of $\hat{A}$ from Hüsler-Reiss model for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level.

	Hüsler-Reiss Copula					
	CBK				PIC	CFG
MSE	$m = 10$	$m = 20$	$m = 30$	$m = 50$		
$\tau = 0.10$	0.0002725	0.0002750	0.0002913	0.0003190	0.0006946	0.0014500
$\tau = 0.25$	0.0009397	0.0003871	0.0002975	0.0002668	0.0005879	0.0009586
$\tau = 0.50$	0.0037731	0.0013207	0.0007105	0.0003523	0.0002858	0.0003303

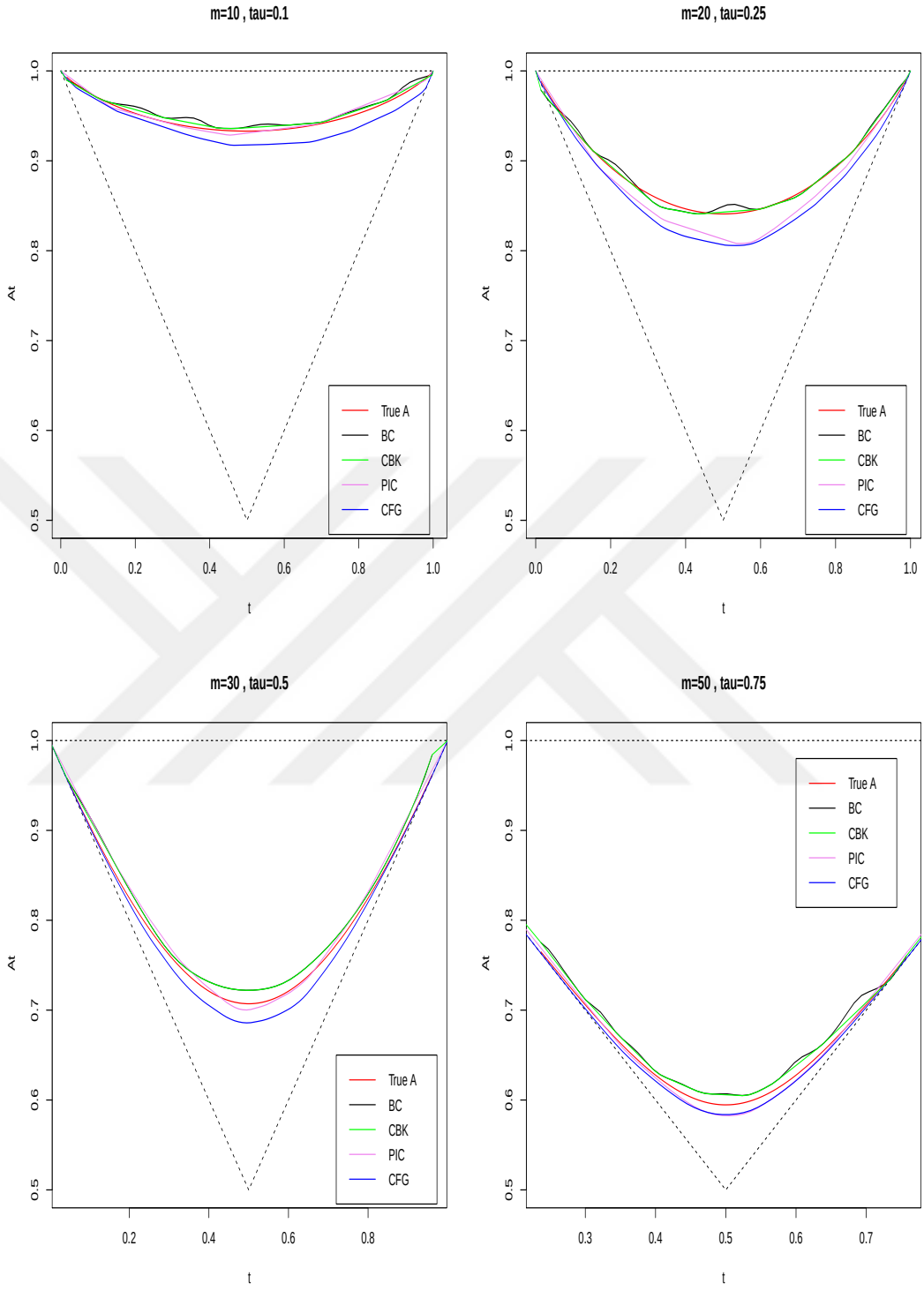


Figure 5.19 Plots based on sample size of 300 from the Logistic copula with $\tau = 0.10$ and $m = 10$ (top left), $\tau = 0.25$ and $m = 20$ (top right), $\tau = 0.50$ and $m = 30$ (bottom left) and $\tau = 0.75$ and $m = 50$ (bottom right).

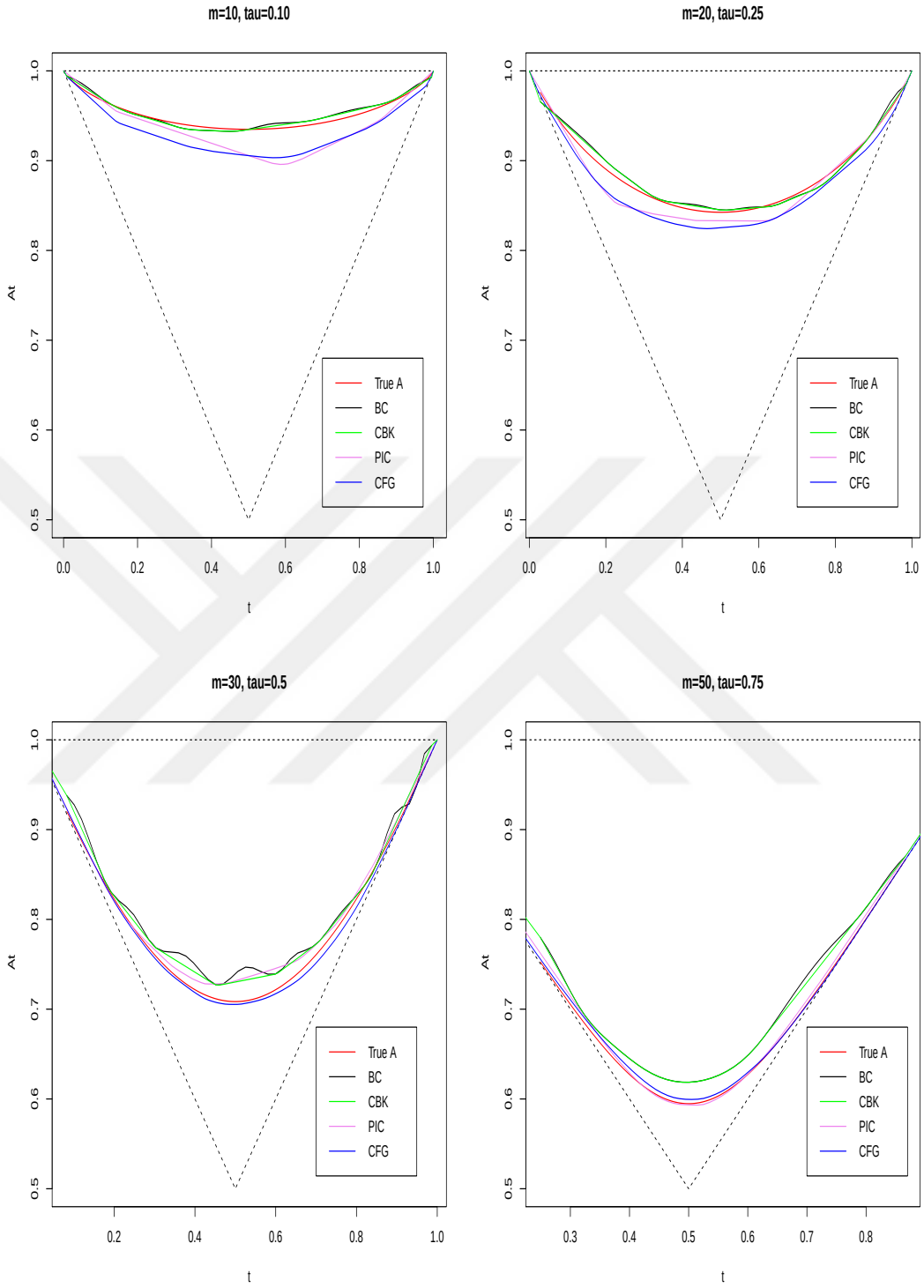


Figure 5.20 Plots based on sample size of 300 from the Negative Logistic copula with $\tau = 0.10$ and $m = 10$ (top left), $\tau = 0.25$ and $m = 20$ (top right), $\tau = 0.50$ and $m = 30$ (bottom left) and $\tau = 0.75$ and $m = 50$ (bottom right).

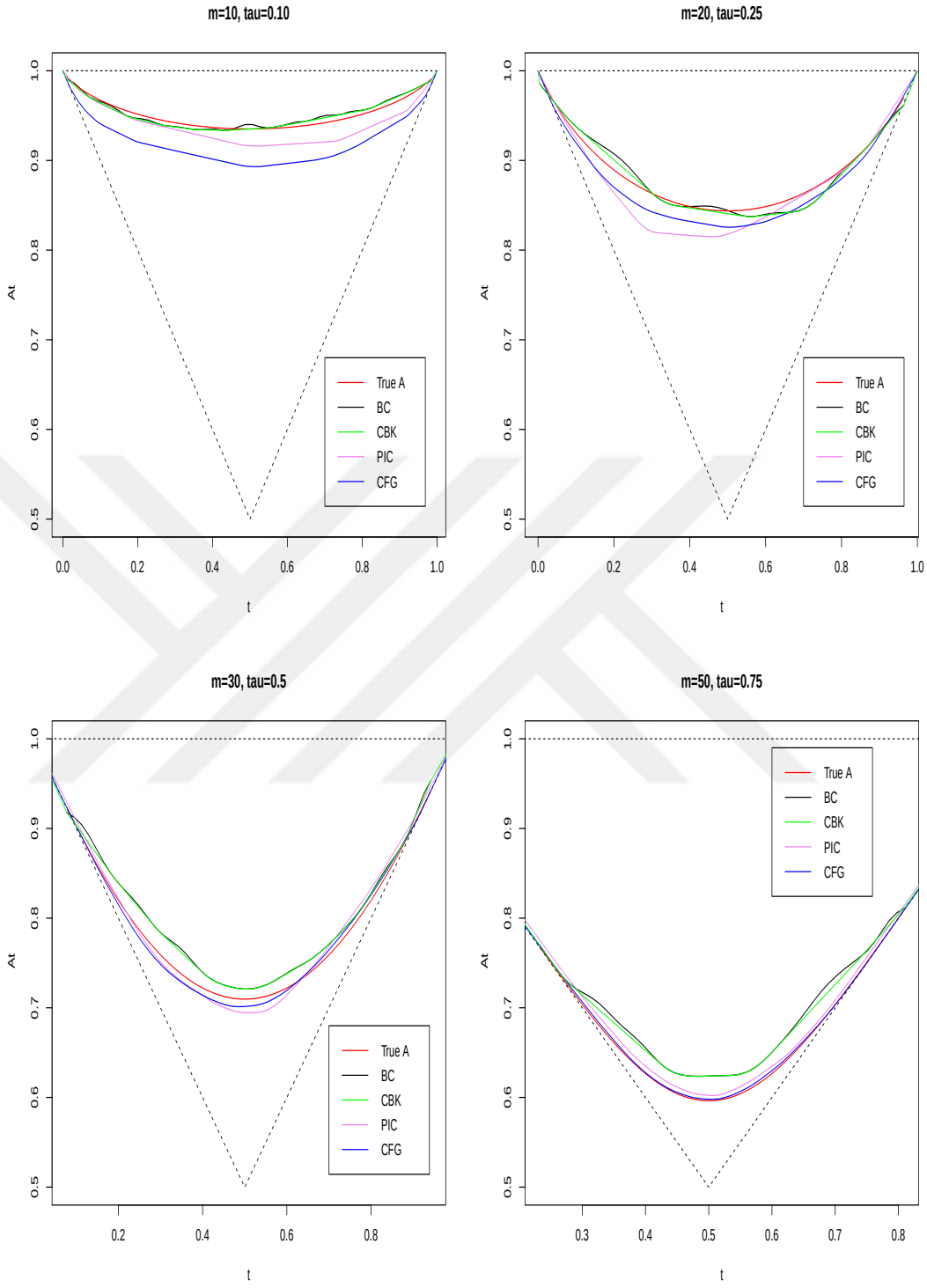


Figure 5.21 Plots based on sample size of 300 from the Hüsler-Reiss copula with $\tau = 0.10$ and $m = 10$ (top left), $\tau = 0.25$ and $m = 20$ (top right), $\tau = 0.50$ and $m = 30$ (bottom left) and $\tau = 0.75$ and $m = 50$ (bottom right).

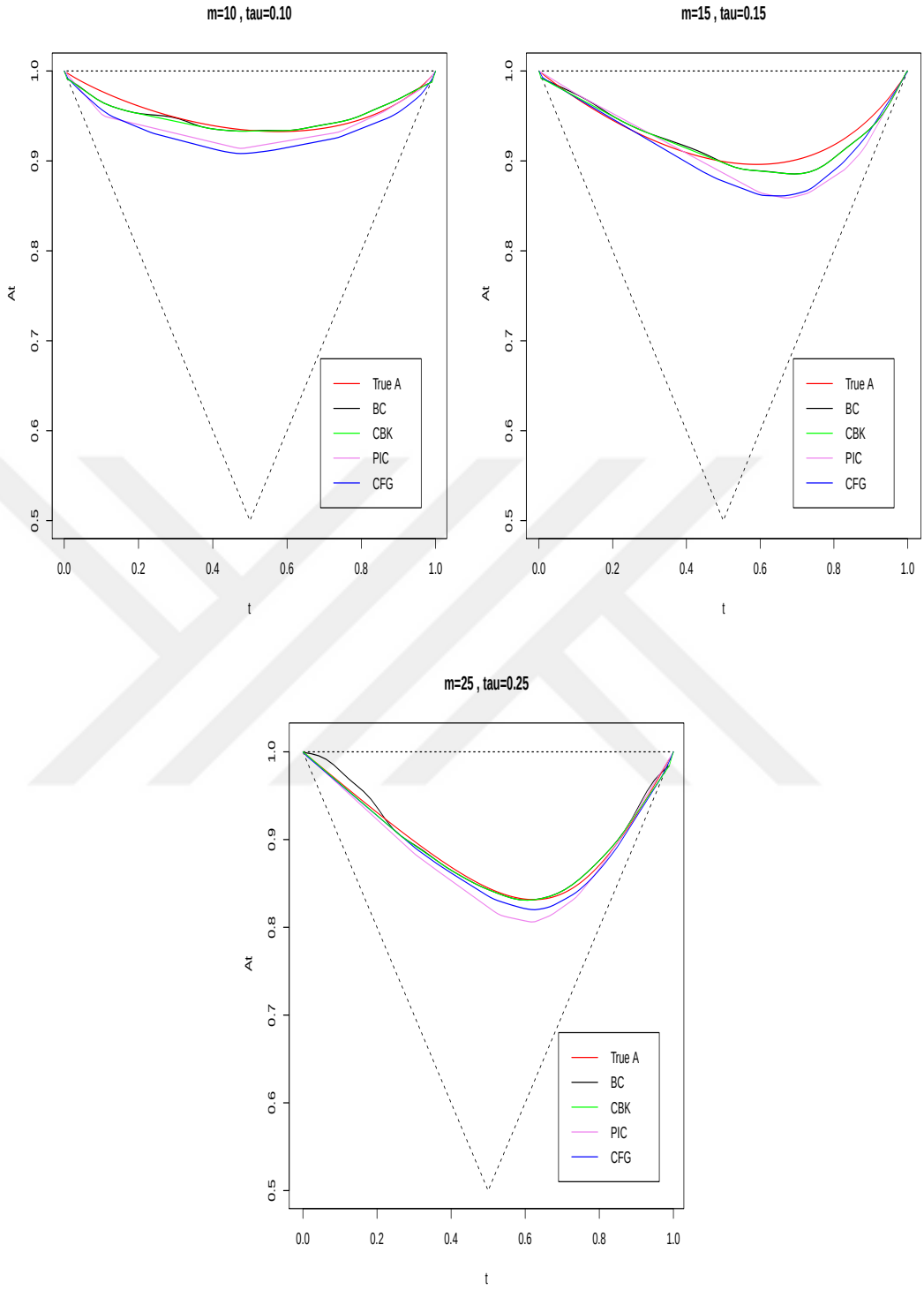


Figure 5.22 Plots based on sample size of 300 from the Asymmetric Logistic copula $(0.35, 0.75)$ with $\tau = 0.10$ and $m = 10$ (top left), $\tau = 0.15$ and $m = 15$ (top right), $\tau = 0.25$ and $m = 25$ (bottom row).

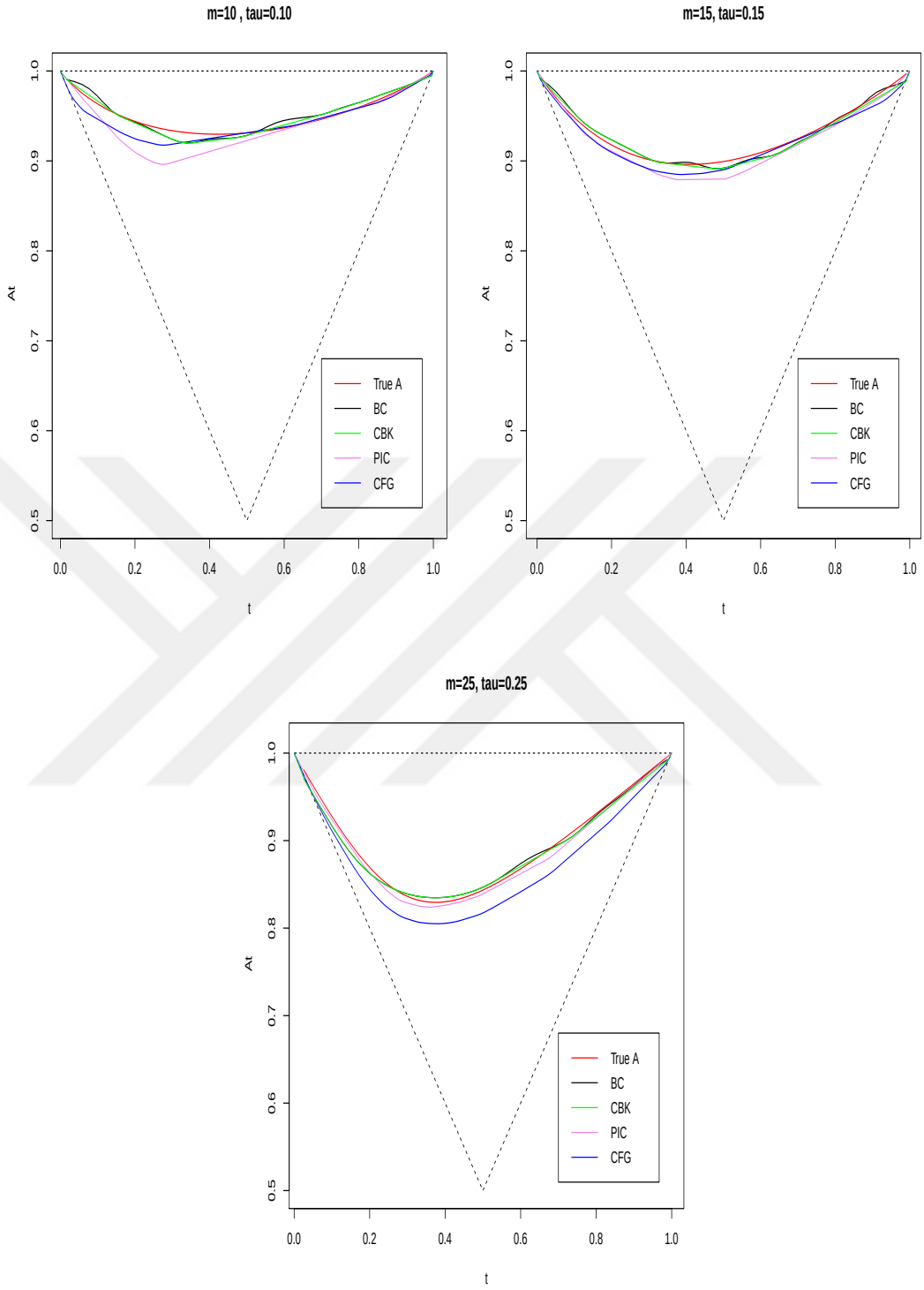


Figure 5.23 Plots based on sample size of 300 from the Asymmetric Logistic copula $(0.75, 0.35)$ with $\tau = 0.10$ and $m = 10$ (top left), $\tau = 0.15$ and $m = 15$ (top right), $\tau = 0.25$ and $m = 25$ (bottom row).

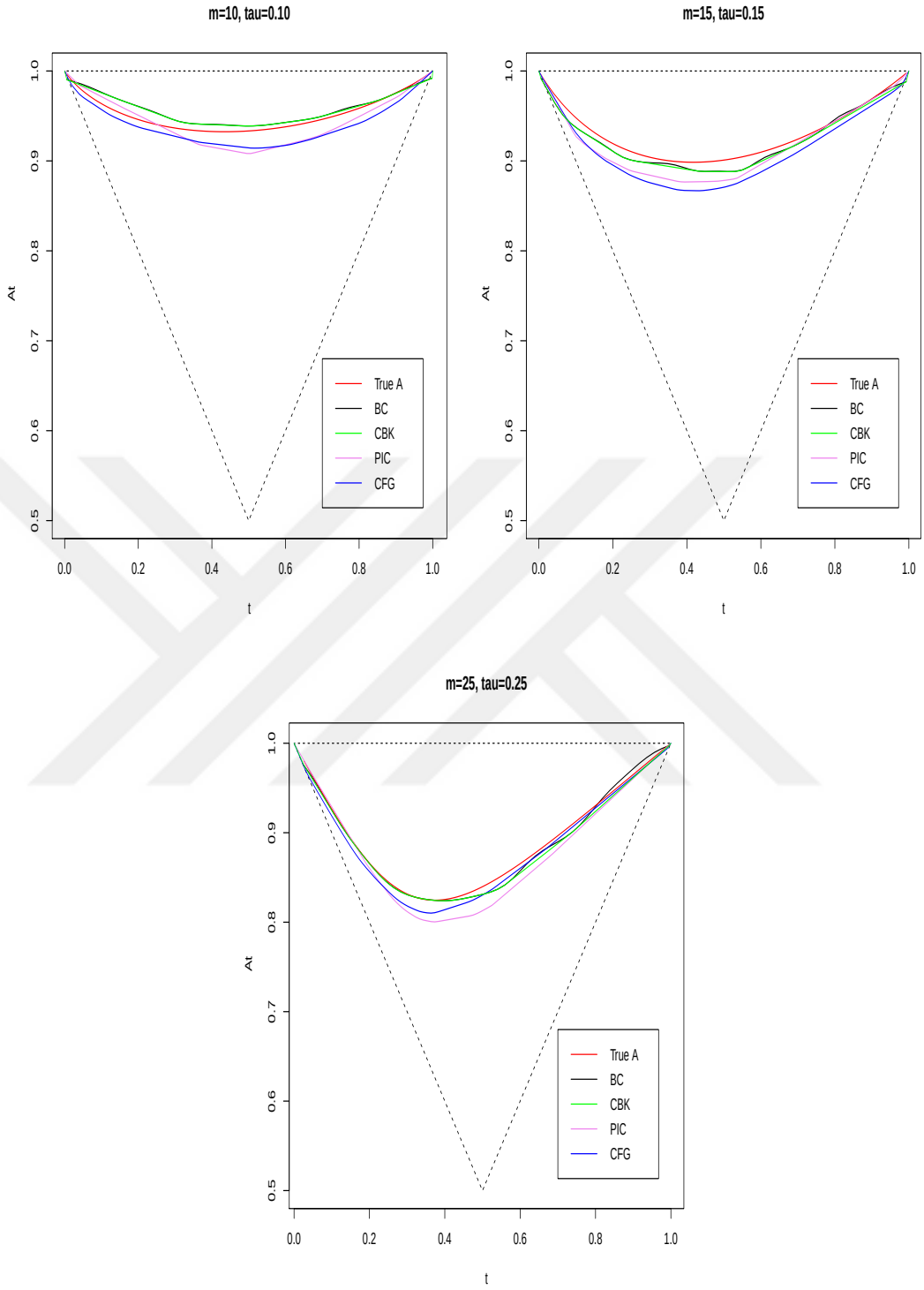


Figure 5.24 Plots based on sample size of 300 from the Asymmetric Negative Logistic copula $(0.35, 0.75)$ with $\tau = 0.10$ and $m = 10$ (top left), $\tau = 0.15$ and $m = 15$ (top right), $\tau = 0.25$ and $m = 25$ (bottom row).

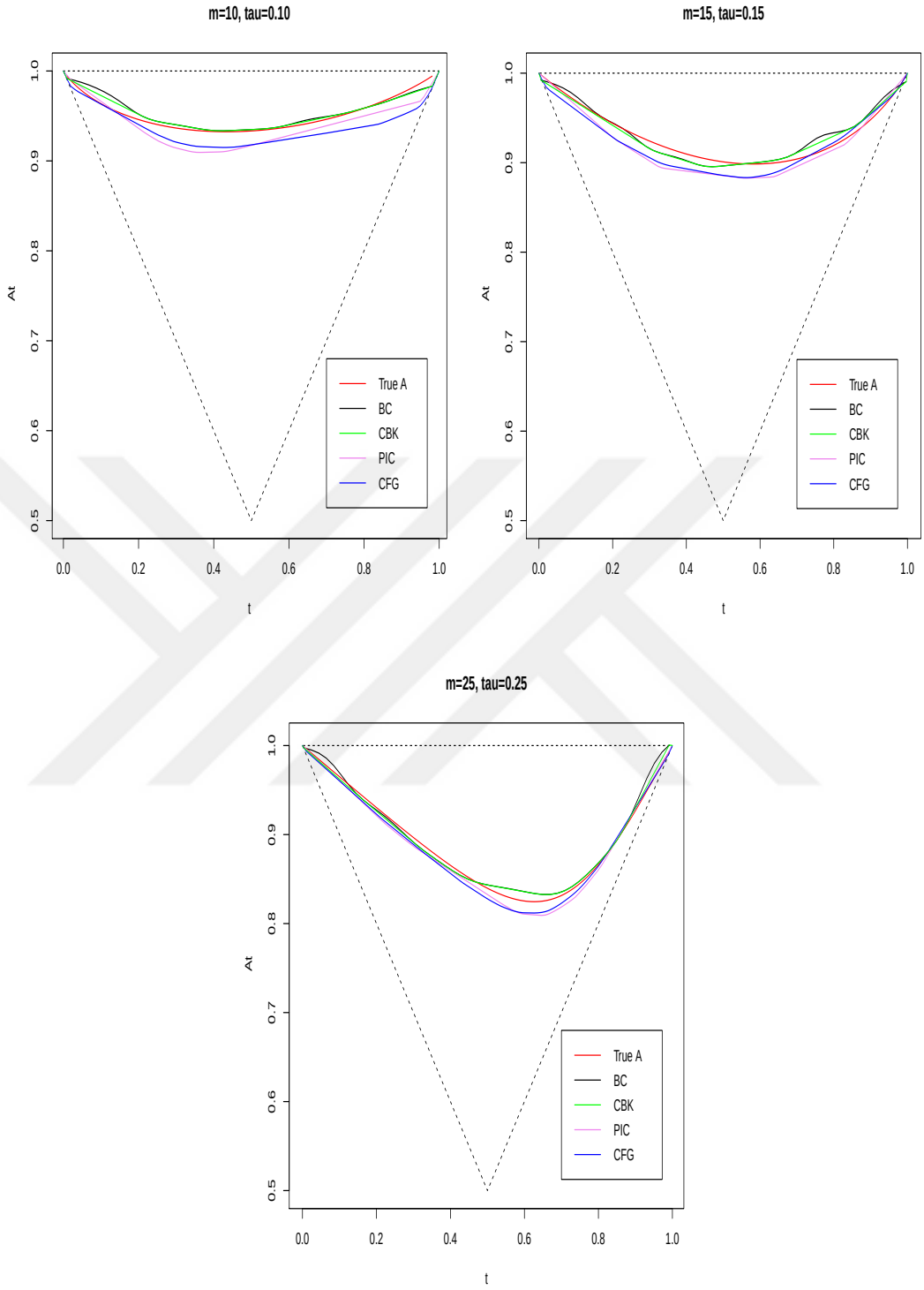


Figure 5.25 Plots based on sample size of 300 from the Asymmetric Negative Logistic copula $(0.75, 0.35)$ with $\tau = 0.10$ and $m = 10$ (top left), $\tau = 0.15$ and $m = 15$ (top right), $\tau = 0.25$ and $m = 25$ (bottom row).

Table 5.15 Median of Mean Squared Error of $\hat{A}$ from Asymmetric Logistic model with parameters $(0.35, 0.75)$ for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level.

	Asymmetric Logistic Copula $(0.35, 0.75)$					
	CBK				PIC	CFG
MSE	$m = 10$	$m = 20$	$m = 30$	$m = 50$		
$\tau = 0.10$	0.0002441	0.0002535	0.0002792	0.0003050	0.0006874	0.0012100
$\tau = 0.15$	0.0003676	0.0002665	0.0002763	0.0002975	0.0007378	0.0010134
$\tau = 0.25$	0.0010711	0.0004524	0.0003448	0.0003012	0.0005870	0.0008475

Table 5.16 Median of Mean Squared Error of $\hat{A}$ from Asymmetric Logistic model with parameters $(0.75, 0.35)$ for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level.

	Asymmetric Logistic Copula $(0.75, 0.35)$					
	CBK				PIC	CFG
MSE	$m = 10$	$m = 20$	$m = 30$	$m = 50$		
$\tau = 0.10$	0.0002513	0.0002621	0.0002814	0.0003011	0.0006708	0.0012820
$\tau = 0.15$	0.0003682	0.0002737	0.0002777	0.0002917	0.0007088	0.0009831
$\tau = 0.25$	0.0011850	0.0004968	0.0003658	0.0003018	0.0005665	0.0008535

Table 5.17 Median of Mean Squared Error of $\hat{A}$ from Asymmetric Negative Logistic model with parameters $(0.35, 0.75)$ for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level.

	Asymmetric Negative Logistic $(0.35, 0.75)$					
	CBK				PIC	CFG
MSE	$m = 10$	$m = 20$	$m = 30$	$m = 50$		
$\tau = 0.10$	0.0002887	0.0002846	0.0002958	0.0003293	0.0006763	0.0013820
$\tau = 0.15$	0.0004376	0.0003023	0.0002950	0.0002957	0.0006712	0.0012450
$\tau = 0.25$	0.0012860	0.0005395	0.0003943	0.0003199	0.0005586	0.0008750

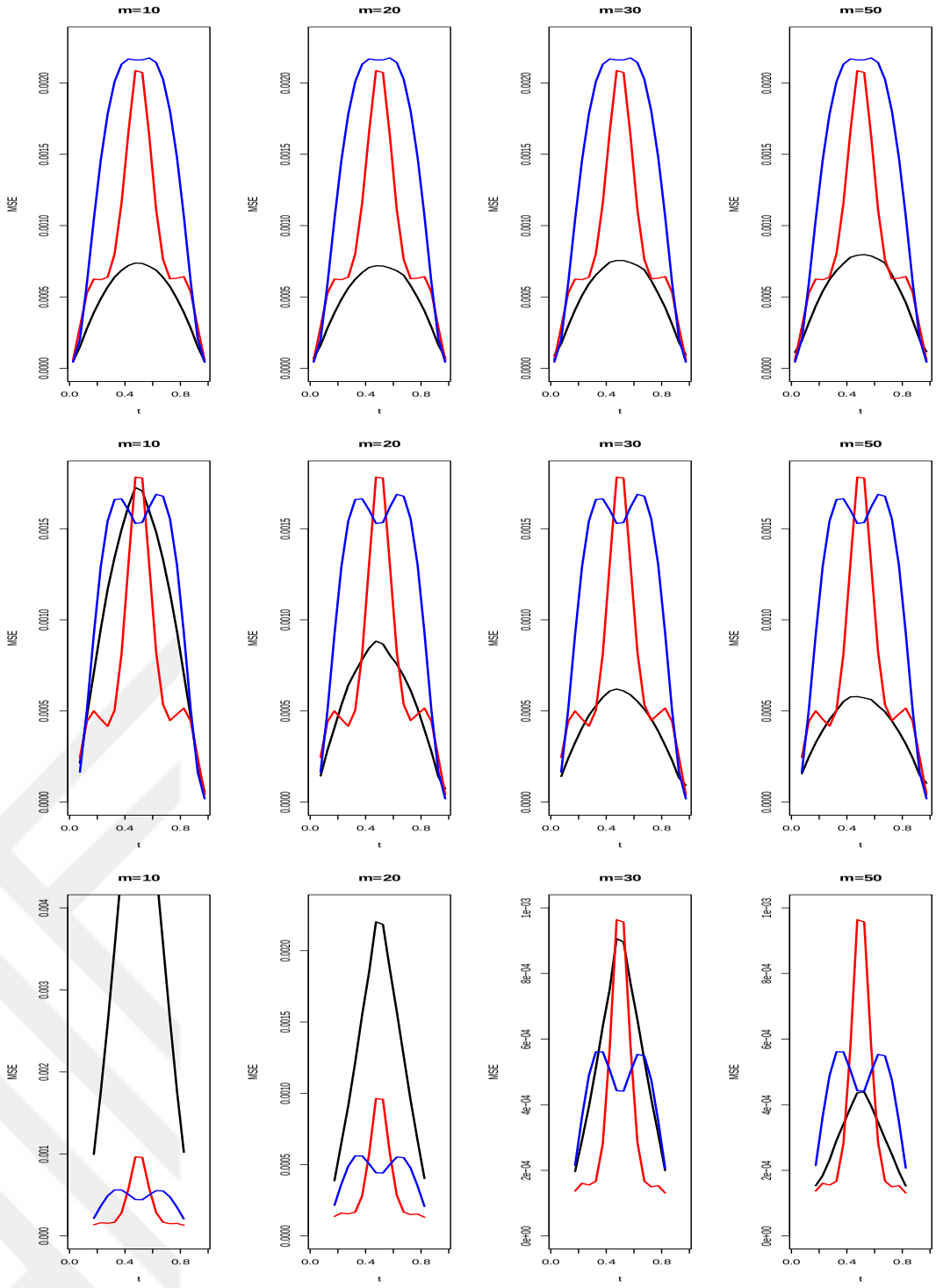


Figure 5.26 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Logistic model when $\tau = 0.1$ (top row), $\tau = 0.25$ (second row), and $\tau = 0.5$ (bottom row).

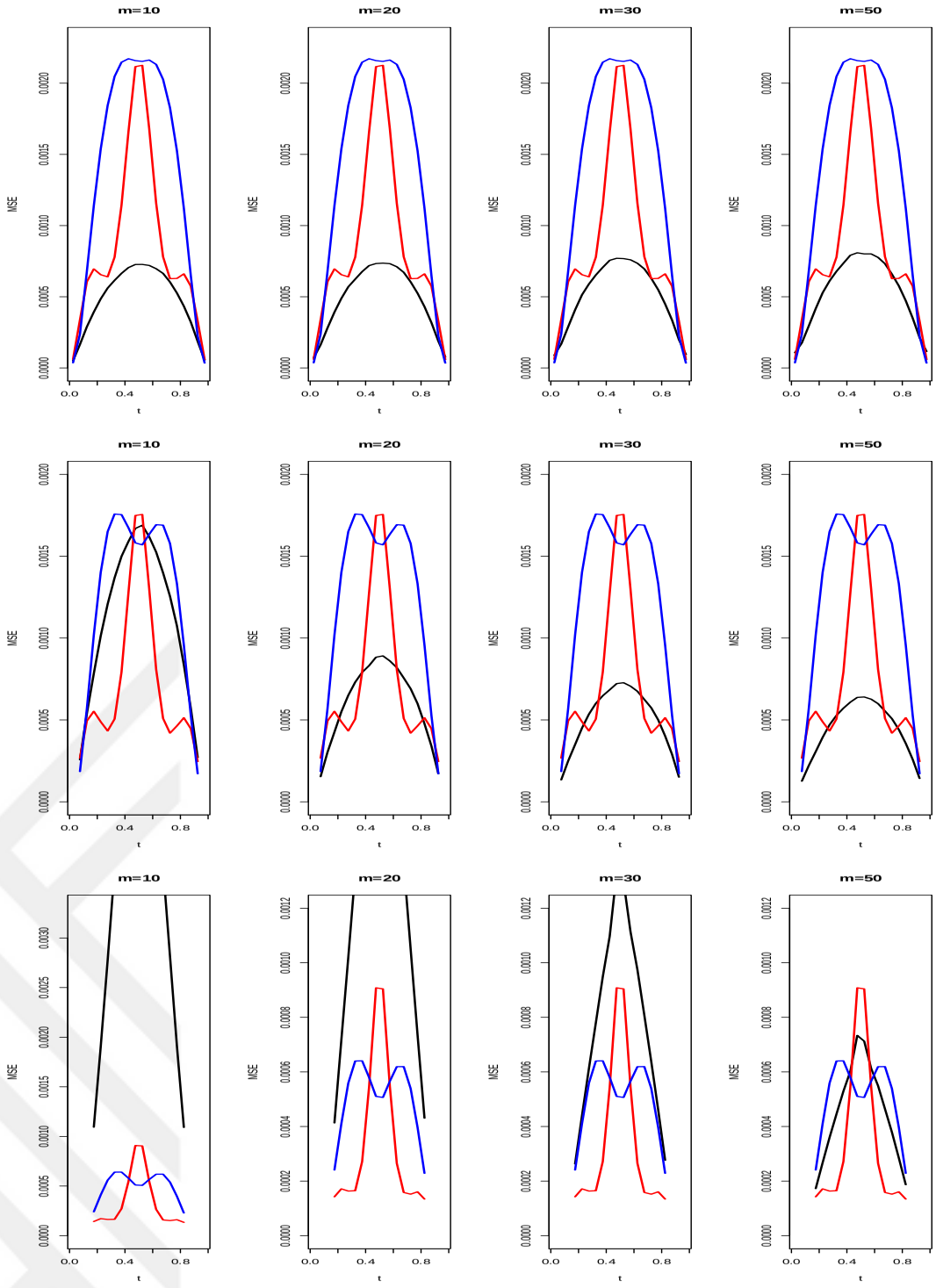


Figure 5.27 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Negative Logistic model when $\tau = 0.1$ (top row), $\tau = 0.25$ (second row), and $\tau = 0.5$ (bottom row).

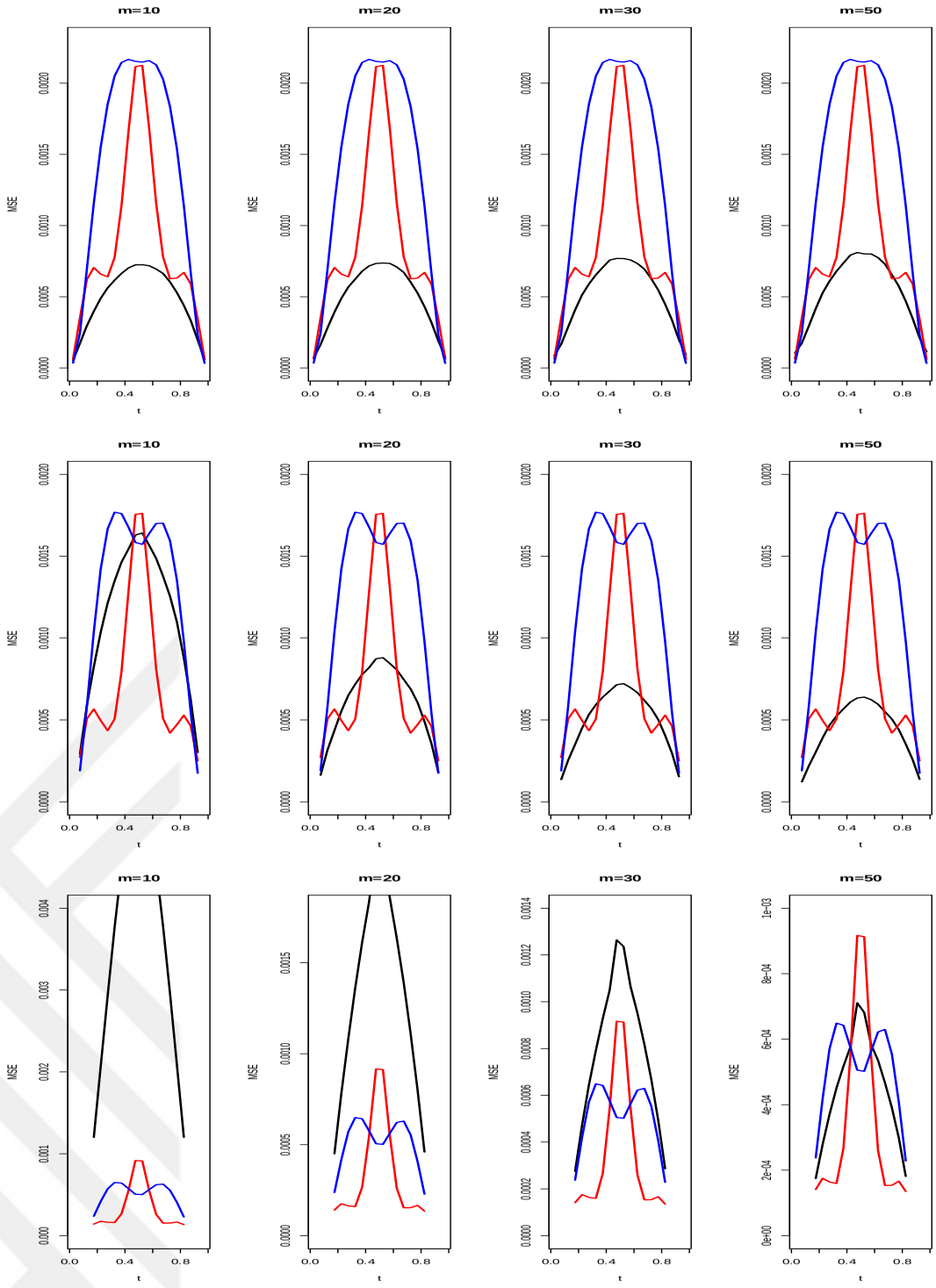


Figure 5.28 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Hüsler-Reiss model when $\tau = 0.1$ (top row), $\tau = 0.25$ (second row), and $\tau = 0.5$ (bottom row).

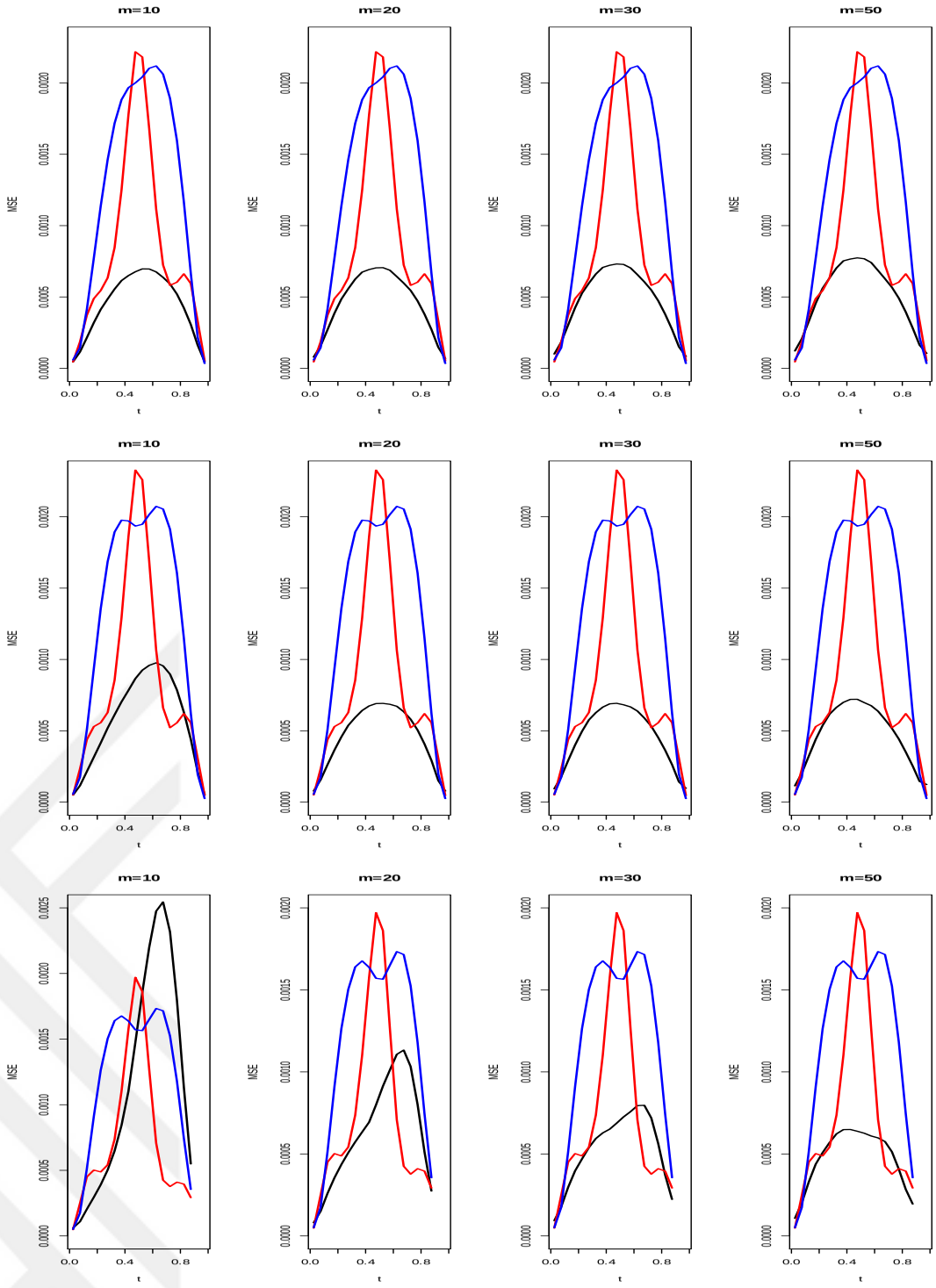


Figure 5.29 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Asymmetric Logistic model ($\alpha = 0.35$, $\beta = 0.75$) when $\tau = 0.1$ (top row), $\tau = 0.15$ (second row), and $\tau = 0.25$ (bottom row).

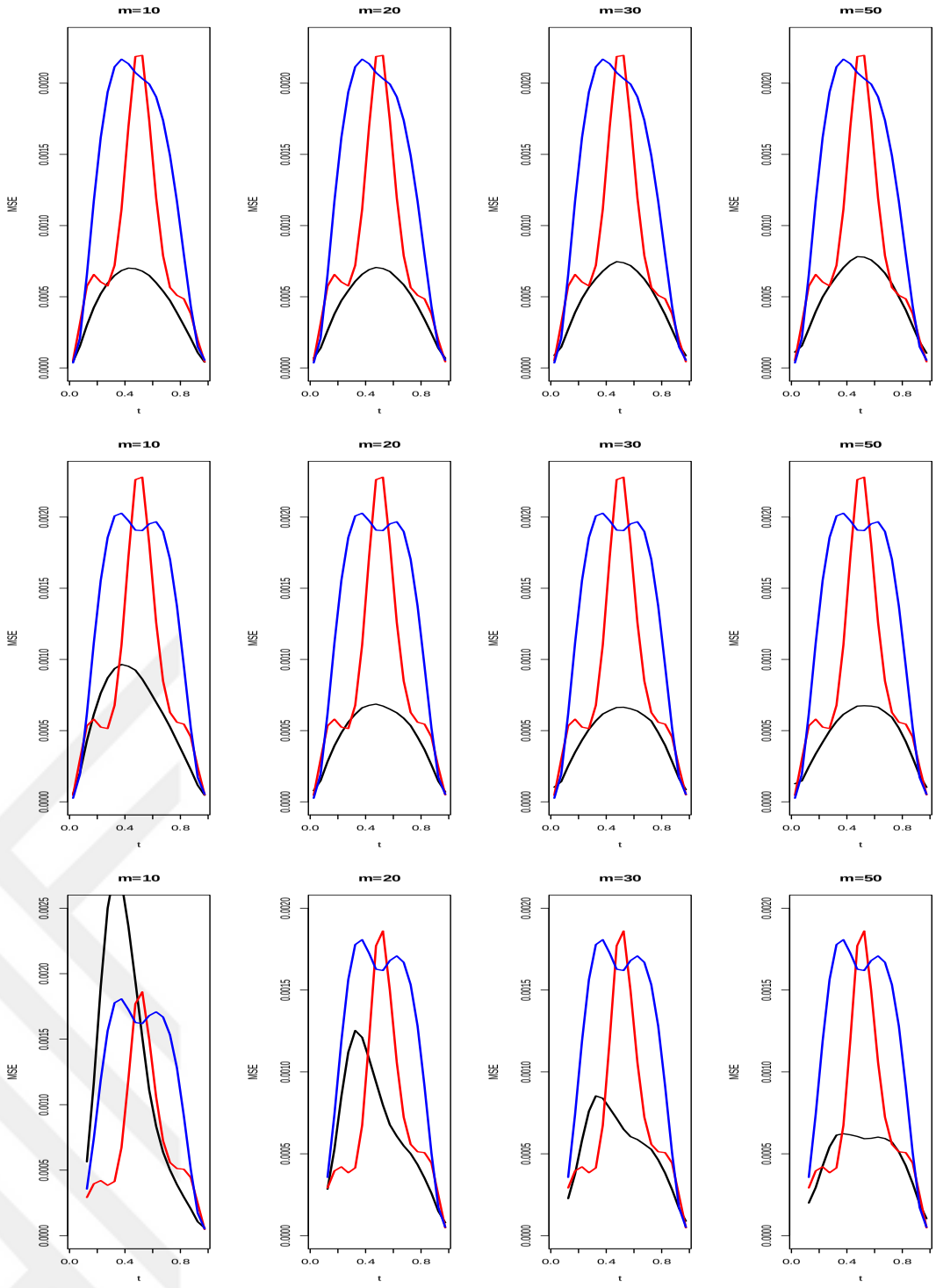


Figure 5.30 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Asymmetric Logistic model ($\alpha = 0.75$, $\beta = 0.35$) when $\tau = 0.1$ (top row), $\tau = 0.15$ (second row), and $\tau = 0.25$ (bottom row).

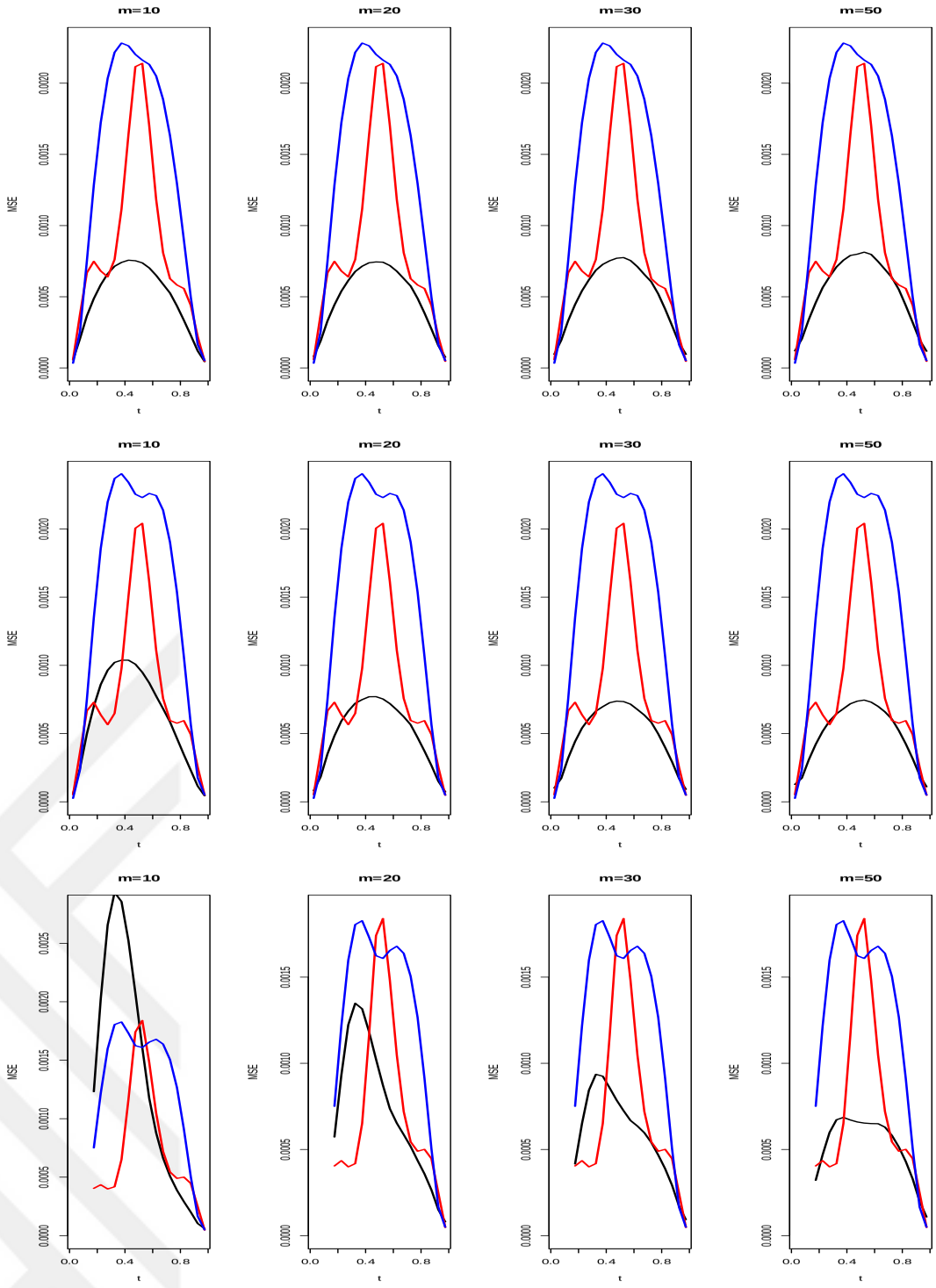


Figure 5.31 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Asymmetric Negative Logistic model ($\alpha = 0.35$, $\beta = 0.75$) when $\tau = 0.1$ (top row), $\tau = 0.15$ (second row), and $\tau = 0.25$ (bottom row).

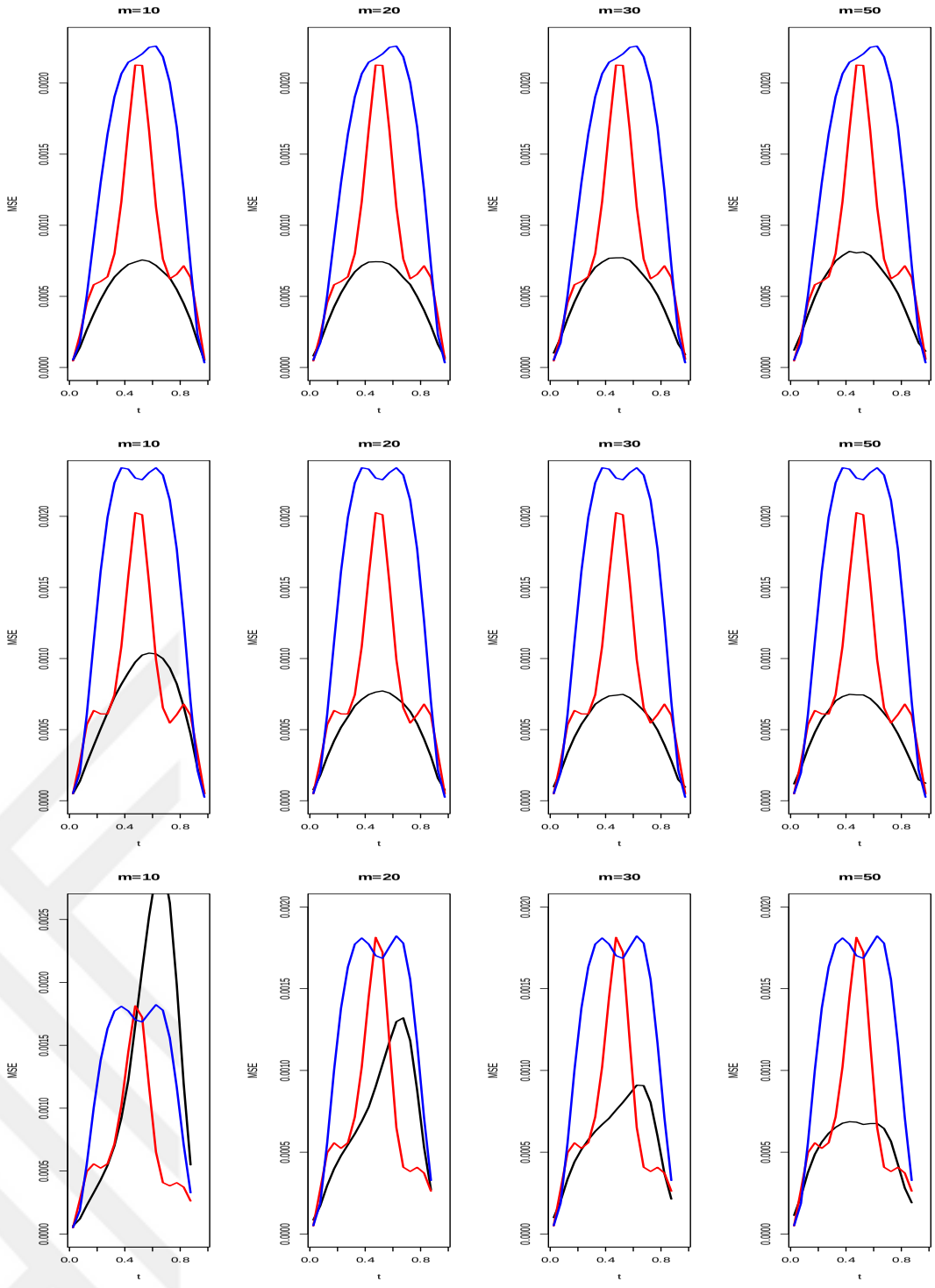


Figure 5.32 MSE of $\hat{A}$ for $\hat{A}_{CBK}$ (black line), $\hat{A}_{CFG}$ (blue line) and $\hat{A}_{PIC}$ (red line) based on 1000 sample size of 300 from Asymmetric Negative Logistic model ($\alpha = 0.75$, $\beta = 0.35$) when $\tau = 0.1$ (top row), $\tau = 0.15$ (second row), and $\tau = 0.25$ (bottom row).

Table 5.18 Median of Mean Squared Error of $\hat{A}$ from Asymmetric Negative Logistic model with parameters $(0.75, 0.35)$ for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level.

	Asymmetric Negative Logistic $(0.75, 0.35)$					
	CBK				PIC	CFG
MSE	$m = 10$	$m = 20$	$m = 30$	$m = 50$		
$\tau = 0.10$	0.0002815	0.0002966	0.0003038	0.0003307	0.0006667	0.0013760
$\tau = 0.15$	0.0004453	0.0002999	0.0002956	0.0003099	0.0006524	0.0012390
$\tau = 0.25$	0.0013210	0.0005478	0.0003878	0.0003134	0.0005547	0.0008831

Table 5.19 Statistically significant comparisons in Median of Mean Squared Error of $\hat{A}$ from Logistic model for monvex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level.

Logistic Copula				
	m=10	m=20	m=30	m=50
$\tau = 0.10$	$A_{CBK} < A_P < A_{CFG}$	$A_{CBK} < A_P < A_{CFG}$	$A_{CBK} < A_P < A_{CFG}$	$A_{CBK} < A_P < A_{CFG}$
$\tau = 0.25$	$A_{CFG} < A_{CBK} < A_P$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$
$\tau = 0.50$	$A_{CFG} < A_P < A_{CBK}$	$A_{CFG} < A_P < A_{CBK}$	$A_{CFG} < A_{CBC} < A_P$	$A_{CBK} \leq A_{CFG} < A_P$
$\tau = 0.75$	$A_{CFG} < A_P < A_{CBK}$	$A_{CFG} < A_P < A_{CBK}$	$A_{CFG} < A_P < A_{CBK}$	$A_{CFG} \leq A_{CBC} < A_P$

Table 5.20 Statistically significant comparisons in Median of Mean Squared Error of $\hat{A}$ from Asymmetric Logistic model $(0.35, 0.75)$ for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level.

Asymmetric Logistic Copula $(0.35, 0.75)$				
	m=10	m=20	m=30	m=50
$\tau = 0.10$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$
$\tau = 0.15$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$
$\tau = 0.25$	$A_{CFG} < A_P < A_{CBK}$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} \leq A_{CFG} < A_P$

Table 5.21 Statistically significant comparisons in Median of Mean Squared Error of $\hat{A}$ from Asymmetric Logistic model $(0.75, 0.35)$ for convex Bernstein copula estimator, Pickands estimator and CFG estimator of different values of dependence level.

Asymmetric Logistic Copula $(0.75, 0.35)$				
	m=10	m=20	m=30	m=50
$\tau = 0.10$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$
$\tau = 0.15$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$
$\tau = 0.25$	$A_{CFG} < A_P < A_{CBK}$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} < A_{CFG} < A_P$	$A_{CBK} \leq A_{CFG} < A_P$

CHAPTER SIX
NEW ESTIMATION OF PICKANDS DEPENDENCE FUNCTION: BAYESIAN APPROACH

6.1 Technical Introduction

This chapter provides a brief review for Bayesian approach to statistical inference and Monte Carlo Markov Chain methods that have been used in this thesis.

The Bayesian paradigm is based on specifying a probability model for the observed data, given a vector of unknown (but non-constant) parameters, leading to the likelihood function, and a probability model for this unknown vector of parameters which is called prior distribution. Inference concerning the model parameter is then based on the posterior distribution which is obtained by Bayes' theorem. (Chen et al., 2000)

In most of the cases, the posterior distribution does not have an analytical closed form, or a quite closed form complex posterior distribution leads to the problems (calculating the integrals or optimization of a function) that do not admit an analytical solution. These problems can be solved by finding an adequate approximation for the problem by sampling from the posterior distribution. This dilemma leads to the following questions: How do we sample from the multivariate posterior distribution when no closed form is available for it? This question had led to an enormous literature for computational methods for sampling from a given multivariate distribution (posterior distribution) as well as methods for approximating the integrals. (Chen et al., 2000)

6.1.1 The Bayesian Approach to Statistical Inference

Let $\mathscr{D} = \{x_1, \ldots, x_n\}$ be a sample from a distribution function with probability density function $f(.|\boldsymbol{\theta})$, and with an unknown parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\boldsymbol{\Theta}$ denotes the parameter space of $\boldsymbol{\theta}$, then the associated likelihood function is

$$L(\boldsymbol{\theta}|\mathscr{D}) = f_{\boldsymbol{\theta}}(x_1, \ldots, x_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) \,. \tag{6.1}$$

This quantity is a fundamental entity for the analysis of the information provided about $\boldsymbol{\theta}$ by the sample $\mathscr{D}$ in both frequentist and Bayesian approaches.

In Bayesian approach, we assume that $\boldsymbol{\theta}$ is random and has a *prior* distribution denoted by $\pi(\boldsymbol{\theta})$. This approach is based upon Bayes' theorem, which, states that

$$\pi(\boldsymbol{\theta}|\mathscr{D}) = \frac{L(\boldsymbol{\theta}|\mathscr{D})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} L(\boldsymbol{\theta}|\mathscr{D})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{6.2}$$

The posterior distribution represents our beliefs about the model parameters after having the observed data $\mathscr{D}$. The prior is concerned with our original beliefs about $\boldsymbol{\theta}$ before observing the data. Thus, Bayes' theorem allows us to update our prior beliefs on the basis of the observed data to obtain our posterior beliefs, which form the basis for our inference.

From (6.2), it is clear that $\pi(\boldsymbol{\theta}|\mathscr{D})$ is proportional to the likelihood multiplied by the prior,

$$\pi(\boldsymbol{\theta}|\mathscr{D}) \propto L(\boldsymbol{\theta}|\mathscr{D})\pi(\boldsymbol{\theta}),$$

and thus it involves a contribution form of the observed data through $L(\boldsymbol{\theta}|\mathscr{D})$, and a contribution form of the prior distribution quantified through $\pi(\boldsymbol{\theta})$. The quantity $m(\mathscr{D}) = \int_{\boldsymbol{\Theta}} L(\boldsymbol{\theta}|\mathscr{D})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the *normalizing constant* of $\pi(\boldsymbol{\theta}|\mathscr{D})$, and is often called the *marginal* distribution of the data or the prior predictive distribution.

6.1.2 Prior Distributions

Specification of the prior distribution of the parameter is obviously a key element of the Bayesian approach, and there are several diverse schools of thought when it comes to assigning priors; these can be loosely categorised into *subjective*, *objective*, and *empirical* or *hierarchical priors* approaches. (Gholami, 2008)

6.1.2.1 Subjective Priors

The subjective Bayesian attempts to encapsulate prior knowledge as fully as possible, be it in the form of previous experimental data or expert knowledge. It is often difficult to articulate qualitative experience or beliefs in mathematical form, but one of the very convenient and analytically favourable class of subjective priors are *conjugate* priors in the *exponential family* of distributions. Generally speaking, a prior is conjugate if the posterior distribution resulting from multiplying the likelihood and prior terms is of the

same form as the prior. (Beal, 2003) Expressed mathematically:

$$\pi(\boldsymbol{\theta}|\tilde{\boldsymbol{\xi}}) = \pi(\boldsymbol{\theta}|\mathscr{D}) \propto L(\boldsymbol{\theta}|\mathscr{D})\pi(\boldsymbol{\theta}|\boldsymbol{\xi}), \tag{6.3}$$

where $\pi(\boldsymbol{\theta}|\boldsymbol{\xi})$ is the prior distribution which is specified by a *hyperparameter*, $\boldsymbol{\xi}$.

Conjugate priors have at least three advantages:

- they often lead to analytically tractable Bayesian integrals,
- if computing the posterior in (6.3) is tractable, then one can be assured that subsequent inferences, based on using the posterior as prior, will also be tractable,
- they have an intuitive interpretation as expressing the results of previous (or indeed imaginary) observations under the model

For more detail in conjugate priors, see Robert (2001).

6.1.2.2 Objective Priors

There are several reasons why a one may want to resort to the use of objective priors (sometimes called *noninformative* priors):

- often one has little expertise and does not want to sway the inference process in any particular direction unknowingly,
- it may be difficult or impossible to elicit expert advice or translate expert opinions into a mathematical form for the prior
- one may want inference to be robust to misspecifications of the prior.

A noninformative prior favored by most Bayesians is socalled *Jeffreys' prior*, which is related to Fisher's information $I(\boldsymbol{\theta})$ by

$$\pi^J(\boldsymbol{\theta}) = |I(\boldsymbol{\theta})|^{1/2},$$

where $|I|$ denotes the determinant of the matrix I. For more details, see Bernardo and Smith (1994).

6.1.2.3 Empirical Bayes and Hierarchical Priors

When there are many common parameters in the vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$, it often makes sense to consider each parameter as being drawn from the same prior distribution. An example of this would be the prior specification of the means of each of the Gaussian components in a mixture model in which there is generally no priori reason to expect any particular component to be different from another. The parameter prior is then formed from integrating with respect to a *hyperprior*, $\pi(\boldsymbol{\xi})$, with hyperparameter $\boldsymbol{\xi}$:

$$\pi(\boldsymbol{\theta}) = \int_{\Xi} \pi(\boldsymbol{\xi}) \prod_{j=1}^{p} \pi(\theta_j|\boldsymbol{\xi}) \, d\boldsymbol{\xi}. \tag{6.4}$$

Therefore, each parameter is independent given the hyperparameter, although they are dependent marginally. Hierarchical priors are useful even when applied only to a single parameter, often offering a more intuitive interpretation for the parameter's role. For example, the precision parameter σ^2 for a Gaussian variable is often given a (conjugate) inverse Gamma prior, which itself has two hyperparameters (a,b) corresponding to the shape and scale of the prior. Interpreting the marginal distribution of the variable in this generative sense is often more intuitively appealing than simply enforcing a Student-*t* prior. Hierarchical priors are often designed using conjugate forms (described above), both for analytical ease and also previous knowledge which can be readily expressed.

The phrase empirical Bayes refers to the practice of optimising the hyperparameters (e.g. $\boldsymbol{\xi}$) of the priors, so as to maximise the marginal likelihood of a data set $L(\boldsymbol{\eta}|\mathscr{D})$. In this way Bayesian approach can be seen as maximum marginal likelihood learning, where there are always distributions over the parameters, but the hyperparameters are optimised just as in maximum likelihood learning. This practice is somewhat suboptimal as it ignores the uncertainty in the hyperparameter. Alternatively, a more coherent approach is to define priors over the hyperparameters and priors on the parameters of those priors, etc., to the point where at the top level one is content to leave those parameters unoptimised. With sufficiently vague priors at the top level, the posterior distributions over intermediate parameters should be determined principally by the data. In this fashion, no parameters are actually ever fit to the data, and all predictions and inferences are based on the posterior distributions over the parameters. (Beal, 2003)

6.1.3 Point Estimation

From a practical perspective, this complex multi-dimensional posterior distribution may be difficult to interpret in terms of the physical process under study. In practice, we often summarize the distribution by calculating a variety of interpretable univariate summary statistics such as posterior means and variances for example. (Brooks, 2003)

Point estimates for parameters of interest are obtained via the specification of a loss function. This approach stems from the gaming roots of the Bayesian approach. The Bayesian approach to parameter estimation is to choose the value which minimizes their expected loss under the posterior distribution. (Brooks, 2003). For example, the quadratic error loss function for estimating the true value $\boldsymbol{\theta}$ by $\widehat{\boldsymbol{\theta}}$ is given by

$$L(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) = \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|^2,$$

which increases in size the further our estimate lies from the truth. With this particular loss function, it turns out that for any problem the estimate (i.e. value of $\widehat{\boldsymbol{\theta}}$) which minimizes the expected value of this loss function over values of $\boldsymbol{\Theta}$ is simply the posterior mean:

$$\widehat{\boldsymbol{\theta}} = \mathbb{E}_{\pi(\boldsymbol{\theta}|\mathscr{D})}[\boldsymbol{\theta}|\mathscr{D}] = \int_{\Theta} \boldsymbol{\theta}\, \pi(\boldsymbol{\theta}|\mathscr{D})\, d\boldsymbol{\theta} = \frac{\int_{\Theta} \boldsymbol{\theta}\, L(\boldsymbol{\theta}|\mathscr{D})\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}}{\int_{\Theta} L(\boldsymbol{\theta}|\mathscr{D})\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}}, \tag{6.5}$$

for a given sample $\mathscr{D}$.

Other loss functions would lead to the other estimates for the parameter of interest such as MAP. The MAP estimator *maximum a posteriori estimator* is defined as

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \pi(\boldsymbol{\theta}|\mathscr{D}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \pi(\boldsymbol{\theta}) L(\boldsymbol{\theta}|\mathscr{D}), \tag{6.6}$$

where the function to maximize is usually provided in closed form.

6.1.4 Bayesian Computation

Although the Bayesian recipe for inference is conceptually simple, a practical problem with this approach is the difficulty associated with exploring and summarizing realistically complex posterior distributions. In most practical problems, the integrals resulting from inference procedure, such as (6.5) and (6.6), do not admit an analytical solution and computational tools are required to approximate them. Furthermore, in the most models and applications, $m(\mathscr{D})$, the marginal distribution of data, does not have a

closed form and the posterior is represented by an unnormalized density, and thus the Bayes factor will not be in the closed form and should be approximated. (Gholami, 2008)

The introduction to the statistical literature of the technique known as *Markov chain Monte Carlo* (MCMC) in the late 1980s overcomes this problem and greatly simplifies the Bayesian analysis of even the most complex data.

6.1.4.1 The Integration Problem

We are given a density function $\pi(.)$, on some state space $\mathscr{X}$, which is possibly unnormalised but at least satisfies $0 < \int_{\mathscr{X}} \pi(x)dx < \infty$. We want to, say, estimate expectations of function $g : \mathscr{X} \to \mathbb{R}$

$$\mathbb{E}_\pi[g(X)] = \int_{\mathscr{X}} g(x)\pi(x)dx.$$

If $\mathscr{X}$ is high-dimensional, and $\pi(.)$ is a complicated function, then direct integration (either analytical or numerical) of this integral is infeasible.

The classical Monte Carlo solution to this problem is to simulate i.i.d. random variables $X_1, X_2, \ldots, X_N \sim \pi(.)$, and then estimate $\mathbb{E}_\pi[g(X)]$ by

$$\widehat{\mathbb{E}}_\pi[g(X)] = \frac{1}{N}\sum_{i=1}^{N} g(X_i) \tag{6.7}$$

This gives an unbiased estimate, having standard deviation of order $O(1/\sqrt{N})$. Furthermore, if $\mathbb{E}_\pi[g^2(X)] < \infty$, then by the classical Central Limit Theorem, the error $\widehat{\mathbb{E}}_\pi[g(X)] - \mathbb{E}_\pi[g(X)]$ will have a limiting normal distribution. The problem, however, is that if π is complicated, then it is very difficult to directly simulate i.i.d. random variables from $\pi(.)$.

The Markov chain Monte Carlo (MCMC) solution is to *simulate* sequences that are neither independent nor identically distributed, but converge in distribution to $\pi(.)$. It can be done by constructing a *Markov chain* on $\mathscr{X}$ which has $\pi(.)$ as a stationary distribution.

6.1.4.2 Markov Chain Monte Carlo Methods

Markov Chain Monte Carlo (MCMC) methods date from the original work of Metropolis et al. (1953), who were interested in methods for the efficient simulation of

the energy levels of atoms in a crystalline structure. The original idea was subsequently generalized by Hastings (1970), but its true potential was not fully realized within the statistical literature until Gelfand and Smith (1990) demonstrated its application to the estimation of integrals commonly occurring in the context of Bayesian statistical inference. (Brooks, 2003)

MCMC methods work by constructing that is known as a *Markov chain* (see Meyn and Tweedie (1993)), which is essentially a series of observations generated one after the other so that the value at any point in the sequence depends only upon the preceding value. The key to MCMC methods is to construct a chain which has the target distribution (posterior distribution) as a stationary distribution.

In particular, the marginal distributions of the sampled values are exactly that corresponding to the posterior distribution of interest, so that the sample mean of the sequence approximates the true mean under the posterior distribution. In the following, we present those methods which will be used in this thesis. For a complete review of these methods, see, Robert and Casella (2004).

6.1.4.3 Gibbs Sampler

The Gibbs sampler may be one of the best known MCMC sampling algorithms in the Bayesian computational literature. The Gibbs sampler is founded on the ideas of Grenander (1983), while the formal term is introduced by Geman and Geman (1984). The primary bibliographical landmark for Gibbs sampling in problems of Bayesian inference is Gelfand and Smith (1990). A similar idea termed as *data augmentation* is introduced by Tanner and Wong (1987).

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ be a p-dimensional vector of parameters and let $\pi(\boldsymbol{\theta}|\mathscr{D})$ be its posterior distribution given data $\mathscr{D}$. Then, the basic scheme of the Gibbs sampler is given as follows:

Gibbs Sampling Algorithm

Step 0. Chose an arbitrary starting point $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \ldots, \theta_p^{(0)})'$, and set $i = 0$.

Step 1. Generate $\boldsymbol{\theta}^{(i+1)} = (\theta_1^{(i+1)}, \ldots, \theta_p^{(i+1)})'$ as follows:

- Generate $\theta_j^{(i+1)} \sim \pi(\theta_j|\theta_{-j}^{(i+1)}, \mathscr{D})$ for $j = 1, \ldots, p$.

Step 2. Set $i = i+1$, and go to **Step 1**.

Note that in the above $\pi(\theta_j|\theta_{-j}, \mathscr{D})$ stands for the full conditional distribution of $\boldsymbol{\theta}$ and $\mathscr{D}$ which is given by $\pi(\theta_j|\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_p\mathscr{D})$.

Thus each component of $\boldsymbol{\theta}$ is visited in the natural order and cycle in this scheme requires generation of p random variates. Gelfand and Smith (1990) show that under certain regularity conditions, the vector sequence $\{\boldsymbol{\theta}^{(i)},\ i = 1, 2, \ldots\}$ has a stationary distribution $\pi(\boldsymbol{\theta}|\mathscr{D})$. Schervish and Carlin (1992) provide a sufficient condition that guarantees geometric convergence. (Chen et al., 2000) For complete review of Gibbs sampling, see Robert and Casella (2004).

6.1.4.4 Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm is developed by Metropolis et al. (1953) and subsequently generalized by Hastings (1970).

Let $q(\boldsymbol{\theta}, \boldsymbol{\eta})$ be a proposal density, which is also termed as a *candidate generating density*, such that

$$\int_{\eta} q(\boldsymbol{\theta}, \boldsymbol{\eta}) d\boldsymbol{\eta} = 1.$$

Also let $\mathscr{U}(0,1)$ denotes the uniform distribution over $(0,1)$. Then, a general version of the Metropolis-hastings algorithm for sampling from the posterior distribution $\pi(\boldsymbol{\theta}|\mathscr{D})$ can be described as follows:

Metropolis-Hastings Algorithm

Step 0. Chose an arbitrary starting point $\boldsymbol{\theta}^{(0)}$ and set $i = 0$.

Step 1. Generate a candidate points $\boldsymbol{\theta}^*$ from $q(\boldsymbol{\theta}^{(i)}, .)$ and u from $\mathscr{U}(0,1)$.

Step 2. Set $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^*$ if $u \leq \rho(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\theta}^*)$ and $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)}$ otherwise, when the acceptance probability is given by

$$\rho(\boldsymbol{\theta}^{(i+1)}, \boldsymbol{\theta}^*) = \min\left\{1, \frac{\pi(\boldsymbol{\eta}|\mathscr{D})q(\boldsymbol{\eta}, \boldsymbol{\theta})}{\pi(\boldsymbol{\theta}|\mathscr{D})q(\boldsymbol{\theta}, \boldsymbol{\eta})}\right\}.$$

Step 3. Set $i = i + 1$, and go to **Step 1**.

The above algorithm is very general. When $q(\boldsymbol{\theta}, \boldsymbol{\eta}) = q(\boldsymbol{\eta})$, the Metropolis-Hasting algorithm reduces to the *independence chain* Metropolis algorithm (see Tierny (1994)).

Another family of proposal densities is given by the form $q(\boldsymbol{\theta}, \boldsymbol{\eta}) = q_1(\boldsymbol{\theta} - \boldsymbol{\eta})$, where $q_1(.)$ is a multivariate density. The candidate $\boldsymbol{\theta}^*$ is drawn according to the process $\boldsymbol{\theta}^* = \boldsymbol{\theta} + \boldsymbol{\omega}$, where $\boldsymbol{\omega}$ is called the increment random variable and follows the distribution q_1. Because the candidate is equal to the current value plus noise, this case is called *random walk chain*.

The performance of a Metropolis-Hastings algorithm depends on the choice of a proposal density q. As discussed in Chib and Greenberg (1995), the spread of proposal density q affects the behavior of the chain in at least two dimension: One is the "acceptance rate" (the percentage of times a move to a new point is made), and the other is the region of the sample space that is covered by the chain. If the spread is extremely large, some of the generated candidates will have a low probability of being accepted. On the other hand, if the spread is chosen too small, the chain will take longer to traverse the support of the density. Both of these situation are likely to be reflected in high autocorrelation across sample values. (Chen 2012) In the context of q_1 (the random walk proposal density), Robert et al. (1997) show that if the target proposal densities are normal, then the scale of latter should be tuned so that the acceptance rate is approximately 0.45 in one-dimensional problems and approximately 0.23 as the number of dimensions approaches infinity, with the optimal acceptance rate being around 0.25 in six dimension. For the independent chain, in which we take $q(\boldsymbol{\theta}, \boldsymbol{\eta}) = q(\boldsymbol{\eta})$, it is important to ensure that the tails of the proposal density $q(\boldsymbol{\eta})$ dominate those of the target density $\pi(\boldsymbol{\theta}|\mathscr{D})$ (Geweke, 1989).

6.1.4.5 The Reversible Jump Technique

A variable dimension model is, to quote Peter Green, a *"model where one of the things you do not know is the number of things you do not know"*. This means that the statistical model under consideration is not defined precisely enough for the dimension of the parameter space to be fixed.

The simulation methods introduced before can not take into account this extra dimension variability and we need more advanced techniques that are capable of moving between parameter spaces of different dimensions to explore whole parameter space. (Gholami, 2008) The standard Metropolis-Hastings algorithm described earlier is incapable of such movements, whereas the *reversible jump algorithm* of Green (1995) is an extension of the standard Metropolis-Hastings algorithm to allow exactly this possibility.

Suppose that we have a countable collection of candidate models $\{\mathscr{M}_j,\ j=1,\ldots,m\}$ where model $\mathscr{M}_j$ has a vector $\boldsymbol{\theta}_j$ of unknown parameters, with dimension p_j, which may vary from model to model. Under model $\mathscr{M}_j$, the posterior distribution of $\boldsymbol{\theta}_j$ takes the form

$$\pi(\boldsymbol{\theta}_j|\mathscr{D},\mathscr{M}_j) \propto L(\boldsymbol{\theta}_j|\mathscr{D},\mathscr{M}_j)\pi(\boldsymbol{\theta}_j|\mathscr{M}_j)\,,$$

where $L(\boldsymbol{\theta}_j|\mathscr{D},\mathscr{M}_j)$ is the likelihood function and $\pi(\boldsymbol{\theta}_j|\mathscr{M}_j)$ is the prior distribution. Then our aim is to sample from the joint distribution of $(j,\boldsymbol{\theta}_j)$ given the data $\mathscr{D}$ which takes form

$$\pi(j,\boldsymbol{\theta}_j|\mathscr{D}) \propto \pi(j)L(\boldsymbol{\theta}_j|\mathscr{D},\mathscr{M}_j)\pi(\boldsymbol{\theta}_j|\mathscr{M}_j)\,,$$

for $j=1,\ldots,m$.

The reversible jump technique does this task by removing the difference in the dimension of the models by supplementing the corresponding parameters $\boldsymbol{\theta}_j$ and $\boldsymbol{\theta}_k$ by auxiliary variables $\boldsymbol{u}$ and $\boldsymbol{u}'$ such that

$$(\boldsymbol{\theta}_j,\boldsymbol{u}) \quad \text{and} \quad (\boldsymbol{\theta}_k,\boldsymbol{u}')$$

are in bijection:

$$(\boldsymbol{\theta}_k,\boldsymbol{u}') = \Psi_{j\to k}(\boldsymbol{\theta}_j,\boldsymbol{u})\,. \tag{6.8}$$

Now we can use the standard Metropolis-Hasting algorithm to move between the two models. This algorithm has been discussed by several authors including Green (1995), Dellaportas et al. (2000) and Robert and Casella (2004). It involves sampling from $\pi(j,\boldsymbol{\theta}_j|\mathscr{D})$, which can be described as follows:

The Reversible Jump Algorithm

If the current state of the chain is $(j,\boldsymbol{\theta}_j)$, then:

Step 1. Propose a new model $\mathscr{M}_{j^\star}$ with probability $p(j^\star|j)$.

Step 2. Generate $\boldsymbol{u}$ from a specified proposal density $q(\boldsymbol{u}|\boldsymbol{\theta}_j, j, j^\star)$.

Step 3. Set $(\boldsymbol{\theta}^\star_{j^\star}, \boldsymbol{u}^\star) = \Psi_{j\to j^\star}(\boldsymbol{\theta}_j, \boldsymbol{u})$, where $\Psi_{j\to j^\star}$ is a bijection between between $(\boldsymbol{\theta}_j, \boldsymbol{u})$ and $(\boldsymbol{\theta}^\star_{j^\star}, \boldsymbol{u}^\star)$, and the length of $\boldsymbol{u}$ and $\boldsymbol{u}^\star$ must satisfy $p_j + \dim(\boldsymbol{u}) = p^\star_j + \dim(\boldsymbol{u}^\star)$.

Step 4. Accept the proposed move to $(\boldsymbol{\theta}^\star_{j^\star}, \boldsymbol{u}^\star)$ with probability

$$\rho = \min\left\{1, \frac{\pi(j^\star, \boldsymbol{\theta}^\star_{j^\star}|\mathscr{D})p(j|j^\star)q(\boldsymbol{u}^\star|\boldsymbol{\theta}^\star_{j^\star}, j^\star, j)}{\pi(j, \boldsymbol{\theta}_j|\mathscr{D})p(j^\star|j)q(\boldsymbol{u}|\boldsymbol{\theta}_j, j, j^\star)} \times \left|\frac{\partial \Psi_{j\to j^\star}(\boldsymbol{\theta}_j, \boldsymbol{u})}{\partial(\boldsymbol{\theta}_j, \boldsymbol{u})}\right|\right\}$$

6.2 The Statistical Model

6.2.1 Cubic B-spline Regression Model

Let $t = \frac{\ln v}{\ln(uv)}$ and then $A(t) = \frac{\ln C(u,v)}{\ln(uv)}$. The transformation from(u,v) to $(t, A(t))$ reduce to a convex curve if and only if, the copula C is the form by (5.1). For more details see Cormier et al. (2014).

Let $\{(X_i, Y_i)\}_{i=1}^n$ be a random sample of size n from an extreme value copula C and marginals F and G. Consider $\hat{F}_n$ and $\hat{G}_n$ be the empirical distributions of F and G which are defined respectively as

$$\hat{F}_n(x_i) = \frac{1}{n}\sum_{j=1}^{n} I(X_j \le x_i)$$

$$\hat{G}_n(y_i) = \frac{1}{n}\sum_{j=1}^{n} I(Y_j \le y_i)$$

Let $\hat{U}_i = \hat{F}_n(X_i)$ and $\hat{V}_i = \hat{G}_n(Y_i)$ for $i = 1, ..., n$. The empirical copula of $\{(\hat{U}_i, \hat{V}_i)\}_{i=1}^n$ is

$$\hat{C}_n(\hat{u}_i, \hat{v}_i) = \frac{1}{n}\sum_{j=1}^{n} I(\hat{U}_j \le \hat{u}_i, \hat{V}_j \le \hat{v}_i).$$

Define $T_i = \frac{\ln(\hat{V}_i)}{\ln(\hat{U}_i\hat{V}_i)}$ and $Y_i = \frac{\ln \hat{C}_n(\hat{U}_i, \hat{V}_i)}{\ln(\hat{U}_i\hat{V}_i)}$. The regression problem is to fit the Pickands dependence function $A(t)$ through the pairs (T_i, Y_i), i.e.

$$y_i = A(t_i) + \sigma\varepsilon_i$$

As we do not consider any specific functional form on A, we model $\{Y_i\}_{i=1}^n$ as

$$Y_i = g(t_i; \beta) + \sigma\varepsilon_i, \quad i = 1, 2, \ldots, n, \quad t \in [0, 1]$$

where $g(t_i;\beta)$ is an arbitrary smooth function and $\varepsilon_i \sim \mathcal{N}(0,1)$. Suppose that m interior knots are $k_1, k_2, \ldots, k_m$ and g is represented by the cubic B-spline basis, as follows

$$g(t;\beta) = \sum_{j=1}^{m+4} \beta_j S_{j,3}(t,\boldsymbol{k})$$

Matrix form of the above model can be written as

$$\mathbf{Y} = X_{m,\boldsymbol{k}}\beta + \sigma\varepsilon$$

where

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{(n\times 1)}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{m+4} \end{bmatrix}_{((m+4)\times 1)}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{(n\times 1)}$$

$$X_{m,\boldsymbol{k}} = \begin{bmatrix} S_{1,3}(t_1,\boldsymbol{k}) & S_{2,3}(t_1,\boldsymbol{k}) & \ldots & S_{m+4,3}(t_1,\boldsymbol{k}) \\ S_{1,3}(t_2,\boldsymbol{k}) & S_{2,3}(t_2,\boldsymbol{k}) & \ldots & S_{m+4,3}(t_2,\boldsymbol{k}) \\ \vdots & \vdots & & \vdots \\ S_{1,3}(t_n,\boldsymbol{k}) & S_{2,3}(t_n,\boldsymbol{k}) & \ldots & S_{m+4,3}(t_n,\boldsymbol{k}) \end{bmatrix}_{(n\times(m+4))}.$$

6.2.2 The Bayesian Model

The Bayesian approach to model fitting and making inferences is based on specifying a probability model to all unknown parameters which is called prior distributions and specifying a probability model for the observed data given the parameters, leading to the likelihood function, and determining the posterior distributions by Bayes theorem.

6.2.2.1 The Priors

To determine the prior distributions for the parameters, we should check the number of unknown parameters of the model. The model parameters are: The number of interior knots, m, knots locations, $\boldsymbol{k}$, model coefficients, $\boldsymbol{\beta}$, and the variance of the errors, σ^2.

If one fixes the number and the locations of the knots, the only unknown parameters in the model are $\boldsymbol{\beta}$ and variance σ^2. The Normal-Inverse-Gamma (NIG) is the popular *conjugate* choice for the prior of $\boldsymbol{\beta}$ and σ^2, which is usually denoted by

$$\pi(\boldsymbol{\beta},\sigma^2) = NIG(\boldsymbol{\mu},\boldsymbol{V},a,b), \quad a,b>0$$

that is;

$$\pi(\boldsymbol{\beta},\sigma^2)=\pi(\boldsymbol{\beta}|\sigma^2)\times\pi(\sigma^2)=\mathscr{N}_p(\boldsymbol{\mu},\sigma^2\boldsymbol{V})\times\mathscr{IG}(a,b)$$

where $\mathscr{IG}(a,b)$ denotes the Inverse-Gamma distribution with parameters a and b. Then, we obtain the prior distribution,

$$\begin{aligned}\pi(\boldsymbol{\beta},\sigma^2) &= \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{V}|^{\frac{1}{2}}(\sigma^2)^{\frac{p}{2}}}\exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{\mu})'\boldsymbol{V}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\right]\\ &\times \frac{b^a}{\Gamma(a)}\left(\frac{1}{\sigma^2}\right)^{a+1}\exp-\frac{b}{\sigma^2}\\ &\propto \left(\frac{1}{\sigma^2}\right)^{a+\frac{p}{2}+1}\exp\left[-\frac{1}{\sigma^2}\left(b+\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu})'\boldsymbol{V}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\right)\right].\end{aligned}$$

If we let $a=b=0$ and $\boldsymbol{V}=n(X'_{m,\boldsymbol{k}}X_{m,\boldsymbol{k}})^{-1}$, the resulting prior is Zellner's G-prior. Following Gholami (2010), for a given number of knots, m, we consider a discrete uniform distribution on the knots locations as follows

$$\pi(\boldsymbol{k}|m)=\binom{K}{m}^{-1},\quad m=1,2,\ldots,M$$

where K is the size of candidate set of knot locations and M is the maximum number of knot sequence allowed. By using the above prior one ensures that all models with dimension m have equal weights. Finally, to determine a prior to m , we consider uniform distribution over all possible values, i.e.

$$\pi(m)=\frac{1}{M},\quad m\in\{1,2,\ldots,M\}.$$

Then the prior for the parameter vector is

$$\pi(\boldsymbol{\theta})=\pi(\boldsymbol{\beta},\sigma^2,\boldsymbol{k},m)=\pi(\boldsymbol{\beta}|\sigma^2,\boldsymbol{k},m)\pi(\sigma^2)\pi(\boldsymbol{k}|m)\pi(m).$$

6.2.2.2 The Likelihood

The likelihood function for the model is defined as the joint probability of the observed data with the given parameters. Since $X_{m,\boldsymbol{k}}$ is fixed, the likelihood function is obtained as

$$f(\boldsymbol{y}|\boldsymbol{\theta})=f(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{k},m)=\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y}-X_{m,\boldsymbol{k}}\boldsymbol{\beta})'(\boldsymbol{y}-X_{m,\boldsymbol{k}}\boldsymbol{\beta})\right]$$

where the error term is distributed as $\mathscr{N}(0,1)$.

6.2.2.3 The Posterior

The first step of the inference procedure is to obtain the joint posterior distribution of $(\boldsymbol{\beta}, \sigma^2)$,

$$\pi(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) \propto \pi(\boldsymbol{\beta}, \sigma^2) f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2)$$

that is;

$$\pi(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) \propto \left(\frac{1}{\sigma^2}\right)^{a^*+\frac{p}{2}+1} \exp\left[-\frac{1}{\sigma^2}\left(b^* + \frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\xi})'\boldsymbol{\Psi}^{-1}(\boldsymbol{\beta}-\boldsymbol{\xi})\right)\right],$$

where

$$\boldsymbol{\xi} = (\boldsymbol{V}^{-1}X'_{m,\boldsymbol{k}}X_{m,\boldsymbol{k}})^{-1}(\boldsymbol{V}^{-1}\boldsymbol{\mu} + X'_{m,\boldsymbol{k}}\boldsymbol{y})$$

$$\boldsymbol{\Psi} = (\boldsymbol{V}^{-1}X'_{m,\boldsymbol{k}}X_{m,\boldsymbol{k}})^{-1}$$

$$a^* = a + \frac{n}{2}$$

$$b^* = b + \frac{1}{2}[\boldsymbol{\mu}'\boldsymbol{V}^{-1}\boldsymbol{\mu} + \boldsymbol{y}'\boldsymbol{y} + \boldsymbol{\xi}'\boldsymbol{\Psi}^{-1}\boldsymbol{\xi}].$$

Note that the marginal posterior distribution is Inverse-Gamma distribution with parameters a^* and b^*. The marginal posterior of $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta}|\boldsymbol{y})$, results from integrating out σ^2 from the above joint posterior distribution as,

$$\int \frac{b^{*a^*}}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Psi}|^{\frac{1}{2}}\Gamma(a^*)}\left(\frac{1}{\sigma^2}\right)^{a^*+\frac{p}{2}+1} \exp\left[-\frac{1}{\sigma^2}\left(b^* + \frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\xi})'\boldsymbol{\Psi}^{-1}(\boldsymbol{\beta}-\boldsymbol{\xi})\right)\right] d\sigma^2.$$

Let $\varphi = b^* + \frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\xi})'\boldsymbol{\Psi}^{-1}(\boldsymbol{\beta}-\boldsymbol{\xi})$, then

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) = \frac{b^{*a^*}}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Psi}|^{\frac{1}{2}}\Gamma(a^*)}\int\left(\frac{1}{\sigma^2}\right)^{a^*+\frac{p}{2}+1} \exp(-\frac{\varphi}{\sigma^2})d\sigma^2$$

by using the properties of Gamma function we have

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) = \frac{b^{*a^*}\Gamma(a^*+\frac{p}{2})}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Psi}|^{\frac{1}{2}}\Gamma(a^*)}\varphi^{-(a^*+\frac{p}{2})}.$$

By simplifying the equation, we see that $\pi(\boldsymbol{\beta}|\boldsymbol{y})$ is a multivariate t density with n degrees of freedom. i.e.

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) = \frac{\Gamma(\frac{n+p}{2})}{\Gamma(\frac{n}{2})\pi^{\frac{p}{2}}|n\Sigma|^{\frac{1}{2}}}\left[1 + \frac{(\boldsymbol{\beta}-\boldsymbol{\xi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}-\boldsymbol{\xi})}{n}\right]^{(-\frac{n+p}{2})}$$

where $n = 2a^*$, $\boldsymbol{\Sigma} = (\frac{b^*}{a^*})\boldsymbol{\Psi}$ and $p = m+4$.

Since we choose the Zellner's G-prior with $\mu = 0$ as a prior distribution of $(\boldsymbol{\beta}, \sigma^2)$, we should make some changes in the obtained posterior distribution and also marginal posterior distributions. Therefore, the posterior of the parameter vector is,

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y} - X_{m,\boldsymbol{k}}\boldsymbol{\beta})'(\boldsymbol{y} - X_{m,\boldsymbol{k}}\boldsymbol{\beta})\right].$$

The marginal posterior distribution of $\boldsymbol{\beta}$ is $(m+4)$-variate Student's t-distribution with n degrees of freedom, i.e.

$$\boldsymbol{\beta}|\boldsymbol{y}, m, \boldsymbol{k} \sim t_{m+4}(n; \boldsymbol{\xi}, \Sigma)$$

where

$$\boldsymbol{\xi} = \frac{n}{n+1}(X'_{m,\boldsymbol{k}}X_{m,\boldsymbol{k}})^{-1}X'_{m,\boldsymbol{k}}\boldsymbol{y},$$
$$\Sigma = \frac{\gamma}{n+1}(X'_{m,\boldsymbol{k}}X_{m,\boldsymbol{k}})^{-1},$$
$$\gamma = \boldsymbol{y}'(I_n - \frac{n}{n+1}X_{m,\boldsymbol{k}}(X'_{m,\boldsymbol{k}}X_{m,\boldsymbol{k}})^{-1}X_{m,\boldsymbol{k}})\boldsymbol{y},$$

then

$$\mathbb{E}[\boldsymbol{\beta}|\boldsymbol{y}, m, \boldsymbol{k}] = \frac{n}{n+1}(X'_{m,\boldsymbol{k}}X_{m,\boldsymbol{k}})^{-1}X'_{m,\boldsymbol{k}}\boldsymbol{Y}$$

and also the marginal posterior distribution of σ^2 given m and $\boldsymbol{k}$ is $IG(\frac{n}{2}, \frac{\gamma}{2})$. Then we obtain

$$\mathbb{E}[\sigma^2|\boldsymbol{y}, m, \boldsymbol{k}] = \frac{\gamma}{n-2}.$$

6.2.3 Posterior Sampling

The posterior distribution, $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, does not have an analytical closed form. Because of the complex form of this posterior, it does not admit an analytical solution, and should be approximated. In order to approximate the posterior distribution, we generate random numbers from posterior distribution by using *Gibbs sampling* method. To complete the sampling process by Gibbs method, we need to generate random numbers from the conditional distribution of $(\boldsymbol{k}, m)$. Note that due to the varying dimension of this marginal posterior, we are not able to use the standard sampling methods. So we use the *Reversible Jump* algorithm (Green 1995). The marginal posterior distribution of the $(\boldsymbol{k}, m)$ is $\boldsymbol{t}_m \in \mathbf{X}_m \subseteq \mathbf{R}^m$. Bayesian inference is then carried out based on the assumption that the

true model is unknown but comes from the class of models $\mathscr{M}_1, \mathscr{M}_2, \ldots$ where $\mathscr{M}_m$ denotes the model with exactly m knots. The overall subspace is $\mathbf{X} = \bigcup_{m=1}^{\infty} \mathbf{X}_m$ where $\mathbf{X}_m$ is the m dimensional space corresponding to model $\mathscr{M}_m$.

Due to the varying dimensionality of our problem, we should design move types between the subspaces of $\mathbf{X}_m$. This process allows the sampler to freely explore the combined parameter space if we use the Reversible Jump algorithm. We use three move types, ***birth***, ***death*** and ***relocation***. ***Birth*** and ***death*** types propose moves between different dimensions while ***relocation*** type proposes moves within a dimension. Throughout the procedure, we assume that the current model is of dimension m.

Birth. With probability $p_{b,m}$, propose to add a new knot at a chosen data point randomly from those which do not currently have a knot.

Death. With probability $p_{d,m}$, propose to remove a randomly chosen knot which is present in the model.

Relocation. With probability $p_{r,m}$, propose to alter a randomly chosen knot, say k_j, by swapping for a randomly chosen knot which is not present in the model.

We choose the proposal probabilities $p_{p,m} = p_{d,m} = p_{r,m} = \frac{1}{3}$ for $m = 2, \ldots, M-1$; $p_{d,1} = 0, p_{b,1} = \frac{1}{3}, p_{r,1} = \frac{2}{3}$ and $p_{b,M} = 0, p_{d,M} = \frac{1}{3}, p_{r,M} = \frac{2}{3}$. These probabilities can be chosen in any way but under the specification that they are more non-informative. The acceptance probability for a Reversible Jump algorithm (Green 1995) for a proposal move from model $\boldsymbol{k}$ (of dimension m) to $\boldsymbol{k}'$ (of dimension m') is given by

$$\begin{aligned}\rho_{\mathscr{M}_m \to \mathscr{M}_{m'}} &= \min\left\{1, \frac{p(y|\boldsymbol{k}')}{p(y|\boldsymbol{k})}\frac{\pi(\boldsymbol{k}')}{\pi(\boldsymbol{k})}\frac{q(\boldsymbol{k}|\boldsymbol{k}')}{q(\boldsymbol{k}'|\boldsymbol{k})}|J|\right\} \\ &= \min\left\{1, LR_{\mathscr{M}_m \to \mathscr{M}_{m'}} \times PR_{\mathscr{M}_m \to \mathscr{M}_{m'}}|J|\right\},\end{aligned}$$

where the ratio of marginal likelihood ($LR_{\mathscr{M}_m \to \mathscr{M}_{m'}}$), prior and proposal ratio ($PR_{\mathscr{M}_m \to \mathscr{M}_{m'}}$) together with a Jacobian term $|J|$, that accounts for the change in scale when moving between models are of potentially different dimensions. In our problem, the marginal likelihood ratio is

$$LR_{\mathscr{M}_m \to \mathscr{M}_{m'}} = \frac{p(\boldsymbol{y}|\boldsymbol{k}')}{p(\boldsymbol{y}|\boldsymbol{k})} = \frac{p(\boldsymbol{y}|\boldsymbol{k}'_{m'}, m')}{p(\boldsymbol{y}|\boldsymbol{k}_m, m)} = \frac{1}{\sqrt{n+1}}\left(\frac{\gamma_{m',\boldsymbol{k}'}}{\gamma_{m,\boldsymbol{k}}}\right)^{\frac{n}{2}}$$

$$PR_{\mathscr{M}_m \to \mathscr{M}_{m'}} = \frac{\pi(m')}{\pi(m)} \frac{\pi(\boldsymbol{k}'_{m'})}{\pi(\boldsymbol{k}_m)} \frac{\pi(\mathscr{M}_m)}{\pi(\mathscr{M}_{m'})} \frac{\pi(\boldsymbol{k}_m|\boldsymbol{k}'_{m'})}{\pi(\boldsymbol{k}'_{m'}|\boldsymbol{k}_m)}.$$

Note that because of the use of linear transformations (Birth, Death and Relocation), the Jacobian term is $|J| = 1$.

The prior and the proposal ratios for birth, death and relocation moves are $\frac{p_{d,m+1}}{p_{b,m}}$, $\frac{p_{b,m-1}}{p_{d,m}}$ and 1, respectively. Then the last move is just a standard update in a *Metropolis-Hasting* sampler. The rest of the algorithm consists of two usual Gibbs steps: updating both the coefficients and the variance which can be done easily. In the following algorithm, we give a pseudo-code that we use for sampling in our model.

0. Initialize the parameters $m^{(0)}, \boldsymbol{k}^{(0)}$

Given $m^{(l)}, \boldsymbol{k}^{(l)}$,

1. Re-set $p_{b,m}, p_{d,m}$ and $p_{r,m}$ considering the current value of $m^{(l)}$

2. Propose $m^{\star} = m^{(l)} + 1, m^{(l)} - 1$ or $m^{(l)}$ with probability $p_{b,m}, p_{d,m}$ or $p_{r,m}$

3. If $m^{\star} = m^{(l)} + 1$
   $m^{(l+1)} =$ ***BIRTH***()
   Else
   If $m^{\star} = m^{(l)} - 1$
   $m^{(l+1)} =$ ***DEATH***()
   Else
   $m^{(l+1)} =$ ***RELOCATION***()

4. Sample $\boldsymbol{\beta}^{(l+1)} | m^{(l+1)}, \boldsymbol{k}^{(l+1)}, \boldsymbol{Y}$

5. Sample $\sigma^{2(l+1)} | m^{(l+1)}, \boldsymbol{k}^{(l+1)}, \boldsymbol{Y}$

In nonparametric settings, the most important parameter to estimate is the functional form of the regression function $A(t) = \eta$ namely $\boldsymbol{\eta} = X\boldsymbol{\beta}$. The posterior distribution of $\boldsymbol{\eta}$ is a mixture of multivariate t-Student distributions

$$\pi(\boldsymbol{\eta}|\boldsymbol{y}) = \sum_{m,\boldsymbol{k}} \pi(m,\boldsymbol{k}|\boldsymbol{y}) \pi(\boldsymbol{\eta}_{m,\boldsymbol{k}}|m,\boldsymbol{k},\boldsymbol{y})$$

where

$$\pi(\boldsymbol{\eta}_{m,\boldsymbol{k}}|m,\boldsymbol{k},\boldsymbol{y}) = X_{m+4}\left(n, X_{m,\boldsymbol{k}}\boldsymbol{\xi}_m, \frac{\gamma_{m,\boldsymbol{k}}}{n} X_{m,\boldsymbol{k}} \boldsymbol{\Sigma}_{\boldsymbol{m},\boldsymbol{k}} X'_{m,\boldsymbol{k}}\right)$$

and

$$\boldsymbol{\xi}_m = \frac{n}{n+1}\left(X'_{m,\boldsymbol{k}} X_{m,\boldsymbol{k}}\right)^{-1} X'_{m,\boldsymbol{k}} \boldsymbol{y}, \qquad \boldsymbol{\Sigma}_{m,\boldsymbol{k}} = \frac{n}{n+1}\left(X'_{m,\boldsymbol{k}} X_{m,\boldsymbol{k}}\right)^{-1}.$$

The overall mean and variance of this mixture distribution can be estimated by using the sample produced from the algorithm.

6.3 Simulation Study

In this section, we consider four families of extreme value distributions; Asymmetric Logistic distribution with $\alpha = 0.35$, $\beta = 0.75$ and $\tau = 0.15$, Galambos, Logistic distributions with $\tau = 0.5$ and Extreme value t-student with four degrees of freedom and with $\tau = 0.5$. In each family, we generate a random sample of size 300 and estimate the Pickands dependence function by running 10000 iterations of algorithm following burn-ins of 5000. In each simulation, the proposed Bayesian Pickands estimator ($\hat{A}_{Bayesian}$) is compared to those of intrinsic versions of the rank-based Pickands and Caperaa-Fougeres-Genest (CFG) estimators (See, Genest and Segers (2009)) and true Pickands function which are denoted by $\hat{A}_{PIC}$, $\hat{A}_{CFG}$ and A, respectively. Figure (6.1) shows the estimated Bayesian Pickands dependence function (blue dashed line) and gray area is a 95% credible interval for the estimated curve for Asymmetric logistic distribution, Galambos distribution, Logistic distribution, and Extreme value t-student, respectively. Pink and green lines present Pickands and CFG estimators, respectively and red line is the true Pickands dependence function. Comparing with the other estimators, Bayesian Pickands estimator is closer to the true Pickands dependence function for all types of extreme-value copulas. For all the models, Bayesian estimator fits best even for the asymmetric model.

Figure (6.2) shows the sequence of m's and the corresponding approximated probability densities. The MAP estimation of m is 2. MAP estimation is the value of the parameter which maximizes the posterior distribution, see Gholami (2010). The upper plot gives the sequence of interior knots and the lower plot gives the corresponding probability densities for asymmetric logistic distribution, Galambos distribution, Logistic distribution and Extreme value t-student distribution, respectively. Figure (6.3) shows the mixing behavior of the σ^2 with corresponding running mean (red line) and its histogram. Figure (6.4) shows the sequence of MSE's against iterations. It is seen when the iteration number getting increased then MSE's are getting smaller as we expected.

In this section we also consider a real-data example which is well-known LOSS/ALAE

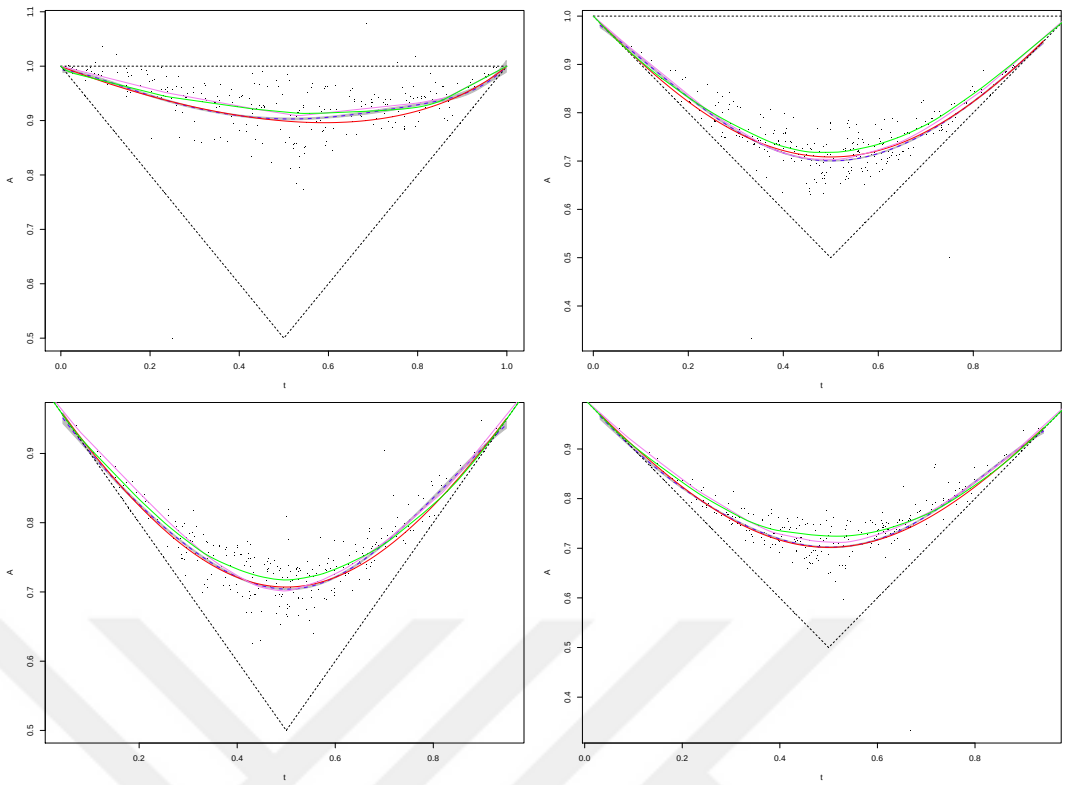


Figure 6.1 Plots with true Pickands dependence function A (red) and three estimates based on 1000 random samples of size 300 from the Asymmetric Logistic ($\alpha = 0.35$, $\beta = 0.75$, $\tau = 0.15$) (top left), Galambos ($\tau = 0.5$) (top right), Logistic ($\tau = 0.5$) (bottom left), Extreme-value t distributions ($\tau = 0.5$) (bottom right). $\hat{A}_{CFG}$ (green), $\hat{A}_{PIC}$ (pink), $\hat{A}_{Bayesian}$ (blue dashed line) and gray area is a 95% credible interval for the estimated curve.

data. Ties are present in the data which provide an example on the dependence between two risks where the dependence is modeled by extreme-value copula. For more details, see Frees and Valdez (1998). Test results are given in Figure (6.5). When we compare our estimator with the alternative estimators, Bayesian estimate of Pickands dependence function again has the best result. It represents the data best satisfying all the constraints of Pickands dependence function. Also, the MAP estimation of m is 2. It is concluded that when the iteration number is getting increased then the MSE is getting smaller.

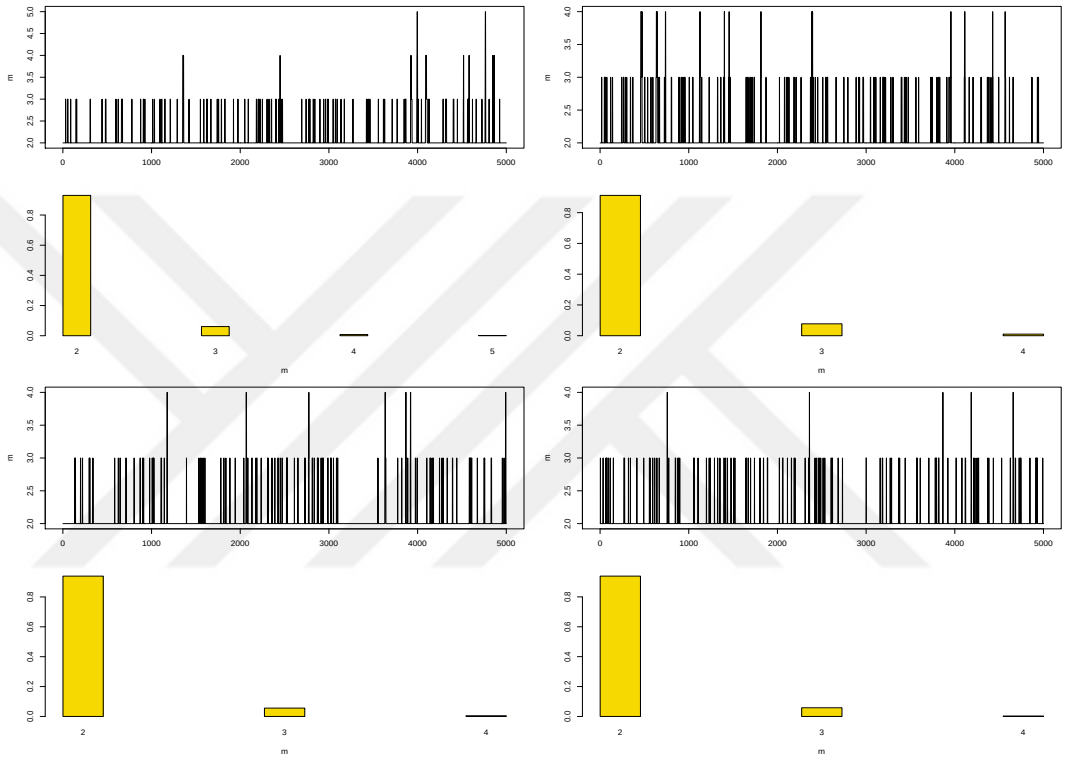


Figure 6.2 Sequence of m's of $\hat{A}_{Bayesian}$ and the corresponding approximated probability densities based on 1000 random samples of size 300 from Asymmetric Logistic ($\alpha = 0.35$, $\beta = 0.75$, $\tau = 0.15$) (top left), Galambos ($\tau = 0.5$) (top right), Logistic ($\tau = 0.5$) (bottom left), Extreme-value t distributions ($\tau = 0.5$) (bottom right).

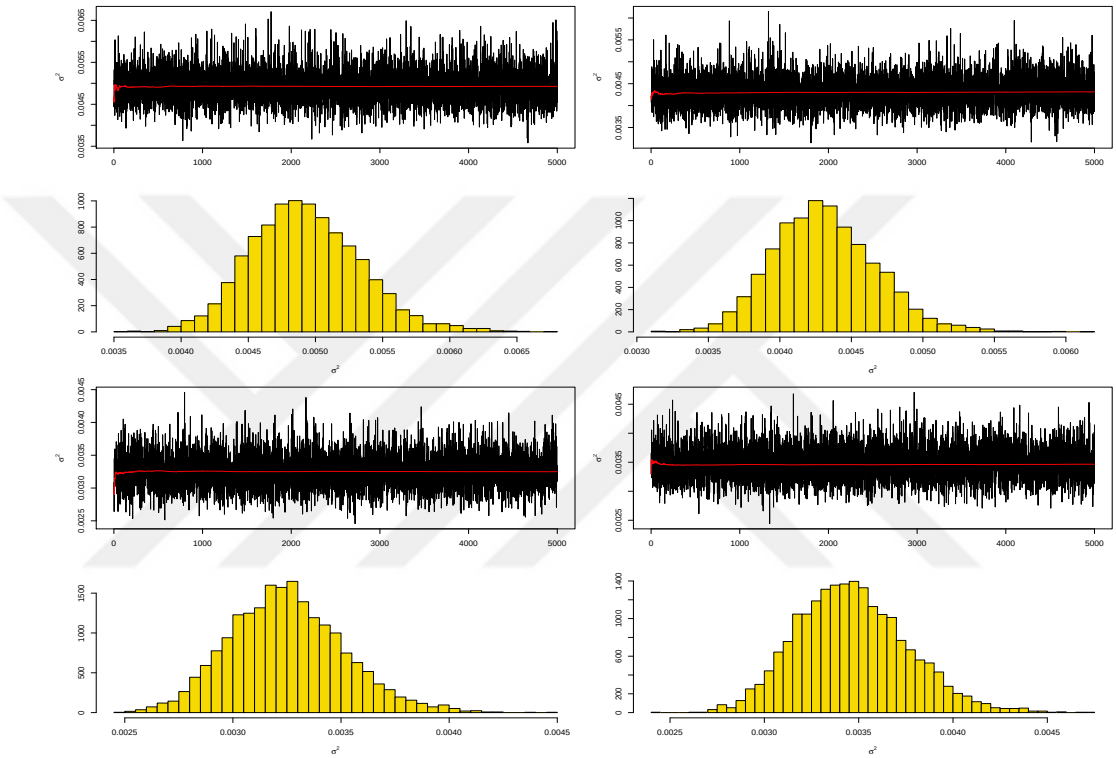


Figure 6.3 Estimated Variance of $\hat{A}_{Bayesian}$ with running mean (red line) and corresponding histogram based on 1000 random samples of size 300 from Asymmetric Logistic ($\alpha = 0.35$, $\beta = 0.75$, $\tau = 0.15$) (top left), Galambos ($\tau = 0.5$) (top right), Logistic ($\tau = 0.5$) (bottom left), Extreme-value t distributions ($\tau = 0.5$ (bottom right).

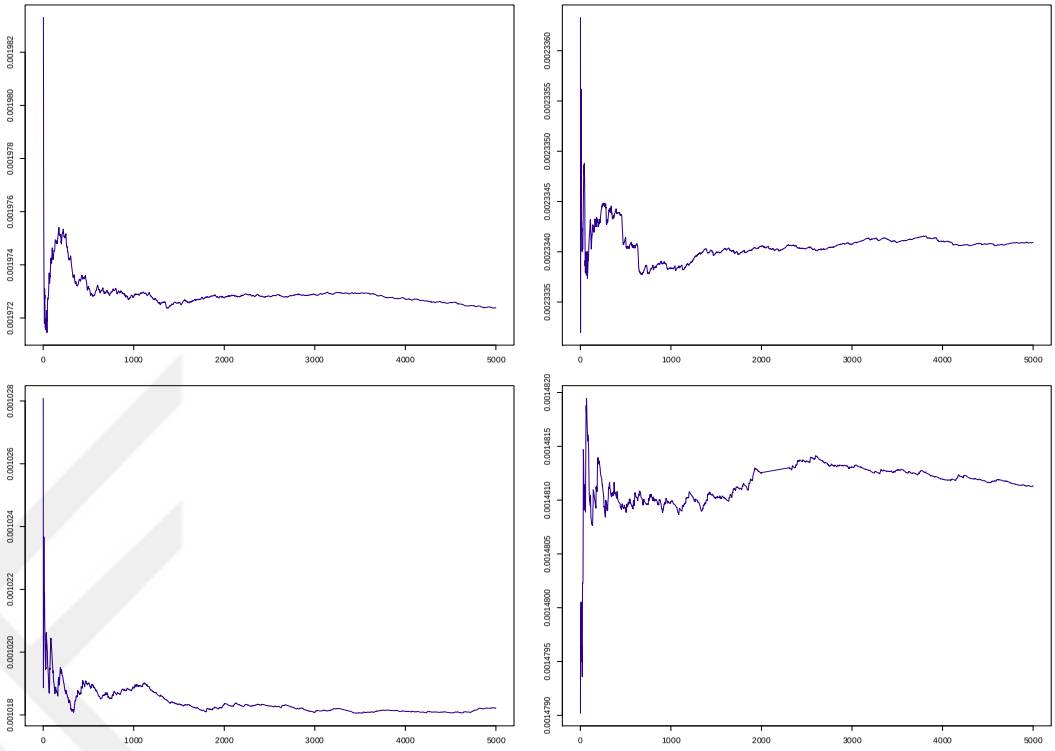


Figure 6.4 Sequence of MSE's of $\hat{A}_{Bayesian}$ against iterations based on 1000 random samples of size 300 from Asymmetric Logistic ($\alpha = 0.35$, $\beta = 0.75$, $\tau = 0.15$) (top left), Galambos ($\tau = 0.5$) (top right), Logistic ($\tau = 0.5$) (bottom left), Extreme-value t distributions ($\tau = 0.5$) (bottom right).

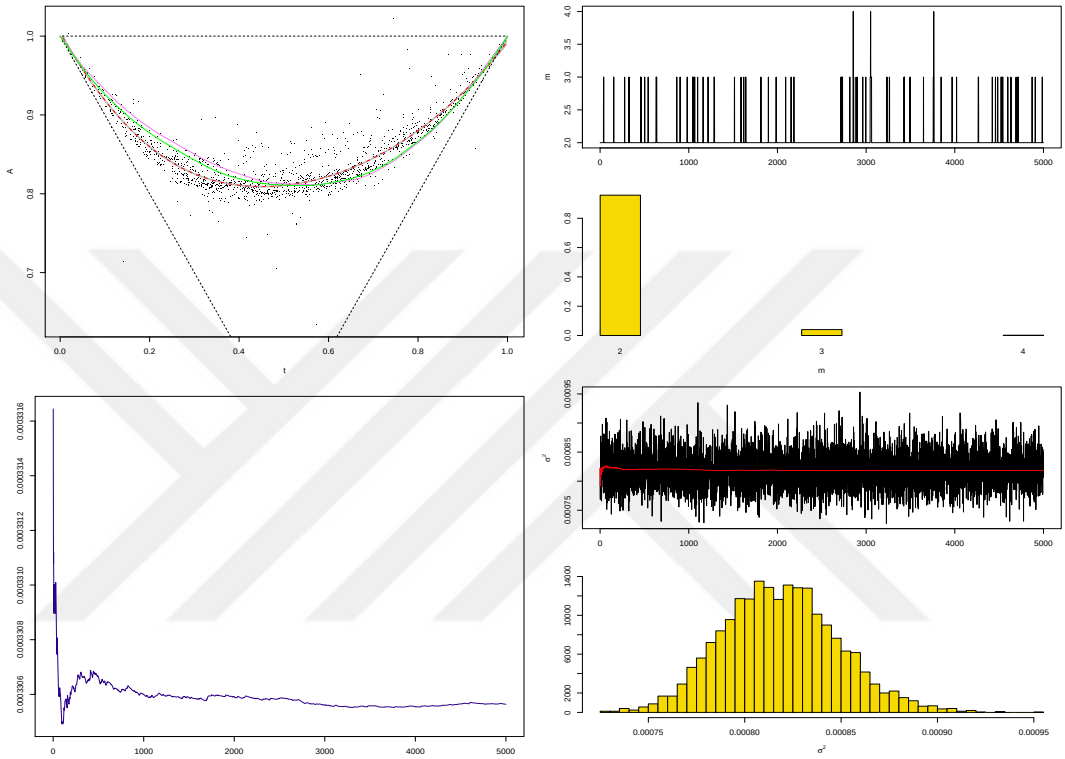


Figure 6.5 Real data analysis: Plots with three estimates of Pickands dependence function A (red); $\hat{A}_{CFG}$ (green), $\hat{A}_{PIC}$ (pink), $\hat{A}_{Bayesian}$ (red dashed line) and gray area is a 95% credible interval for the estimated curve (top left), sequence of m's of $\hat{A}_{Bayesian}$ and the corresponding approximated probability densities (top right), Estimated Variance of $\hat{A}_{Bayesian}$ with running mean (red line) and corresponding histogram (bottom left), Sequence of MSE's of $\hat{A}_{Bayesian}$ against iterations (bottom right).

CHAPTER SEVEN
CONCLUSIONS

A pointwise estimation process is suggested to estimate the Pickands dependence function using Bernstein copula approach. The estimation method was presented for the first time by Cormier, Genest and Neslehova (2014). The points that they have used to estimate the Pickands dependence function is calculated by empirical copula while the points we obtain is calculated by empirical Bernstein copula. Since the Bernstein polynomials have flexible form, the points obtained by our approximation are closer to the true dependence function. In order to evaluate visually whether the random pair has an extreme value dependence structure or not, a plot is developed. Through the plots, the proposed Pickands estimator and the usual Pickands empirical estimator are compared. For the low level of dependence and polynomial degree, the proposed estimator has a better result. Also when the degree of dependence is increased, then the polynomial degree should be increased for a better fit. MSEs of the pairs based on usual empirical copula (T,Z) and our new pairs based on empirical Bernstein copula (T,Q) based on 1000 sample size of 250 from five well known extreme value distributions are given. It is presented that the MSE's of the proposed estimator are smaller than the usual one, especially for small values of τ and low degrees of Bernstein polynomials. A basic measure of the departure of the points from the Pickands estimator of emprical Bernstein copula and the emprical copula is obtained. A significant comparison is also given with test of extremeness through a simulation study. MSE's of our pairs (T,Q) are significantly smaller than the pairs (T,Z). A better performance is observed for the proposed estimator because of flexible form of the Bernstein copula.

A Kernel based estimation process is suggested to estimate the Pickands dependence function using Bernstein copula approach. Almost all the constraints of Pickands dependence function are satisfied through estimation process. After constructing estimator $\hat{A}(t)$ of A that is not necessarily convex, we use *Convhull* function in *R* to reconstruct a bounded convex estimator. Through the plots, the proposed Pickands estimator and the other estimators are compared. Based on the properties of the Bernstein polynomials such as shape preserving, flexibility and etc., our estimator is closer to the true dependence function comparing with Pickands estimator and CFG estimator. For the

low level of dependence and polynomial degree, the proposed estimator has a better result. Also when the degree of dependence is increased, then the polynomial degree should be increased for a better fit. MSEs of $\hat{A}$ with $\hat{A}_{CBK}$, $\hat{A}_{PIC}$ and $\hat{A}_{CFG}$ based on 1000 sample size of 300 from logistic and asymmetric logistic, Negative Logistic, Asymmetric Negative Logistic and Hüsler-Reiss distributions are given. It is presented that the MSE's of the proposed estimator are smaller than the competitors, especially for small values of τ and low levels of Bernstein polynomials. A better performance is observed for the proposed estimator because of flexible form of the Bernstein copula.

A Bayesian estimation process is suggested to estimate the Pickands dependence function for extreme-value copulas through cubic-B-spline regression function. In the Bayesian approach, the explicit form of the likelihood function, the prior and the posterior distributions for vector of parameters are obtained. Since the posterior distribution, $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, does not have an analytical closed form, it is approximated. Through the approximation process, Gibbs sampling method is used to generate random numbers from posterior distribution. Due to the varying dimensionality of the process, move types which we use birth, relocation and death are determined that allows us to freely explore the combined parameter space using Reversible Jump algorithm. The Bayesian method works well because it performs on the possible knot points which are defining functions of our model. A simulation algorithm is presented to measure the performance of the new Bayesian estimator of Pickands dependence function. We consider four families of extreme value distributions; Asymmetric logistic distribution with $\alpha = 0.35$, $\beta = 0.75$ and $\tau = 0.15$, Galambos, Logistic distributions with $\tau = 0.5$ and Extreme value t-student with 4 degrees of freedom and $\tau = 0.5$. In each family, we generate a random sample of size 300 and estimate the Pickands dependence function by running 10000 iterations of algorithm following burn-ins of 5000. Through the simulation study, the proposed Bayesian Pickands estimator and the other estimators are compared. The Bayesian Pickands estimator is closer to the true dependence function comparing with Pickands estimator, CFG estimator. MSEs of $\hat{A}_{Bayesian}$ are getting smaller when the iteration number is increased. As can be observed from estimated variance and corresponding histogram plots the variance assumption is satisfied. Also, The MAP estimation of m is 2. A real-data example which is well-known LOSS/ALAE data is applied to the Bayesian procedure, comparing with the other well-known estimators. Bayesian estimator has a

better result for the data and it also satisfy all the constraints of Pickands dependence function.

REFERENCES

Aguilera, N., Forzani, L., & Morin, P. (2011). On uniform consistent estimators for convex regression. *Journal of Non-Parametric Statistics, 23* (4), 897-908.

Baker, R. (2008). An order-statistics-based method for constructing multivariate distributions with fixed marginals. *Journal of Multivariate Analysis, 99* (10), 2312-2327.

Barber, C. B., Dobkin D. P., & Huhdanpaa H. T. (1996). The Quickhull Algorithm for Convex Hulls. *ACM Transactions on Mathematical Software, 22* (4), 469-483.

Beal, M. J. (2003). *Variational algorithms for approximate Bayesian inference*. Ph.D. Thesis, University College London, London.

Bergahus, B., Bücher, A., & Dette, H. (2012). *Minimum distance estimation of Pickands dependence function for multivariate distributions*, Working Paper.

Bernardo, J., & Smith, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley.

Bernstein, S. N. (1912). Démonstration du théorème de weierstrass fondée sur le calcul dés probabilités. *Kharkov Mathematical Society, 12* (2), 1-2.

Brooks, S. P. (2003). Bayesian computation: A statistical revolution. *Royal Society Publishing, 361* (1813).

Bücher, A., Dette, H., & Volgushev, S. (2011). New estimators of the Pickands dependence function and a test for extreme-value dependence. *Annals of Statistics, 39*, 1963-2006.

Capraà, P., Fougéres, A. L., & Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika, 84*, 567-577.

Chen, M., Shao, Q., & Ibrahim, J. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician, 49*, 327-335.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Springer Series in Statistics.

Cormier, E., Genest, C., & Nevslehova, J. G. (2014). Using B-splines for nonparametric inference on bivariate extreme-value copulas. *Extremes, 17*, 633-659.

Deheuvels, P. (1991). On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statistics and Probability Letters, 12*, 429-439.

Dellaportas, P., Forster, J., & Ntzoufras, I. (2000). Bayesian variable selection using the Gibbs sampler. *Generalized Linear Models: A Bayesian Perspective, Marcel Dekker, New York*, 271-286.

Demarta, S., & McNeil, A. J. (2005). The t copula and related copulas. *International Statistical Review, 73*, 111-129.

Galambos, J. (1975). Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association, 70*, 674-680.

Galambos, J. (1987). *The asymptotic theory of extreme order statistics* (2nd ed.). New York: Wiley.

Gelfand, A., & Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85*, 398-409.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6,* 721-741.

Genest, C., & MacKay, J. (1986a). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canadian Journal of Statistics, 14*, 145-159.

Genest, C., & MacKay, J. (1986b). The joy of copulas: Bivariate distributions with two uniform marginals. *The American Statistical Association, 40,* 280-285.

Geweke, J. F. (1989). Bayesian inference in econometrics using Monte Carlo integration. *Econometrica*, *23,* 156-163.

Gholami, G. H. (2008). *Change-point problems in regression: A Bayesian approach*. Ph.D. Thesis, Paris Dauphine University, Paris.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian determination. *Biometrika, 82,* 711-732.

Grenander, U. (1983). *Tutorials in Pattern synthesis*. Brown University, Division of Applied Mathematics.

Gudendorf, G., & Segers, J. (2011). Nonparametric estimation of an extreme-value copula in arbitrary dimensions. *Journal of Multivariate Analysis, 102,* 37-47.

Gumbel, E. J. (1960). Distributions des valeurs extremes en plusieurs dimensions. *Publications de ´l Instute Statistique de ´l Université de Paris, 9,* 171-173.

Hall, P., & Tajvidi, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli, 6,* 835-844.

Härdle, W. (1990). *Applied nonparametric regression*. Cambridge: Cambridge University Press.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*, 97-109.

Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika*, *73*, 671-678.

Hüsler, J. (1986). Extreme values of non-stationary random sequences. *Journal of Applied Probability, 23*, 937-950.

Janssen, P., Swanepoel, J., & Veraverbeke, N. (2008). Large sample behavior of the Bernstein copula estimator. *Journal of Statistical Planing and Inference*, *142* (5), 1189-1197.

Joe, H. (1990). Multivariate concordance. *Journal of Multivariate Analysis*, *35*, 12-30.

Li, X., Mikusinski, P., & Taylor, M. D. (1997). On approximation of copulas. *Journal of Mathematical Analysis and Application*, *225*, 608-623.

Marcon, G., Padoan, S. A., Naveau, P., & Muliere, P. (2014). Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials. *arXiv Areprint*, arXiv: *1405* (5228).

Metropolis, H., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, *21*, 1087.

Meyn, S., & Tweedie, R. (1993). *Markov chains and stochastic stability*. New York: Springer.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications, 9,* 141-142.

Nelsen, R. B. (2006). *An introduction to copulas* (2nd ed.). New York: Springer.

Nikoloulopoulos, A. K., Joe, H., & Li, H. (2009). Extreme value properties of multivariate t copulas. *Extremes*, *12* (2), 129-148.

Phillips, G. M. (2003). *Interpolation and approximation by polynomials*. New York: Springer.

Pickands, J. (1981). Multivariate extreme value distribution. *In Proceedings of the 43rd Session of the I.S.I. International Statistical Institute*, 859-878.

Reiss, R. D. (1989). *Approximate distributions of order statistics: With applications to nonparametric statistics*. New York: Springer Series in Statistics.

Robert, C. (2007). *The Bayesian Choice* (2nd ed.). New York: Springer.

Robert, C., & Casella, G. (2004). *Monte Carlo Statistical Methods*. New York: Springer.

Sancetta, A., & Satchell, S. (2004). The Bernstein copula and its application to modeling and approximations of multivariate distributions. *Econometric Theory*, (3), 535-562.

Schervish, M. J., & Carlin, B. P. (1992). On the convergence of successive substitution sampling. *Journal of Computational and Graphical Statistics, 1*, 111-127.

Segers, J. (2007). *Non-parametric inference for bivariate extreme-value copulas*. In: Extreme Value Distributions (M. Ahsanullah and S. Kirmani, eds.), Nova Science Publishers, 181-203.

Sklar, A. (1959). Fonctions de répartition àn dimensions et leurs marges. *Publications de ´l Instute Statistique de ´l Université de Paris, 8*, 229-231.

Tanner, M., & Wong, W. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association, 82*, 528-549.

Tawn, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika, 75*, 397-415.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics, 22*, 1701-1762.

Watson, G. S. (1964). Smooth regression analysis. *Sankhya, the Indian Journal of Statistics, 26*, 359-372.

Zhang, D., Wells, M. T., & Peng, L. (2008). Nonparametric estimation of the dependence function for a multivariate extreme value distribution. *Journal of Multivariate Analysis, 99*, 577-588.