

## ABSTRACT

ALMELOR, GLENN. Bayesian Inference for Extreme Values. (Under the direction of Dr. Subhashis Ghoshal).

Extreme value theory (EVT) provides a framework for analyzing extreme values in various fields of application such as insurance, engineering, finance, among others. One classical problem in EVT is the estimation of the extreme value index (EVI), whose sign enables classification of the underlying probability distributions into three so-called maximum domains of attraction (MDA) - the Fréchet, reverse Weibull, and Gumbel distribution families. The Hill and the Pickands estimators for the EVI are well-known frequentist estimators in the EVT literature. We propose a Bayesian method for inference on the EVI, which is approximated by a sequence of functionals of the underlying distribution. The constructed functionals are different depending on whether the EVI is positive, negative, or zero. For the non-zero EVI case, a Dirichlet process prior on the distribution of the observations is assigned and the induced posterior distribution of the sequence of functionals is utilized to infer the EVI. A Bernstein-von Mises Theorem for the posterior distribution of the EVI is established that provides a posterior contraction rate and asymptotic frequentist coverage of a Bayesian  $(1 - \alpha)$ -credible interval. For the zero EVI case, a test based on a modified Bayes factor is designed to test the hypothesis that the EVI is zero. A power of the empirical likelihood is used and the marginal likelihood under the null hypothesis is computed by the Hamiltonian Monte Carlo method. A comprehensive simulation study demonstrates the effectiveness of the proposed Bayesian methods.

Within the Gumbel MDA lie an important class of the Weibull-type distributions that includes well-known families of distributions, such as the normal, gamma, logistic, and Weibull distributions. A defining feature of this class is the Weibull tail coefficient (WTC) which measures the decay rate of the associated survival function to 0. Empirical estimators cannot be directly used to reliably estimate the Weibull tail coefficient due to insufficient observations in the extreme tail. Familiar estimators, such as the ones proposed by Girard and Beirlant, are based on the increments of empirical quantiles at designated levels. Similar to the approach for the EVI inference, a proposed Bayesian approach infers the WTC by approximating it by a sequence of functionals of the underlying distribution, to which a Dirichlet process prior is assigned. We likewise obtain the posterior contraction rate and derive a Bernstein-von Mises theorem to establish that a Bayesian  $(1 - \alpha)$ -credible interval for the Weibull tail coefficient has coverage approximately equal to  $1 - \alpha$ . The Bayesian bootstrap distribution closely approximates the Dirichlet process posterior, enabling the faster implementation of

the proposed method. An extensive simulation study is conducted to assess the finite-sample accuracy of the proposed method and compare its performance with that of the frequentist approaches.

Extreme quantiles in a regression setting are notoriously difficult to estimate due to data sparsity in the extremes. In this situation, the usual conditional quantile estimator is often unreliable at extreme quantile levels. He et al. (2020) designed an approach that obtains a conditional quantile estimator at an extreme quantile level by scaling up the conditional quantile estimator at an intermediate quantile level, using an estimator for the WTC as an exponent to the scaling factor. Utilizing the same scaling idea, we propose a Bayesian approach by leveraging the misspecified asymmetric Laplace density (ALD) posterior to infer the intermediate quantile and combining with the earlier Bayesian approach to infer the WTC  $\theta$ . The frequentist and Bayesian approaches are then compared using mean square error and coverage metrics. Additionally, the conditional tail expectation at the extreme level are also calculated and compared under both approaches.

© Copyright 2026 by Glenn Almelor

All Rights Reserved

Bayesian Inference for Extreme Values

by  
Glenn Almelor

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

Operations Research

Raleigh, North Carolina  
2026

APPROVED BY:

---

Dr. Min Kang

---

Dr. Ralph Smith

---

Dr. Brian Reich

---

Dr. Subhashis Ghoshal  
Chair of Advisory Committee

## DEDICATION

I dedicate this dissertation to my wife Marie and my daughter Vanessa for their patience and support, as well as their role as an inspiration in completing my research.

## **BIOGRAPHY**

Glenn Almelor obtained his Bachelor of Science in Mathematics (Honors Program) from the Ateneo de Manila University in Quezon City, Philippines where he was born. When he embarked on his actuarial career, he attained the Fellow of the Society of Actuaries and Chartered Financial Analyst designations. He also has earned Master of Science degrees in Mathematics from the University of Minnesota at Twin Cities and Mathematical Finance from the University of North Carolina at Charlotte.

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Ghoshal not only for his depth of knowledge and guidance, but also for the exceptional patience he demonstrated throughout our collaboration. His support was particularly invaluable given that I conducted this research on a part-time basis.

I would also like to express my sincere gratitude to Drs. Kang, Reich, and Smith for their valuable feedback and constructive insights, which significantly strengthened the overall quality of this dissertation.

# TABLE OF CONTENTS

<b>List of Tables</b> . . . . .	<b>vii</b>
<b>List of Figures</b> . . . . .	<b>ix</b>
<b>Authorship Statement</b> . . . . .	<b>xiii</b>
<b>Chapter 1 INTRODUCTION</b> . . . . .	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Extreme Value Theory . . . . .	6
1.3 Bayesian Nonparametrics . . . . .	8
1.4 Strong Approximations . . . . .	10
1.5 Quantile Regression . . . . .	11
<b>Chapter 2 BAYESIAN INFERENCE ON THE EXTREME VALUE INDEX</b>	<b>13</b>
2.1 Introduction . . . . .	13
2.2 Setup and Methodology . . . . .	14
2.2.1 $EVI \neq 0$ . . . . .	15
2.2.2 $EVI = 0$ . . . . .	16
2.3 Convergence Results . . . . .	17
2.4 Numerical Results . . . . .	23
2.5 Proofs . . . . .	30
<b>Chapter 3 BAYESIAN INFERENCE ON THE WEIBULL TAIL COEFFICIENT</b>	<b>37</b>
3.1 Introduction . . . . .	37
3.2 Setup and Methodology . . . . .	38
3.3 Convergence Results . . . . .	40
3.4 Numerical Results . . . . .	44
3.5 Proofs . . . . .	49
<b>Chapter 4 EXTREME QUANTILE ESTIMATION FOR WEIBULL-TYPE DISTRIBUTIONS</b>	<b>58</b>
4.1 Introduction . . . . .	58
4.2 Setup and Methodology . . . . .	59
4.2.1 He-Wang-Tong (HWT) Approach . . . . .	59
4.2.2 Bayesian Approach . . . . .	61
4.2.3 Conditional Tail Expectation . . . . .	62
4.3 Numerical Results . . . . .	63
4.3.1 Adaptive Metropolis-Hastings (AMH) Method . . . . .	63
4.3.2 bayesQR (BQR) Method . . . . .	64
4.3.3 Simulated Dataset Results . . . . .	64
4.3.4 Real Life Dataset Results . . . . .	78

<b>Chapter 5 CONCLUSION</b> . . . . .	<b>79</b>
5.1 Summary . . . . .	79
5.2 Further Research . . . . .	81
<b>References</b> . . . . .	<b>82</b>
<b>APPENDICES</b> . . . . .	<b>86</b>
Appendix A Chapter 4 Supplementary Material - Data Model 1 . . . . .	87
A.1 Data Model 1 - Normal . . . . .	87
A.2 Data Model 1 - Logistic . . . . .	92
A.3 Data Model 1 - Weibull . . . . .	96
Appendix B Chapter 4 Supplementary Material - Data Model 2 . . . . .	108
B.1 Data Model 2 - Beta . . . . .	108
B.2 Data Model 2 - Trig . . . . .	119
B.3 Data Model 2 - Poly . . . . .	131

## LIST OF TABLES

Table 2.1	Selected EVI inference simulation results for Pareto distribution with parameters $(t, \alpha)$ . . . . .	24
Table 2.2	Selected EVI inference simulation results for Student's $t$ -distribution with $\nu$ degrees of freedom. . . . .	25
Table 2.3	Selected EVI inference simulation results for Fréchet Distribution with parameters (loc, scale, shape). . . . .	25
Table 2.4	Selected EVI inference simulation results for Snedecor's $F$ -distribution with $\nu_1, \nu_2$ degrees of freedom. . . . .	26
Table 2.5	Selected EVI inference simulation results for uniform distribution on $[a, b], a < b$ . . . . .	26
Table 2.6	Selected EVI inference simulation results for beta distribution with parameters $(\alpha, \beta)$ . . . . .	27
Table 2.7	Selected EVI inference simulation results for triangular distribution with parameters $a < b < c$ . . . . .	27
Table 3.1	Selected WTC inference simulation results for a Weibull distribution with shape $\alpha$ . . . . .	46
Table 3.2	Selected WTC inference simulation results for an extended Weibull distribution with $\mu = 1, \nu = 0.9$ and shape $\sigma$ . . . . .	47
Table 3.3	Selected WTC inference simulation results for a generalized power Weibull distribution with $\nu = 0.9, \sigma = 1$ , and shape $\sigma$ . . . . .	48
Table 3.4	Selected WTC inference simulation results for the standard logistic distribution. . . . .	48
Table 3.5	Selected WTC inference simulation results for the standard normal distribution. . . . .	48
Table 3.6	Selected WTC inference simulation results for a gamma distribution with scale $\theta = 1$ and shape $\alpha$ . . . . .	49
Table 4.1	Coverage and length results on WTC for simulations from DM 1 - Gamma, with $\theta = 1.0$ using the HWT and AMH methods under quantile set A. . . . .	68
Table 4.2	Coverage and length results on extreme quantile and CTE for simulations under DM 1 - Gamma, with $\theta = 1.0$ using the HWT and AMH methods under quantile set A. . . . .	71
Table 4.3	Coverage and length results on WTC for simulations from DM 1 - Gamma, with $\theta = 1.0$ using the HWT and AMH methods under quantile set B. . . . .	71
Table 4.4	Coverage and length results on extreme quantile and CTE for simulations under DM 1 - Gamma, with $\theta = 1.0$ using the HWT and AMH methods under quantile set B. . . . .	74
Table 4.5	Coverage and length results on WTC for simulations from DM 1 - Beta, with $\theta = 0.5$ using the HWT and AMH methods under quantile set A. . . . .	74

Table 4.6	Coverage and length results on extreme quantile and CTE for simulations under DM 1 - Beta, with $\theta = 0.5$ using the HWT and AMH methods under quantile set A. . . . .	77
-----------	--	----

## LIST OF FIGURES

Figure 2.1	Modified Bayes factors for various distributions using training and validation data, $n = 50$ , $B = 100$ . . . . .	29
Figure 4.1	Extreme quantile and CTE plots for a sample simulation from DM 1 - Gamma, with $\theta = 1.0$ using the HWT and AMH methods under quantile set A. . . . .	68
Figure 4.2	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Gamma, with $\theta = 1.0$ using the HWT and AMH methods under quantile set A. . . . .	69
Figure 4.3	RMSE and bias box plots on WTC for simulations from DM 1 - Gamma, with $\theta = 1.0$ using the HWT and AMH methods under quantile set A. . . . .	70
Figure 4.4	Extreme quantile and CTE plots for a sample simulation from DM 1 - Gamma, with $\theta = 1.0$ using the HWT and AMH methods under quantile set B. . . . .	70
Figure 4.5	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Gamma, with $\theta = 1.0$ using the HWT and AMH methods under quantile set B. . . . .	72
Figure 4.6	RMSE and bias box plots on WTC for simulations from DM 1 - Gamma, with $\theta = 1.0$ using the HWT and AMH methods under quantile set B. . . . .	73
Figure 4.7	Extreme quantile and CTE plots for a sample simulation from DM 1 - Beta, with $\theta = 0.5$ using the HWT and AMH methods under quantile set A. . . . .	73
Figure 4.8	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Beta, with $\theta = 0.5$ using the HWT and AMH methods under quantile set A. . . . .	75
Figure 4.9	RMSE and bias box plots on WTC for simulations from DM 1 - Beta, with $\theta = 0.5$ using the HWT and AMH methods under quantile set A. . . . .	76
Figure 4.10	Extreme quantile and CTE of diamond price as function of carat under the HWT and AMH methods. . . . .	78
Figure A.1	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Normal, with $\theta = 0.5$ using the HWT and AMH methods under quantile set A. . . . .	88
Figure A.2	RMSE and bias box plots on WTC for simulations from DM 1 - Normal, with $\theta = 0.5$ using the HWT and AMH methods under quantile set A. . . . .	89
Figure A.3	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Normal, with $\theta = 0.5$ using the HWT and AMH methods under quantile set B. . . . .	90
Figure A.4	RMSE and bias box plots on WTC for simulations from DM 1 - Normal, with $\theta = 0.5$ using the HWT and AMH methods under quantile set B. . . . .	91
Figure A.5	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Logistic, with $\theta = 1$ using the HWT and AMH methods under quantile set A. . . . .	92

Figure A.6	RMSE and bias box plots on WTC for simulations from DM 1 - Logistic, with $\theta = 1$ using the HWT and AMH methods under quantile set A. .	93
Figure A.7	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Logistic, with $\theta = 1$ using the HWT and AMH methods under quantile set B. . . . .	94
Figure A.8	RMSE and bias box plots on WTC for simulations from DM 1 - Logistic, with $\theta = 1$ using the HWT and AMH methods under quantile set B. .	95
Figure A.9	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Weibull, with $\theta = 0.5$ using the HWT and AMH methods under quantile set A. . . . .	96
Figure A.10	RMSE and bias box plots on WTC for simulations from DM 1 - Weibull, with $\theta = 0.5$ using the HWT and AMH methods under quantile set A. .	97
Figure A.11	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Weibull, with $\theta = 0.5$ using the HWT and AMH methods under quantile set B. . . . .	98
Figure A.12	RMSE and bias box plots on WTC for simulations from DM 1 - Weibull, with $\theta = 0.5$ using the HWT and AMH methods under quantile set B. .	99
Figure A.13	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Weibull, with $\theta = 1$ using the HWT and AMH methods under quantile set A. . . . .	100
Figure A.14	RMSE and bias box plots on WTC for simulations from DM 1 - Weibull, with $\theta = 1$ using the HWT and AMH methods under quantile set A. .	101
Figure A.15	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Weibull, with $\theta = 1$ using the HWT and AMH methods under quantile set B. . . . .	102
Figure A.16	RMSE and bias box plots on WTC for simulations from DM 1 - Weibull, with $\theta = 1$ using the HWT and AMH methods under quantile set B. .	103
Figure A.17	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Weibull, with $\theta = 1.5$ using the HWT and AMH methods under quantile set A. . . . .	104
Figure A.18	RMSE and bias box plots on WTC for simulations from DM 1 - Weibull, with $\theta = 1.5$ using the HWT and AMH methods under quantile set A. .	105
Figure A.19	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Weibull, with $\theta = 1.5$ using the HWT and AMH methods under quantile set B. . . . .	106
Figure A.20	RMSE and bias box plots on WTC for simulations from DM 1 - Weibull, with $\theta = 1.5$ using the HWT and AMH methods under quantile set B. .	107
Figure B.1	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Beta, with $\theta = 0.5$ using the HWT and AMH methods under quantile set B. . . . .	109
Figure B.2	RMSE and bias box plots on WTC for simulations from DM 2 - Beta, with $\theta = 0.5$ using the HWT and AMH methods under quantile set B. .	110

Figure B.3	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Beta, with $\theta = 1$ using the HWT and AMH methods under quantile set A. . . . .	111
Figure B.4	RMSE and bias box plots on WTC for simulations from DM 2 - Beta, with $\theta = 1$ using the HWT and AMH methods under quantile set A. . . . .	112
Figure B.5	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Beta, with $\theta = 1$ using the HWT and AMH methods under quantile set B. . . . .	113
Figure B.6	RMSE and bias box plots on WTC for simulations from DM 2 - Beta, with $\theta = 1$ using the HWT and AMH methods under quantile set B. . . . .	114
Figure B.7	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Beta, with $\theta = 1.5$ using the HWT and AMH methods under quantile set A. . . . .	115
Figure B.8	RMSE and bias box plots on WTC for simulations from DM 2 - Beta, with $\theta = 1.5$ using the HWT and AMH methods under quantile set A. . . . .	116
Figure B.9	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Beta, with $\theta = 1.5$ using the HWT and AMH methods under quantile set B. . . . .	117
Figure B.10	RMSE and bias box plots on WTC for simulations from DM 2 - Beta, with $\theta = 1.5$ using the HWT and AMH methods under quantile set B. . . . .	118
Figure B.11	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Trig, with $\theta = 0.5$ using the HWT and AMH methods under quantile set A. . . . .	119
Figure B.12	RMSE and bias box plots on WTC for simulations from DM 2 - Trig, with $\theta = 0.5$ using the HWT and AMH methods under quantile set A. . . . .	120
Figure B.13	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Trig, with $\theta = 0.5$ using the HWT and AMH methods under quantile set B. . . . .	121
Figure B.14	RMSE and bias box plots on WTC for simulations from DM 2 - Trig, with $\theta = 0.5$ using the HWT and AMH methods under quantile set B. . . . .	122
Figure B.15	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Trig, with $\theta = 1$ using the HWT and AMH methods under quantile set A. . . . .	123
Figure B.16	RMSE and bias box plots on WTC for simulations from DM 2 - Trig, with $\theta = 1$ using the HWT and AMH methods under quantile set A. . . . .	124
Figure B.17	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Trig, with $\theta = 1$ using the HWT and AMH methods under quantile set B. . . . .	125
Figure B.18	RMSE and bias box plots on WTC for simulations from DM 2 - Trig, with $\theta = 1$ using the HWT and AMH methods under quantile set B. . . . .	126
Figure B.19	RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Trig, with $\theta = 1.5$ using the HWT and AMH methods under quantile set A. . . . .	127
Figure B.20	RMSE and bias box plots on WTC for simulations from DM 2 - Trig, with $\theta = 1.5$ using the HWT and AMH methods under quantile set A. . . . .	128

Figure B.21 RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Trig, with $\theta = 1.5$ using the HWT and AMH methods under quantile set B. . . . .	129
Figure B.22 RMSE and bias box plots on WTC for simulations from DM 2 - Trig, with $\theta = 1.5$ using the HWT and AMH methods under quantile set B. . . . .	130
Figure B.23 RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Poly, with $\theta = 0.5$ using the HWT and AMH methods under quantile set A. . . . .	131
Figure B.24 RMSE and bias box plots on WTC for simulations from DM 2 - Poly, with $\theta = 0.5$ using the HWT and AMH methods under quantile set A. . . . .	132
Figure B.25 RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Poly, with $\theta = 0.5$ using the HWT and AMH methods under quantile set B. . . . .	133
Figure B.26 RMSE and bias box plots on WTC for simulations from DM 2 - Poly, with $\theta = 0.5$ using the HWT and AMH methods under quantile set B. . . . .	134
Figure B.27 RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Poly, with $\theta = 1$ using the HWT and AMH methods under quantile set A. . . . .	135
Figure B.28 RMSE and bias box plots on WTC for simulations from DM 2 - Poly, with $\theta = 1$ using the HWT and AMH methods under quantile set A. . . . .	136
Figure B.29 RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Poly, with $\theta = 1$ using the HWT and AMH methods under quantile set B. . . . .	137
Figure B.30 RMSE and bias box plots on WTC for simulations from DM 2 - Poly, with $\theta = 1$ using the HWT and AMH methods under quantile set B. . . . .	138
Figure B.31 RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Poly, with $\theta = 1.5$ using the HWT and AMH methods under quantile set A. . . . .	139
Figure B.32 RMSE and bias box plots on WTC for simulations from DM 2 - Poly, with $\theta = 1.5$ using the HWT and AMH methods under quantile set A. . . . .	140
Figure B.33 RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Poly, with $\theta = 1.5$ using the HWT and AMH methods under quantile set B. . . . .	141
Figure B.34 RMSE and bias box plots on WTC for simulations from DM 2 - Poly, with $\theta = 1.5$ using the HWT and AMH methods under quantile set B. . . . .	142

# AUTHORSHIP STATEMENT

Contributions of Glenn Almelor and coauthors are listed below for each chapter.

## **Chapter 1**

1. Glenn Almelor: performed literature review; wrote the draft.
2. Dr. Subhashis Ghoshal: proposed the agenda and general approach; reviewed and edited the draft.
3. Dr. Ralph Smith: recommended changes to address typographical errors.

## **Chapter 2**

1. Glenn Almelor: performed literature review; carried out the proofs; coded, ran, and summarized numerical experiments; wrote the draft.
2. Dr. Subhashis Ghoshal: proposed the agenda and general approach, including use of Hamiltonian Monte Carlo (HMC) method and modified Bayes factors; reviewed and edited the draft.

## **Chapter 3**

1. Glenn Almelor: performed literature review; carried out the proofs, coded, ran, and summarized numerical experiments; wrote the draft.
2. Dr. Subhashis Ghoshal: proposed the agenda and general approach, including use of strong approximations, Dirichlet process priors, and Bayesian bootstrap; reviewed and edited the draft.

## **Chapter 4**

1. Glenn Almelor: performed literature review; carried out the proofs; coded, ran, and summarized numerical experiments; wrote the draft.
2. Dr. Subhashis Ghoshal: proposed the agenda and general approach, including quantile regression model assumptions and conditional tail expectation (CTE) approximations; reviewed and edited the draft.

## **Chapter 5**

1. Glenn Almelor: wrote the draft.

2. Dr. Subhashis Ghoshal: recommended topics to include in Further Research subsection; reviewed and edited the draft.

**Use of generative artificial intelligence:** No generative artificial intelligence was used to write this dissertation.

# CHAPTER

## 1

# INTRODUCTION

## 1.1 Introduction

The study of extreme values of stochastic events is an essential endeavor with practical consequences. For example, insurance companies have a significant interest in estimating extremely large (catastrophic) insurance claim amounts arising from natural disasters (Embrechts et al. 1997, 1999). In the Netherlands, the Dutch government builds and maintains dikes high enough so that a water breach is sufficiently improbable (de Haan and Ferreira 2006). Testing the maximum pesticide content across a series of samples is used to determine whether safety and health standards are met (Asmussen and Steffensen 2020). Extreme value theory (EVT) provides a framework for addressing these and other problems, as covered in classic textbooks such as Leadbetter et al. (1983), Galambos (1987), and de Haan and Ferreira (2006), all of which provide comprehensive introductions to EVT.

In a one-dimensional setting, EVT usually starts with a random sample  $X_1, \dots, X_n$  of independent and identically distributed random variables from some unknown cdf  $F$ , with a goal of finding the distribution of either the maximum or minimum of a random sample. While the distribution of the max of the random sample,  $[F(x)]^n$ , converges to a degenerate distribution as  $n \rightarrow \infty$ , a judicious choice of scalars  $a_n > 0$  and  $b_n \in \mathbb{R}$  may allow  $[F(a_n x + b_n)]^n$  to converge to a nondegenerate limit distribution  $G(x)$ , which is called an

extreme value (EV) distribution. Fisher and Tippett (1928) and Gnedenko (1943) have shown that the EV distribution has the form  $G_\gamma(ax + b)$  with  $a > 0$ ,  $b \in \mathbb{R}$ , and where

$$G_\gamma(x) = \exp[-(1 + \gamma x)^{-1/\gamma}], \quad 1 + \gamma x > 0. \quad (1.1)$$

Furthermore, the sign of the extreme value index  $\gamma$  characterizes the (non-degenerate) EV distributions into three classes. If  $\gamma > 0$ , then  $F$  is said to be in the maximum domain of attraction (MDA) of the Fréchet distribution family. If  $\gamma < 0$ , then  $F$  is in the MDA of the negative Weibull distribution family. Finally, if  $\gamma = 0$ , then  $F$  is in the MDA of the Gumbel distribution.

Parameter estimation of  $\gamma$  has been extensively studied, with well-known estimators including the Hill estimator (Hill 1975) and the Pickands estimator (Pickands III 1975). de Haan and Ferreira (2006) describes MLE, moment, and probability-weighted moment estimators for  $\gamma$ .

From a Bayesian perspective, Tokdar et al. (2022) proposed a semiparametric approach based on the generalized Pareto form of tilt, assigned a Gaussian process-based prior for the probability density of the central factor, and derived the posterior contraction rates for  $\gamma$  and the whole density. Padoan and Rizzelli (2022) proposed a semiparametric Bayesian framework for multivariate max-stable distributions to establish posterior consistency. An empirical Bayes procedure for inference on the block maxima law was proposed, and consistency and asymptotic normality of the posterior distributions of the EVI were established in Padoan and Rizzelli (2024).

We propose a Bayesian approach to estimate the EVI  $\gamma$  without making a specific parametric assumption on  $F$  with EVI  $\gamma$ . When  $\gamma$  is positive, we approximate  $\gamma$  by a sequence of functional  $\psi_n(F)$ . We assign the familiar Dirichlet process prior on  $F$  and make an inference on  $\gamma$  by the induced posterior distribution of  $\psi_n(F)$ . Monte Carlo sampling is used to compute posterior characteristics. The resulting posterior distribution automatically quantifies the parameter's uncertainty; i.e., we obtain an estimate along with credible intervals that capture the spread of the posterior. Using the strong approximation technique, we establish a Bernstein-von Mises theorem (Van Der Vaart 1998) for the posterior distribution of  $\gamma$ . This theorem yields a contraction rate for the posterior distribution of  $\gamma$  at its true value; the posterior distribution of  $\gamma$  is almost surely approximately normal with center at  $\psi_n(\mathbb{F}_n)$ , while the sampling theoretic distribution of  $\psi_n(\mathbb{F}_n)$  is also asymptotically normal with center at  $(1+1/\gamma)/2$ , and the two scales are the same. Consequently, the asymptotic frequentist coverage of a credible interval matches its credibility. Thus, the Bayesian uncertainty quantification is justified by the frequentist yardstick.

The case  $\gamma < 0$  is handled by the same approach as that for  $\gamma > 0$ , with the main difference being that a different sequence of functionals  $\zeta_n(\cdot)$  is used for approximations instead of  $\psi_n(\cdot)$ .

For the zero EVI case, we begin with Theorem 1.2.2 in de Haan and Ferreira (2006) for a characterization that suggests a test of the hypothesis  $\gamma = 0$ . A Bayes factor, the original idea of which is proposed in Jeffreys (1935) and further covered in Kass and Raftery (1995), is constructed to allow for comparison of models  $M_0$ , having a uniform prior on the simplex, and  $M_1$ , which includes an additional constraint that's aligned with the defining characterization. However, the nonparametric models with the Dirichlet process prior are undominated, so a traditional Bayes factor does not exist. To address the problem, the Bayes factor is constructed based on the empirical likelihood (Owen 1988) with weights at the observations. Posterior samples from the restricted model  $M_1$  can be obtained by the Hamiltonian Monte Carlo method (Neal 2012).

One of the primary challenges in extreme value modeling is the scarcity of data in the extremes. Hence, extrapolation from observed data is necessary (Coles 2001). Asymptotic arguments are typically used to support EV models, and such arguments will also be presented in this paper.

Within the MDA of the Gumbel distribution, a distribution  $F$  with the property that for some  $\theta > 0$ ,

$$\lim_{z \rightarrow \infty} \frac{\log(1/\bar{F}(\zeta z))}{\log(1/\bar{F}(z))} = \zeta^{1/\theta} \quad \text{for all } \zeta > 0, \quad (1.2)$$

is said to have a Weibull-type tail, and  $\theta$  is called the Weibull tail coefficient (WTC) of  $F$ . Familiar distributions such as the normal, gamma, logistic, and Weibull distributions all have Weibull-type tails. The WTC is scale-invariant; it assumes the same value in a scale family. Furthermore, the WTC measures the decay rate of  $\bar{F}$ , so that higher values of  $\theta$  imply slower decay of the survival function.

Thus far, estimation methods for the WTC proposed in the literature are based on the empirical distribution function  $\mathbb{F}_n$ . Broniatowski (1993) estimated the Weibull tail coefficient based on a weighted average of the  $k_n$  upper order statistics of the random sample, where  $1 \leq k_n \leq n$  is such that  $k_n/n \rightarrow 0$  as  $n \rightarrow \infty$ . Beirlant et al. (1995) developed an estimator based on the mean residual function  $E(X - x|X > x)$ . Girard (2004) suggested a Hill-type estimator (Hill 1975) for the tail coefficient. He et al. (2020) devised an estimator for the Weibull tail coefficient in a regression setting. Additional estimators were proposed in Gardes and Girard (2006, 2008, 2016).

Similar to the approach taken in estimating the EVI  $\gamma$ , a Bayesian approach is employed to

estimate the WTC  $\theta$  without making a specific parametric assumption on  $F$ . We approximate  $\theta$  by a sequence of functional  $\psi_n(F)$ . We assign the familiar Dirichlet process prior on  $F$  and make an inference on  $\theta$  by the induced posterior distribution of  $\psi_n(F)$ , which can be easily obtained by Monte Carlo sampling, especially when we further approximate the Dirichlet process posterior by the Bayesian bootstrap. The Bayesian approach automatically quantifies the uncertainty in the parameter through the posterior distribution. This provides credible intervals in addition to an estimate of the parameter. Using the strong approximation technique, we establish a Bernstein-von Mises theorem (Van Der Vaart 1998) for the posterior distribution of  $\theta$ . The result gives a rate of contraction for the posterior distribution of  $\theta$  at its true value; the posterior distribution of  $\theta$  is almost surely approximately normal with center at  $\psi_n(\mathbb{F}_n)$ , while  $\psi_n(\mathbb{F}_n)$  is also asymptotically normal with center at  $(2\theta)^{-1}$ , and the two scales are the same. Consequently, the asymptotic frequentist coverage of a credible interval agrees with the credibility. Thus, the Bayesian uncertainty quantification is justified by the frequentist yardstick. The following three main steps are involved in the proof of the Bernstein-von Mises theorem:

- using the strong approximation theory of the Dirichlet process posterior, establishing a normal approximation to the posterior distribution of  $\theta$  conditional on the data and appropriately controlling the scaling constants by judicious choice of some sequences defining the functional  $\psi_n$ ;
- using the strong approximation theory of empirical processes, establishing a normal approximation to the sampling distribution of  $\psi_n(\mathbb{F}_n)$  centered at the true value of  $\psi_n(F)$ ;
- controlling the approximation error of  $\psi_n(F)$  in approximating  $\theta$ .

One application of extreme value theory is quantile regression, first introduced by Koenker and Bassett Jr. (1978). Quantile regression can be thought of as the quantile counterpart of the popular mean regression. Chernozhukov (2005) applied EVT to develop extreme quantile regression, establishing asymptotic results under both intermediate and extreme quantile levels. He et al. (2020) proposed the use of the WTC as an exponent to scale the intermediate quantile to estimate the extreme quantile. As noted earlier, well-known distributions such as the normal, gamma, logistic, and Weibull not only have an extreme value index  $\gamma = 0$  but also have a Weibull tail coefficient  $\theta > 0$ . Thus, residuals that approximately follow any of these common distributions allow for an extreme quantile to be inferred.

On the Bayesian side, Yu and Moyeed (2001) argued that, regardless of the original distribution of the underlying data, the asymmetric Laplace likelihood can be used in

Bayesian quantile regression. Furthermore, Sriram et al. (2013) showed posterior consistency of the corresponding model parameters under certain sufficient conditions. However, Yang et al. (2016) noted that the posterior variance-covariance requires adjustment in order to create asymptotically valid interval estimates.

With an asymmetric Laplace posterior,  $\beta$  can be written as a scale mixture of normals, which facilitates Monte Carlo sampling, see Kotz et al. (2001). Kozumi and Kobayashi (2011) developed a Gibbs sampling algorithm for quantile regression models based on the location-scale mixture representation of the ALD.

Value-at-Risk (VaR) (McNeil et al. 2015) quantiles are often used to set risk management thresholds in finance and insurance applications. While well-known and widely used, VaR as a risk measure does not satisfy the coherence property proposed in Artzner et al. (1999); in particular, it does not satisfy the subadditivity condition of coherence, which is more recognized as the principle of risk diversification. By contrast, the conditional tail expectation (CTE), also known as Tail Value-at-Risk (TVaR), is a coherent measure that has found meaningful use in the actuarial and insurance fields, for example, in the calculation of statutory reserves and capital.

The main proposals in this thesis are as follows:

- to present a Bayesian approach to inferring the extreme value index  $\gamma$ , thereby enabling modeling of the EV distribution;
- to recommend a Bayesian method to infer the Weibull tail coefficient  $\theta$ ;
- to propose a Bayesian analog to infer extreme quantiles and conditional tail expectations for a class of light-tailed distributions.

For the rest of this chapter, the preliminary concepts and results are covered, setting the stage for the major results covered in the next three chapters. In Chapter 2, a Bayesian inference approach for the extreme value index is discussed. This covers the key ideas in Almelor and Ghosal (2026a). The third chapter describes a Bayesian inference approach for the Weibull tail index, applicable to certain distributions in the Gumbel MDA. Almelor and Ghosal (2026b) contains the main results in this chapter. Chapter 4 covers Bayesian quantile regression for distributions in the Gumbel maximum domain of attraction and builds upon the results in the previous chapter. Chapter 5 summarizes the main conclusions and outlines possible directions for further research.

## 1.2 Extreme Value Theory

Let  $X_1, \dots, X_n$  be independent and identically distributed (i.i.d.) observations (random variables) arising from a continuous and strictly increasing unknown distribution  $F$ . We denote the survival function as  $\bar{F}$ , so that  $\bar{F}(x) = 1 - F(x)$ ,  $x \in \mathbb{R}$ . For brevity, we write the data as  $D_n = (X_1, \dots, X_n)$ . From this random sample, we denote the corresponding order statistics as  $X_{1:n} \leq \dots \leq X_{n:n}$ .

We're interested in estimating the distribution of the sample maxima. Suppose there exists sequences of constants  $a_n > 0$  and  $b_n \in \mathbb{R}$  such that

$$\frac{\max(X_1, \dots, X_n) - b_n}{a_n}$$

has a nondegenerate limit distribution as  $n \rightarrow \infty$ , i.e.,

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x),$$

for every continuity point of  $G$ , where  $G$  is a non-degenerate function.  $G$  is called an extreme value (EV) distribution.

**Theorem 1.** *The class of extreme value distributions is  $G_\gamma(ax + b)$  for some  $a > 0$  and  $b \in \mathbb{R}$  where*

$$G_\gamma(x) = \begin{cases} \exp(-(1 + \gamma x)^{-1/\gamma}), & \text{if } \gamma \neq 0 \\ \exp(e^{-x}), & \text{otherwise} \end{cases} \quad (1.3)$$

where  $(1 + \gamma x) > 0$ .

**Definition 1 (EVI).** The parameter  $\gamma$  is called the *extreme value index* (EVI).

The sign of  $\gamma$  allows the classification of the underlying distribution  $F$  to belong to a maximum domain of attraction (MDA). When  $\gamma > 0$ ,  $F$  is said to be in the Fréchet MDA. When  $\gamma < 0$ ,  $F$  is in the reverse Weibull MDA. When  $\gamma = 0$ ,  $F$  is in the Gumbel MDA.

There are various characterizations associated with the EVI  $\gamma$  and the MDA where  $F$  belongs. Two such characterizations in de Haan and Ferreira (2006) that will be used in our Bayesian approach to EVI inference are as follows:

**Theorem 2.** *The distribution  $F$  is in the MDA of the EV distribution  $G_\gamma$  if and only if*

- for  $\gamma > 0$ :  $\omega = \sup\{x : F(x) < 1\} = \infty$  and

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-1/\gamma}$$

for all  $x > 0$ ;

- $\gamma < 0$ ,  $\omega$  is finite and

$$\lim_{t \rightarrow 0} \frac{\bar{F}(\omega - tx)}{\bar{F}(\omega - t)} = x^{-1/\gamma}$$

for all  $x > 0$ ;

- for  $\gamma = 0$ :  $\omega$  can be finite or infinite and

$$\lim_{t \rightarrow \omega} \frac{\bar{F}(t + xf(t))}{\bar{F}(t)} = e^{-x}$$

for all real  $x$ , where  $f$  is a suitable positive function.

The first two parts in the above theorem would be used in the Bayesian inference of  $\gamma$ . For the Gumbel case, we use the last part of the following theorem.

**Theorem 3.** *The distribution  $F$  is in the MDA of the EV distribution  $G_\gamma$  if and only if*

- for  $\gamma > 0$ :  $\omega = \sup\{x : F(x) < 1\} = \infty$  and

$$\lim_{t \rightarrow \infty} \frac{\int_t^\infty x^{-1} \bar{F}(x) dx}{\bar{F}(t)} = \gamma;$$

- $\gamma < 0$ ,  $\omega$  is finite and

$$\lim_{t \rightarrow 0} \frac{\int_{\omega-t}^\omega (\omega - x)^{-1} \bar{F}(x) dx}{\bar{F}(\omega - t)} = -\gamma;$$

- for  $\gamma = 0$ :  $\omega$  can be finite or infinite,  $\int_x^\omega \int_t^\omega \bar{F}(s) ds dt < \infty$  and the function  $h$  defined by

$$h_F(t) := \frac{\bar{F}(t) \int_t^\omega \int_u^\omega \bar{F}(s) ds du}{\left( \int_t^\omega \bar{F}(s) ds \right)^2}$$

satisfies  $\lim_{t \rightarrow \omega} h(t) = 1$ .

As  $F$  is generally unknown,  $\gamma$  has to be estimated, which in turn enables the estimation of the EV distribution. Many estimators for  $\gamma$  have been proposed, and in our research, we focus on three familiar ones.

- **Hill Estimator:** The Hill estimator (Hill 1975) for  $\gamma > 0$  is given by

$$\hat{\gamma}_H = \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n-i:n} - \log X_{n-k:n},$$

where we choose  $k = k(n) \rightarrow \infty$  and  $k/n \rightarrow 0$  to ensure consistency of the estimator.

Since the Hill estimator is applicable only for  $\gamma > 0$ , a somewhat, but not fully, complementary estimator was proposed by Falk (1995), which is applicable for  $\gamma < -1/2$ .

- **Falk Estimator:** The Falk estimator for  $\gamma < -1/2$  is given by

$$\hat{\gamma}_F = \frac{1}{k} \sum_{i=1}^{k-1} \log(X_{n:n} - X_{n-i:n}) - \log(X_{n:n} - X_{n-k:n}),$$

where, in addition to the conditions on  $k = k(n)$  described for Hill, we also assume  $k^\eta / \log n \rightarrow \infty$  for some small  $\eta$ .

Another estimator was proposed by Pickands III (1975).

- **Pickands Estimator:** The Pickands estimator for  $\gamma \in \mathbb{R}$  is given by

$$\hat{\gamma}_P = \frac{1}{\log 2} \log \frac{X_{n-k:n} - X_{n-2k:n}}{X_{n-2k:n} - X_{n-4k:n}},$$

where all conditions on  $k$  imposed for the Hill estimator also apply to the Pickands estimator.

There are many more estimators in the literature, but we will focus on these three estimators for comparison with our proposed Bayesian approach.

### 1.3 Bayesian Nonparametrics

A Bayesian method for inference on a parameter such as  $\gamma$  or  $\theta$  can be established by assigning an appropriate prior on  $F$  (equivalently, the underlying probability measure) and using the induced posterior distribution of an approximating sequence of functionals of  $F$  to make an inference on the parameter. Since we make no parametric assumption on  $F$ , we consider a Dirichlet Process prior (Ferguson 1973; Ghosal and van der Vaart 2017) because of its large support and the tractability of the posterior distribution. We recall that a random measure  $P$  on a sample space  $(\mathfrak{X}, \mathcal{X})$  has a *Dirichlet Process* distribution  $\text{DP}(\alpha)$  with base measure  $\alpha$  if for every finite measurable partition  $A_1, \dots, A_k$  of  $\mathfrak{X}$ ,

$$(P(A_1), \dots, P(A_k)) \sim \text{Dir}(k; \alpha(A_1), \dots, \alpha(A_k));$$

here,  $\text{Dir}(k; a_1, \dots, a_k)$  stands for the Dirichlet distribution on the unit  $k$ -simplex with density at  $\mathbf{x} = (x_1, \dots, x_{k-1})$  given by

$$\frac{\Gamma(a_1 + \dots + a_k)}{\Gamma(a_1) \cdots \Gamma(a_k)} x_1^{a_1-1} x_2^{a_2-1} \cdots x_{k-1}^{a_{k-1}-1} (1 - x_1 - \dots - x_{k-1})^{a_k-1}, \quad \mathbf{x} \in \mathbb{D}_{k-1},$$

where  $\mathbb{D}_{k-1} = \{(x_1, \dots, x_{k-1}) : \min_i x_i \geq 0, \sum_{i=1}^{k-1} x_i \leq 1\}$ , the representation of the unit simplex in  $\mathbb{R}^k$  in  $(k-1)$ -dimensions. Since we deal with probability measures on  $\mathbb{R}$ , we can represent the base measure by  $\alpha = MG$  for some c.d.f.  $G$  and  $M = \alpha(\mathbb{R}) > 0$ , the precision parameter.

We denote the empirical distribution based on the random sample  $D_n$  from  $F$  by  $\mathbb{F}_n(x) = n^{-1} \sum_{i=1}^n \mathbf{1}\{X_i \leq x\}$ , and the corresponding empirical process by  $\mathbb{J}_n(x) := \sqrt{n}(\mathbb{F}_n(x) - F(x))$ ,  $x \in \mathbb{R}$ . It is well known that the Dirichlet process has the conjugacy property: the posterior is again a Dirichlet process with the updated base measure  $MG + n\mathbb{F}_n$ . Then, the posterior distribution of  $\theta$  can be obtained by sampling  $F$  from its posterior and computing  $\psi_n(F)$  for each realization. We view  $F$  drawn from its posterior distribution  $\text{DP}(MG + n\mathbb{F}_n)$  as a stochastic process and write  $\mathbb{F}_n^*$  for it. Then we study the distribution of the functional  $\psi_n(\mathbb{F}_n^*)$ . The Dirichlet process posterior, centered by the empirical distribution and normalized by  $\sqrt{n}$ , is denoted by  $\mathbb{J}_n^*(x) = \sqrt{n}(\mathbb{F}_n^*(x) - \mathbb{F}_n(x))$ ,  $x \in \mathbb{R}$ .

While it is possible to obtain approximate samples of  $F$  distributed with a Dirichlet process distribution using various approximate sampling schemes (see Ghosal and van der Vaart (2017) for details), a remarkable simplification is obtained by considering the noninformative limit of  $\text{DP}(MG + n\mathbb{F}_n)$  as  $M \rightarrow 0$ , resulting in the Bayesian bootstrap (BB) distribution  $\text{DP}(n\mathbb{F}_n)$  (Rubin 1981), which is representable as  $\sum_{i=1}^n W_i \delta_{X_i}$ , where  $(W_1, \dots, W_n) \sim \text{Dir}(n; 1, \dots, 1)$ . Observe that the expectation of the BB process, conditional on  $(X_1, \dots, X_n)$ , is the empirical distribution function  $\mathbb{F}_n$ . We write  $\mathbb{F}_n^\#$  for the stochastic process  $F$  following the BB distribution. The BB differs from the Dirichlet process posterior by  $O(n^{-1})$ , that is,  $\|\mathbb{F}_n^* - \mathbb{F}_n^\#\|_\infty = O(n^{-1})$  a.s. This implies that the asymptotic properties of the procedures based on the Dirichlet process posterior and the Bayesian bootstrap are identical. Since the base measure's effect dissipates, its choice does not significantly impact the results. The most significant benefit of using the BB is that it depends only on  $n$  variables, so samples from it can be readily obtained. On the other hand, sampling from the Dirichlet process involves truncating its stick-breaking representation, which involves infinitely many variables and therefore requires truncation. The simplification also frees us from choosing a base measure to implement the procedure. The BB is the basis for the simulation that is discussed in 2.4. The BB process, centered by the empirical distribution and normalized by  $\sqrt{n}$ , is denoted by  $\mathbb{J}_n^\#(x) = \sqrt{n}(\mathbb{F}_n^\#(x) - \mathbb{F}_n(x))$ ,  $x \in \mathbb{R}$ .

## 1.4 Strong Approximations

A Kiefer process is a stochastic process  $\{K(x, y) : x \in [0, 1], y \geq 0\}$  satisfying

$$K(x, y) = W(x, y) - xW(1, y)$$

for some two-parameter Wiener process  $W(x, y)$ . A few useful properties of Kiefer processes include:

- $E[K(x, y)] = 0$  for  $x \in [0, 1], y \geq 0$
- $E[K(x_1, y_1)K(x_2, y_2)] = (x_1 \wedge x_2 - x_1x_2)(y_1 \wedge y_2)$  for  $x_1, x_2 \in [0, 1], y_1, y_2 \geq 0$ .

The following strong approximations are used to establish asymptotic results.

- *Strong approximation for the empirical process* (Theorem 4 of Komlós et al. (1975)): There exists a Kiefer process  $K_1$  such that

$$\mathbb{J}_n(x) = n^{-1/2}K_1(F(x), n) + \kappa_n(x), \quad x \in \mathbb{R}, \quad (1.4)$$

and  $\sup\{|\kappa_n(x)| : x \in \mathbb{R}\} = O(n^{-1/2} \log^2 n)$  a.s. as  $n \rightarrow \infty$ , and

$$\sup\{|\mathbb{J}_n(x)| : x \in \mathbb{R}\} = O(\sqrt{\log \log n}) \quad \text{a.s.} \quad (1.5)$$

- *Strong approximation for the posterior Dirichlet process* (Lemma 6.3 of Lo (1987)): There exists a Kiefer process  $K_2$  independent of  $K_1$  above, conditional on the data

$$\mathbb{J}_n^*(x) = n^{-1/2}K_2(\mathbb{F}_n(x), n) + \delta_n(x), \quad x \in \mathbb{R}, \quad (1.6)$$

with  $\sup\{|\delta_n(x)| : x \in \mathbb{R}\} = O(n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4})$  a.s. as  $n \rightarrow \infty$ . Moreover, conditional on almost all observations, as  $n \rightarrow \infty$ , uniformly in  $x$ ,

$$\mathbb{J}_n^*(x) = n^{-1/2}K_2(F(x), n) + O(n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}), \quad (1.7)$$

with  $F$  replacing  $\mathbb{F}_n$  and

$$\sup\{|\mathbb{J}_n^*(x)| : x \in \mathbb{R}\} = O(\sqrt{\log \log n}) \quad \text{a.s.} \quad (1.8)$$

- *Strong approximation for the Bayesian bootstrap process:* Specializing the Dirichlet process posterior to the BB process, we also have

$$\mathbb{J}_n^\#(x) = n^{-1/2}K_2(\mathbb{F}_n(x), n) + O(n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}), \quad (1.9)$$

$F$  can replace  $\mathbb{F}_n$  above, and

$$\sup\{|\mathbb{J}_n^\#(x)| : x \in \mathbb{R}\} = O(\sqrt{\log \log n}) \quad \text{a.s.} \quad (1.10)$$

Csorgo and Revesz (1981) is a good reference for strong approximation topics.

## 1.5 Quantile Regression

Koenker and Bassett Jr. (1978) introduced the concept of quantile regression, as an alternative to the more common mean regression, which is more appropriate in many situations, including highly skewed or heavy-tailed distributions. Let  $\mathbf{X}$  be a  $p$ -dimensional covariate with first element equal to 1 and  $Y$  be the dependent variable. We consider the linear quantile regression model

$$q_Y(\tau|\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}_\tau \quad (1.11)$$

for some given quantile level  $\tau \in (0, 1)$ , where  $\boldsymbol{\beta}_\tau$  is an unknown  $p$ -dimensional quantile coefficient. Given data  $(\mathbf{X}_i, Y_i), i = 1, \dots, n$ , a consistent estimator for  $\boldsymbol{\beta}_\tau$  is given by

$$\hat{\boldsymbol{\beta}}_\tau = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i^T \boldsymbol{\beta}), \quad (1.12)$$

where  $\rho_\tau(u) = u(\tau - \mathbf{1}_{u < 0})$  is the quantile loss function.

Recall that the asymmetric Laplace density with parameters  $\tau \in (0, 1)$  and  $\sigma > 0$  is given by

$$f_\tau(u) = \frac{\tau(1-\tau)}{\sigma} \exp\{-\rho_\tau(u)/\sigma\}. \quad (1.13)$$

If we let  $u = y_i - \mathbf{x}_i^T \boldsymbol{\beta}$ , we see that the minimization in (1.12) is the same as the maximization of the likelihood obtained by the product of terms in (1.13).

Shifting to a Bayesian perspective, let  $\pi(\boldsymbol{\beta}_\tau)$  be a prior on  $\boldsymbol{\beta}_\tau$ , so that the posterior of  $\boldsymbol{\beta}_\tau$

can be written as

$$p(\boldsymbol{\beta}_\tau | D_n) \propto \pi(\boldsymbol{\beta}_\tau) \exp \left\{ - \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\tau) / \sigma \right\}. \quad (1.14)$$

By Yu and Moyeed (2001), the choice of the improper uniform prior on  $\boldsymbol{\beta}_\tau$  results in a proper posterior distribution. Hence, it is reasonable to consider the working likelihood

$$L(\boldsymbol{\beta}_\tau | D_n) = \frac{\tau^n (1 - \tau)^n}{\sigma^n} \exp \left\{ - \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\tau) / \sigma \right\}. \quad (1.15)$$

This means that even if  $\boldsymbol{\beta}_\tau$  does not follow an asymmetric Laplace distribution (i.e., misspecified) in reality, we can still sample from the posterior of  $\boldsymbol{\beta}_\tau$  by assuming it follows ALD, and Sriram et al. (2013) demonstrated posterior consistency of the model parameters under certain sufficient conditions with the above likelihood setup. However, as noted by Yang et al. (2016), the posterior variance-covariance must be adjusted to obtain asymptotically valid interval estimates.

For a univariate random variable  $Z$ , we recall the mean excess function  $e(u) = E[Z - u | Z > u]$ . If  $q_Z(\tau) := \inf\{z : \bar{F}_Z(z) \leq \tau\}$  is the  $\tau$  quantile of  $Z$  of interest, then the conditional tail expectation at  $\tau$  is given by  $e(q_Z(\tau)) + q_Z(\tau)$ . These concepts extend naturally in the regression data setting.

## CHAPTER

# 2

# BAYESIAN INFERENCE ON THE EXTREME VALUE INDEX

## 2.1 Introduction

Our goal in this chapter is to propose a Bayesian framework for which an extreme value index of an EV distribution can be inferred. The approaches between non-zero EVI and zero EVI are substantially different and thus merit separate discussions.

As discussed in Chapter 1, the characterization in Theorem 2 is the starting point for inferring the non-zero EVI  $\gamma$ . Depending on the sign of  $\gamma$ , we derive separate approximations for  $\gamma$  in terms of the unknown underlying distribution  $F$ . A sequence of functionals approximating  $\gamma$  is constructed in terms of integrals with integration limits approaching infinity at different speeds. The sequence of functionals for  $\gamma > 0$  is different from the sequence for  $\gamma < 0$ .

We apply a Dirichlet process prior on  $F$  and use the induced posterior of the functional of  $F$  for inference on  $\gamma$ . We simplify the posterior sampling scheme by replacing the DP prior by its noninformative limit, also known as the Bayesian bootstrap. This facilitates the MCMC sampling from the approximate posterior distribution of  $\gamma$ .

The value  $\gamma = 0$  occupies a distinguished position with many common distributions falling in the MDA of the Gumbel distribution. To test the hypothesis  $\gamma = 0$ , we use a

characterization for  $F$  in the MDA of the Gumbel distribution given in Theorem 3, namely,  $h_F(t) \rightarrow 1$  as  $t \rightarrow \omega$ , where

$$h_F(t) := \frac{\bar{F}(t) \int_t^\omega \int_u^\omega \bar{F}(s) ds du}{\left(\int_t^\omega \bar{F}(s) ds\right)^2}. \quad (2.1)$$

This allows us to compare the plausibility of two models,  $M_0$  and  $M_1$ , respectively representing the unrestricted model (either  $\gamma \geq 0$  or  $\gamma \leq 0$ ) and the null hypothesis  $\gamma = 0$  for the Gumbel MDA. In the Bayesian approach, comparisons are based on the Bayes factor. However, the nonparametric models with the Dirichlet process prior are undominated, so a traditional Bayes factor does not exist. To address the problem, we define a Bayes factor based on the empirical likelihood  $\prod_{i=1}^n W_i$  with weights  $W_i$  at the observation  $X_i$ ,  $i = 1, \dots, n$ . We place the uniform prior on the simplex for  $(W_1, \dots, W_n)$  in Model  $M_0$ . For the model  $M_1$ , the uniform prior restricted to a region of the form

$$\{(w_1, \dots, w_n) : 1 - \delta_n \leq h_F(t_n) \leq 1 + \delta_n, F = \sum_{i=1}^n W_i \delta_{X_i}\} \quad (2.2)$$

with a suitably large  $t_n$  and small  $\delta_n$ . We obtain posterior samples from the restricted model  $M_1$  by the Hamiltonian Monte Carlo method. We compute the marginal likelihoods of the models  $M_0$  and  $M_1$  using the harmonic mean technique of Newton and Raftery (1994). For a technical reason about the convergence of the integral, we use a power  $\alpha$  slightly less than 1 of the empirical likelihood. The marginal likelihood for  $M_0$  is computed analytically, while the Monte Carlo method is adopted to compute the marginal likelihood of  $M_1$ .

The rest of this chapter is organized as follows. We formally describe the setup and the methodology in the next section. In Section 2.3, we derive convergence results for the case  $\gamma \neq 0$ . Section 2.4 presents results from a simulation study that calculates coverage and other metrics for various distributions under the Fréchet and reverse Weibull MDAs. We also present calculations of the Bayes factors for data from distributions within either Gumbel or non-Gumbel MDAs. The proofs are given in Section 2.5.

## 2.2 Setup and Methodology

From independent and identically distributed (i.i.d.) observations  $X_1, \dots, X_n$  arising from a continuous distribution  $F$  with extreme value index  $\gamma$ , denote the corresponding order statistics as  $X_{1:n} \leq \dots \leq X_{n:n}$ .

### 2.2.1 EVI $\neq 0$

Theorem 1.2.1 in de Haan and Ferreira (2006) gives the following characterization for all  $x > 0$ :

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-1/\gamma}, \text{ if } \gamma > 0, \quad (2.3)$$

$$\lim_{t \rightarrow 0} \frac{\bar{F}(\omega - tx)}{\bar{F}(\omega - t)} = x^{-1/\gamma}, \text{ if } \gamma < 0. \quad (2.4)$$

Dividing the terms in (2.3) by  $x$  and integrating over  $[1, 1+v]$  for some small  $v > 0$  and for large  $t$ , we obtain, using the generalized binomial theorem, that

$$\int_1^{1+v} \frac{\bar{F}(tx)}{x\bar{F}(t)} dx \approx \gamma[1 - (1+v)^{-1/\gamma}] = v - \frac{1}{2}(1+1/\gamma)v^2 + O(v^3).$$

This leads to the relation

$$\frac{1}{2}(1+1/\gamma) \approx \frac{1}{v} \left[ 1 - \frac{1}{v} \int_1^{1+v} \frac{\bar{F}(tx)}{x\bar{F}(t)} dx \right] + O(v). \quad (2.5)$$

By making substitutions  $t = a_n$  and  $z = a_n x$ , this motivates the consideration of the functional

$$\psi_n(F) = \frac{1}{v_n} \left[ 1 - \frac{1}{v_n} \int_{a_n}^{a_n(1+v_n)} \frac{\bar{F}(z)}{z\bar{F}(a_n)} dz \right], \quad n = 1, 2, \dots, \quad (2.6)$$

where  $a_n \rightarrow \infty$  and  $v_n \rightarrow 0$  as  $n \rightarrow \infty$ . Similarly, we construct the functional

$$\zeta_n(F) = \frac{1}{\epsilon_n} \left[ 1 - \frac{1}{\epsilon_n} \int_{\omega - \epsilon_n}^{\omega - \epsilon_n + \epsilon_n^2} \frac{\bar{F}(x)}{(\omega - x)\bar{F}(\omega - \epsilon_n)} dx \right], \quad n = 1, 2, \dots, \quad (2.7)$$

where  $\epsilon_n \rightarrow 0$ .

We develop a Bayesian method by assigning the celebrated Dirichlet process prior (Ferguson 1973; Ghosal and van der Vaart 2017) on  $F$  (which is equivalent to a prior on the underlying probability measure  $P$ ), updating to its conjugate posterior and using the induced posterior distribution of  $\psi_n(F)$  for inference on  $\gamma$  if  $\gamma > 0$  (respectively, the induced posterior of  $\zeta_n(F)$  if  $\gamma < 0$ ). The large support and the structure of the Dirichlet process posterior allow fast computation via independent sampling and asymptotic analysis of the induced posterior distributions of functionals via a strong representation of the Dirichlet process.

As described earlier, the inference on the EVI  $\gamma$  proceeds by approximately representing it by a sequence of functionals  $\psi_n(F)$  or  $\zeta_n(F)$ , depending on the sign of  $\gamma$ , which is assumed to

be given. Then, a reasonable estimator is obtained by plugging in the empirical distribution in place of  $F$ . For the Bayesian approach,  $F$  is given a prior distribution, and the inference is based on the induced posterior distribution of either  $\psi_n(F)$  or  $\zeta_n(F)$ .

### 2.2.2 EVI = 0

When  $\gamma = 0$ , we recall that  $h_F(t)$  defined in (2.1) tends to 1 as  $t \uparrow \omega$ . This limiting condition allows us to propose a test based on a modified Bayes factor, comparing two models  $M_0$  and  $M_1$ , where  $M_0$  denotes  $F$  unrestricted and  $M_1$  denotes  $F$  in the MDA of the Gumbel distribution. Since the statistical models are undominated, a traditional Bayes factor does not exist. We, therefore, consider an empirical likelihood as the basis for the Bayes factor, that is, a weight  $W_i$  is attached to the observation  $X_i$ ,  $i = 1, \dots, n$ , where  $W_i \geq 0$ ,  $\sum_{i=1}^n W_i = 1$ , that is,  $F$  is fully supported by the observed points. However, since the observations are already used, there is no prior distribution on  $(W_1, \dots, W_n)$ ; instead, the BB posterior distribution is directly imposed on  $F$  in the unrestricted model  $M_0$ . This is equivalent to the uniform posterior distribution for  $(W_1, \dots, W_n)$  on the unit simplex  $\mathcal{S}_n$  in  $\mathbb{R}^n$  with constant density  $(n-1)!$ . We then use the harmonic mean representation of the marginal likelihood by Newton and Raftery (1994) to obtain the marginal likelihood of the model  $M_0$ . For the model  $M_1$  restricted by the Gumbel MDA, the same method is used except that now the BB posterior is limited to the region qualitatively described by (2.1). Substituting  $\sum_{i=1}^n W_i \delta_{X_i}$  for  $F$ , the function  $h_F$  in (2.1) reduces to

$$\begin{aligned}
h_F(t) = & \left\{ \left( \sum_{i=r(t)}^n W_i \right)^2 \left( \frac{1}{2} X_{r(t):n}^2 - X_{r(t):n} t + \frac{1}{2} t^2 \right) \right. \\
& + \left. \left( \sum_{i=r(t)}^n W_i \right) \sum_{j=r(t)}^{n-1} \left( \sum_{i=j+1}^n W_i \right) \left[ \frac{1}{2} X_{j+1:n}^2 - \frac{1}{2} X_{j:n}^2 - (X_{j+1:n} - X_{j:n}) t \right] \right\} \\
& \div \left\{ \left( \sum_{i=r(t)}^n W_i \right) (X_{r(t):n} - t) + \sum_{j=r(t)}^{n-1} \left( \sum_{i=j+1}^n W_i \right) (X_{j+1:n} - X_{j:n}) \right\}^2,
\end{aligned}$$

where  $r(t) \in \{1, \dots, n\}$  such that  $X_{r(t)-1:n} \leq t < X_{r(t):n}$ .

To allow the expression to differ from the limit mildly, we use a threshold sequence  $\epsilon_n \rightarrow 0$  to describe the region corresponding to  $M_1$  to be

$$\mathcal{R}_n = \left\{ (w_1, \dots, w_n) \in \mathcal{S}_n : 1 - \epsilon_n < h_F(t_n) < 1 + \epsilon_n, F = \sum_{i=1}^n w_i \delta_{X_i} \right\} \quad (2.8)$$

for some sequence  $t_n \uparrow \omega$  as  $n \rightarrow \infty$ . Now the Bayes factor  $B_{10}$  favoring Model  $M_1$  over  $M_0$  is given by the ratio of the harmonic mean of  $\prod_{i=1}^n W_i$  with respect to the uniform distribution on  $\mathcal{R}_n$  and  $\mathcal{S}_n$  respectively. The latter can be evaluated analytically, but the integral is infinite; hence, a modification is needed to make the integral convergent. To this end, we replace the likelihood with a power of it,  $\alpha < 1$ , a practice occasionally used in Bayesian statistics for estimation and model selection; see Friel and Pettitt (2008); O’Hagan (1995) and others. In the present context, replacing the likelihood by its  $\alpha$ -power and using the BB posterior is equivalent to using the Bayesian bootstrap clone posterior (c.f., Lo (1991)) corresponding to the Gamma distribution with shape parameter  $2 - \alpha$  as the weighting distribution, and using the usual likelihood for computing the marginal likelihood of the models.

The modified marginal likelihood in Model  $M_0$  can be calculated analytically:

$$\frac{1}{p_{M_0}(D_n)} = \int_{\mathcal{S}_n} (n-1)! (\prod_{i=1}^n w_i)^{-\alpha} dw_1 \cdots dw_n = \frac{(n-1)! [\Gamma(1-\alpha)]^n}{\Gamma(n(1-\alpha))}. \quad (2.9)$$

To sample from the modified empirical likelihood, we will use the Hamiltonian Monte Carlo (Neal 2012). This method controls the random-walk behavior in the well-known Metropolis algorithm, allowing faster movement within the target (posterior) distribution (Gelman et al. 2013).

In the next section, we shall obtain the convergence rate of the estimators and the limiting coverage and length of the credible intervals.

## 2.3 Convergence Results

**Definition 2.** We shall call a distribution  $F$  in the Fréchet MDA to be  $(a_n, v_n, m_n)$ -regular for sequences  $a_n \rightarrow \infty$ ,  $v_n \rightarrow 0$  and  $m_n \rightarrow 0$  if

$$\sup_{z \in [1, 1+v_n]} \left| \frac{\bar{F}(a_n z)}{\bar{F}(a_n)} - z^{-1/\gamma} \right| = O(m_n). \quad (2.10)$$

Similarly, we define a distribution  $F$  in the reverse Weibull MDA to be  $(\epsilon_n, m_n)$ -regular for sequences  $\epsilon_n \rightarrow 0$  and  $m_n \rightarrow 0$  if

$$\sup_{z \in [1-\epsilon, 1]} \left| \frac{\bar{F}(\omega - \epsilon_n z)}{\bar{F}(\omega - \epsilon_n)} - z^{-1/\gamma} \right| = O(m_n). \quad (2.11)$$

The standard distributions in the Fréchet and reverse Weibull maximum domains of attraction listed earlier satisfy these respective regularity conditions, as described in the

following proposition.

**Proposition 1.** *The following distributions are in the Fréchet MDA with  $\gamma$  and  $m_n$  as specified below:*

- Pareto with shape parameter  $\alpha$ :  $\gamma = 1/\alpha$ ,  $m_n = 0$ ;
- Fréchet with shape parameter  $\alpha$ :  $\gamma = 1/\alpha$ ,  $m_n = a_n^{-\alpha}$ ;
- Student's  $t$  with degrees of freedom  $\nu$ :  $\gamma = 1/\nu$ ,  $m_n = a_n^{-4}$ ;
- Snedecor's  $F$  with degrees of freedom  $\nu_1, \nu_2$ :  $\gamma = 2/\nu_2$ ,  $m_n = a_n^{-2}$ .

*The following distributions are in the reverse Weibull MDA with  $\gamma$  and  $m_n$  as specified below:*

- Uniform:  $\gamma = -1$ ,  $m_n = 0$ ;
- Beta with shape parameters  $(\alpha, \beta)$ :  $\gamma = -1/\beta$ ,  $m_n = \epsilon_n$ ;
- Triangular:  $\gamma = -1/2$ ,  $m_n = 0$ .

**Proposition 2.** *Let  $a_n \rightarrow \infty$ ,  $v_n \rightarrow 0$ , and  $\psi_n$  be defined by (2.6). If  $F$  is  $(a_n, v_n, m_n)$ -regular with  $m_n = o(v_n)$ , then*

$$\psi_n(F) = (1 + 1/\gamma)/2 + O(v_n + m_n/v_n). \quad (2.12)$$

*Similarly, let  $\epsilon_n \rightarrow 0$  and  $\zeta_n$  be defined by (2.7). If  $F$  is  $(\epsilon_n, m_n)$ -regular with  $m_n = o(\epsilon_n)$ , then*

$$\zeta_n(F) = -(1 + 1/\gamma)/2 + O(\epsilon_n + m_n/\epsilon_n). \quad (2.13)$$

Denote the true distribution of the observations  $X_1, \dots, X_n$  by  $F_0$  with extreme value index  $\gamma_0$ .

**Corollary 1** (Consistency). *Let  $a_n \rightarrow \infty$ ,  $v_n \rightarrow 0$ ,  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ ,  $\psi_n$  be defined by (2.6) and  $\zeta_n$  be defined by (2.7). For  $\gamma > 0$ , if  $F_0$  satisfies (2.10), then*

- (i)  $\psi_n(\mathbb{F}_n)$  is a consistent estimator for  $(1 + 1/\gamma)/2$  at  $\gamma = \gamma_0$ ;
- (ii) the induced posterior distribution of  $\psi_n(F)$  is consistent for  $(1 + 1/\gamma)/2$  at  $\gamma = \gamma_0$ .

*For  $\gamma < 0$ , if  $F_0$  satisfies (2.11), then*

(iii)  $\zeta_n(\mathbb{F}_n)$  is a consistent estimator for  $-(1 + 1/\gamma)/2$  at  $\gamma = \gamma_0$ ;

(iv) the induced posterior distribution of  $\psi_n(F)$  is consistent for  $-(1 + 1/\gamma)/2$  at  $\gamma = \gamma_0$ .

Let  $\phi_{n,F,H,\psi}(t) = \psi_n(F + tH)$ , so that  $\phi_{n,F,H,\psi}(0) = \psi_n(F)$ . Then  $\phi'_{n,F,H,\psi}(0) = \psi'_{n,F}(H)$ , the Gateaux derivative of  $\psi_n$  at  $F$  in the direction of the càdlàg function  $H$  on  $\mathbb{R}$ . By the Taylor series expansion, we have

$$\phi_{n,F,H,\psi}(t) = \phi_{n,F,H,\psi}(0) + \phi'_{n,F,H,\psi}(0)t + \phi''_{n,F,H,\psi}(\eta_n)t^2/2,$$

for some  $\eta_n$  in between 0 and  $t$ . Making the substitutions  $t = 1/\sqrt{n}$  and  $H = n^{-1/2}K_1(F_0(\cdot), n) + \kappa_n$ , the above Taylor expansion can be expressed as

$$\begin{aligned} \psi_n(\mathbb{F}_n) - \psi_n(F_0) &= n^{-1/2}\psi'_{n,F_0}(n^{-1/2}K_1(F_0(\cdot), n) + \kappa_n) + (2n)^{-1}\phi''_{n,F_0,H,\psi}(\eta_n) \\ &= n^{-1/2}\psi'_{n,F_0}(n^{-1/2}K_1(F_0(\cdot), n)) + n^{-1/2}\psi'_{n,F_0}(\kappa_n) + (2n)^{-1}\phi''_{n,F_0,H,\psi}(\eta_n). \end{aligned}$$

This allows us to get an approximate distribution of  $\psi_n(\mathbb{F}_n) - \psi_n(F_0)$  by looking at the distribution of the terms on the right-hand side. To that end, we want to show that the last two terms on the right are stochastically bounded. In effect, the distribution of  $\psi_n(\mathbb{F}_n) - \psi_n(F_0)$  only depends on  $\psi'_{n,F_0}(n^{-1/2}K_1(F_0(\cdot), n))$  for large  $n$ . The above expansion also holds when we replace  $\psi_n$  with  $\zeta_n$  instead.

**Lemma 1.** *For any càdlàg function  $H$  on  $\mathbb{R}$ , the Gateaux derivative of  $\psi_n$  at  $F$  in the direction of  $H$  is given by*

$$\psi'_{n,F}(H) = \frac{\bar{F}(a_n) \int_{a_n}^{b_n} x^{-1} H(x) dx - H(a_n) \int_{a_n}^{b_n} x^{-1} \bar{F}(x) dx}{v_n^2[\bar{F}(a_n)]^2}, \quad (2.14)$$

while the Gateaux derivative of  $\zeta_n$  at  $F$  in the direction of  $H$  is given by

$$\zeta'_{n,F}(H) = \frac{\bar{F}(\omega - \epsilon_n) \int_{\omega - \epsilon_n}^{\omega - \epsilon_n + \epsilon_n^2} \frac{H(x)}{\omega - x} dx - H(\omega - \epsilon_n) \int_{\omega - \epsilon_n}^{\omega - \epsilon_n + \epsilon_n^2} \frac{\bar{F}(x)}{\omega - x} dx}{\epsilon_n^2[\bar{F}(\omega - \epsilon_n)]^2}. \quad (2.15)$$

The remainder term in the Taylor series expansion involves the second derivative of  $\phi$ , the formula for which is given below.

**Lemma 2.** *For any càdlàg function  $H$  on  $\mathbb{R}$ , the second derivative  $\phi''_{n,F,H,\psi}(t)$  for the case  $\gamma > 0$  is given by*

$$\frac{2\bar{F}(a_n)H(a_n) \int_{a_n}^{b_n} x^{-1} H(x) dx - 2[H(a_n)]^2 \int_{a_n}^{b_n} x^{-1} \bar{F}(x) dx}{v_n^2[\bar{F}(a_n) - tH(a_n)]^3} \quad (2.16)$$

and the second derivative  $\phi''_{n,F,H,\zeta}(t)$  for the case  $\gamma < 0$  is given by

$$\frac{2\bar{F}(\omega - \epsilon_n)H(\omega - \epsilon_n) \int_{\omega - \epsilon_n}^{\omega - \epsilon_n + \epsilon_n^2} \frac{H(x)}{\omega - x} dx - 2[H(\omega - \epsilon_n)]^2 \int_{\omega - \epsilon_n}^{\omega - \epsilon_n + \epsilon_n^2} \frac{\bar{F}(x)}{\omega - x} dx}{\epsilon_n^2 [\bar{F}(\omega - \epsilon_n) - tH(\omega - \epsilon_n)]^3}. \quad (2.17)$$

The following result provides a formula for the variance of the linear term in the Taylor expansion.

**Lemma 3.** *The variance  $\sigma_{n,\psi}^2$  of  $\psi'_{n,F_0}(n^{-1/2}K_1(F_0(\cdot), n))$  is given by*

$$\sigma_{n,\psi}^2 = \frac{1}{v_n^4} \left\{ \frac{1}{[\bar{F}_0(a_n)]^2} \int_{a_n}^{b_n} \int_{a_n}^{b_n} \frac{F_0(x) \wedge F_0(y) - F_0(x)F_0(y)}{xy} dx dy - \frac{1}{[\bar{F}_0(a_n)]^3} \left( \int_{a_n}^{b_n} x^{-1} \bar{F}_0(x) dx \right)^2 F_0(a_n) \right\}$$

and is of the order of  $v_n^{-2}p_n^{-1}$ .

The variance  $\sigma_{n,\zeta}^2$  of  $\zeta'_{n,F_0}(n^{-1/2}K_1(F_0(\cdot), n))$  is given by

$$\sigma_{n,\zeta}^2 = \frac{1}{\epsilon_n^4} \left\{ \frac{1}{[\bar{F}_0(\omega - \epsilon_n)]^2} \int_{\omega - \epsilon_n}^{\omega - \epsilon_n + \epsilon_n^2} \int_{\omega - \epsilon_n}^{\omega} \frac{F_0(x) \wedge F_0(y) - F_0(x)F_0(y)}{(\omega - x)(\omega - y)} dx dy - \frac{1}{[\bar{F}_0(\omega - \epsilon_n)]^3} \left( \int_{\omega - \epsilon_n}^{\omega - \epsilon_n + \epsilon_n^2} (\omega - x)^{-1} \bar{F}_0(x) dx \right)^2 F_0(\omega - \epsilon_n) \right\}.$$

and is of the order of  $\epsilon_n^{-2}p_n^{-1}$ .

We now arrive at one of the main objectives of this paper: establishing the asymptotic normality of the plug-in estimator. For the convergence theory to work,  $v_n$ ,  $\epsilon_n$ , and  $a_n$  must be chosen to control the remainder terms in the Taylor expansions to achieve asymptotic normality. For the Fréchet MDA, we choose  $a_n$  such that  $p_n = \bar{F}(a_n) \asymp n^{-\tau}$  for some  $0 < \tau < 1/2$  and  $v_n \asymp n^{-\xi}$  for some  $\xi > 0$ . For the Reverse Weibull MDA, we set  $p_n = \bar{F}(\omega - \epsilon_n) \asymp n^{-\tau}$  and  $\epsilon_n \asymp n^{-\xi'}$  for some  $\xi' > 0$ .

**Theorem 4** (Asymptotic normality of plug-in estimator). *If  $F_0$  is  $(a_n, v_n, m_n)$ -regular,  $\tau < 1/3$ ,  $\xi > 1/6$  and  $m_n = o(n^{\tau/2-1/2})$ , then*

$$\frac{\sqrt{n}}{\sigma_{n,\psi}} [\psi_n(\mathbb{F}_n) - (1 + 1/\gamma_0)/2] \rightsquigarrow N(0, 1). \quad (2.18)$$

Similarly, if  $F_0$  is  $(\epsilon_n, m_n)$ -regular,  $\tau < 1/3$ ,  $\xi' > 1/6$  and  $m_n = o(n^{\tau/2-1/2})$ , then

$$\frac{\sqrt{n}}{\sigma_{n,\zeta}} [\zeta_n(\mathbb{F}_n) + (1 + 1/\gamma_0)/2] \rightsquigarrow \text{N}(0, 1). \quad (2.19)$$

We experimented with different choices of  $\tau$ ,  $\xi$ , and  $\xi'$  in our simulation study. Since  $F$  is unknown, we proxy the choice of  $a_n$  for  $\psi_n$  by using the empirical distribution function; choose  $a_n$  as the empirical  $p_n$ -quantile  $q_n(p_n)$ . By the law of the iterated logarithm for the empirical process,

$$\bar{F}_0(q_n(p_n)) = \bar{\mathbb{F}}_n(q_n(p_n)) + O(\sqrt{(\log \log n)/n}) \asymp p_n \quad (2.20)$$

almost surely, since  $\bar{\mathbb{F}}_n(q_n(p_n)) = p_n + O(1/n)$  and  $p_n \gg (n/\log \log n)^{-1/2}$ . Hence, with probability one, the implementable choice  $a_n = q_n(p_n)$  satisfies the requirement. For  $\zeta_n$ , since  $\omega$  is unknown, we use the maximum random sample value  $X_{n:n}$  as the estimator for  $\omega$ , which then leads to the choice of  $X_{n:n} - \epsilon_n$  for  $a_n$ .

Consider a Dirichlet process prior  $F \sim \text{DP}(MG)$ , where  $G$  is a cumulative probability distribution function and  $M > 0$  is the precision parameter. Then the posterior process for  $F$  is the distribution of  $\mathbb{F}_n^* \sim \text{DP}(MG + n\mathbb{F}_n)$  and its BB approximation is given by  $\mathbb{F}_n^\# \sim \text{DP}(n\mathbb{F}_n)$ . With the above choices for  $a_n$  and  $v_n$  to define the functionals  $\psi_n$  and  $\zeta_n$ , the posterior distribution for  $\gamma$  is the distribution of either  $\psi_n(\mathbb{F}_n^*)$  or  $\zeta_n(\mathbb{F}_n^*)$ , and the BB distribution is the distribution of  $\psi_n(\mathbb{F}_n^\#)$  or  $\zeta_n(\mathbb{F}_n^\#)$ . The associated plug-in empirical estimator  $\psi_n(\mathbb{F}_n)$  or  $\zeta_n(\mathbb{F}_n)$  centers the limiting posterior and BB distributions. By sampling from the distributions of  $\psi_n(\mathbb{F}_n^*)$ ,  $\zeta_n(\mathbb{F}_n^*)$ ,  $\psi_n(\mathbb{F}_n^\#)$ , and  $\zeta_n(\mathbb{F}_n^\#)$ , we may obtain Bayesian (and BB) estimators as the posterior mean or median, and also posterior equal-tail credible intervals. The following result describes the computation of the BB distribution for  $\gamma$ , whose proof is evident from the definitions of the BB process and the functionals  $\psi_n$  or  $\zeta_n$ .

**Proposition 3.** *Let  $\mathbb{F}_n^\# = \sum_{i=1}^n W_i \delta_{X_{i:n}}$  where  $X_{1:n} < \dots < X_{n:n}$  are the order statistics of the data  $D_n = (X_1, \dots, X_n)$ , and  $\sum_{i=1}^n W_i = 1$  with  $W_i > 0$  for  $i = 1, \dots, n$ . Then,*

- (i)  $\bar{\mathbb{F}}_n^\#(x) = \sum_{i=j+1}^n W_i$  if  $X_{j:n} \leq x < X_{j+1:n}$  for some  $j = j(x)$ ;
- (ii)  $\int_{X_{j:n}}^{X_{j+1:n}} x^{-1} \bar{\mathbb{F}}_n^\#(x) dx = [\sum_{i=j+1}^n W_i] [\log X_{j+1:n} - \log X_{j:n}]$ ;
- (iii) if  $\omega < \infty$ , then

$$\int_{X_{j:n}}^{X_{j+1:n}} (\omega - x)^{-1} \bar{\mathbb{F}}_n^\#(x) dx = \left( \sum_{i=j+1}^n W_i \right) [\log(\omega - X_{j:n}) - \log(\omega - X_{j+1:n})]$$

(iv)  $\psi_n(\mathbb{F}_n^\#) = v_n^{-1}[1 - v_n^{-1}(N_1 + N_2 + N_3)/D]$ , where

$$\begin{aligned} N_1 &= [\sum_{i=r}^n W_i][\log X_{r:n} - \log a_n], \\ N_2 &= \sum_{j=r}^{s-1} [\sum_{i=j+1}^n W_i](\log X_{j+1:n} - \log X_{j:n}), \\ N_3 &= [\sum_{i=s+1}^n W_i](\log b_n - \log X_{s:n}), \\ D &= \sum_{i=r}^n W_i, \end{aligned}$$

and  $r$  and  $s$  are such that  $X_{r-1:n} < a_n \leq X_{r:n}$  and  $X_{s:n} \leq b_n < X_{s+1:n}$ ;

(v)  $\zeta_n(\mathbb{F}_n^\#) = \epsilon_n^{-1}[1 - \epsilon_n^{-1}(N_1 + N_2 + N_3)/D]$ , where

$$\begin{aligned} N_1 &= [\sum_{i=r}^n W_i][\log \epsilon_n - \log(\omega - X_{r:n})], \\ N_2 &= \sum_{j=r}^{n-2} [\sum_{i=j+1}^n W_i][\log(\omega - X_{j:n}) - \log(\omega - X_{j+1:n})], \\ N_3 &= 0, \\ D &= \sum_{i=r}^n W_i, \end{aligned}$$

and  $r$  is such that  $X_{r-1:n} \leq \omega - \epsilon_n < X_{r:n}$ .

The asymptotic behavior of the Dirichlet process posterior distribution (and, as a special case, the Bayesian bootstrap distribution) of the induced functional  $\psi_n$  or  $\zeta_n$ , is summarized nicely by the following theorem, the proof for which essentially mimic that of the earlier theorem on the asymptotic normality of plug-in estimator.

**Theorem 5** (Bernstein-von Mises theorem). *Under the same conditions in Theorem 4, conditional on the data  $D_n$ , almost surely  $[F_0]$ ,*

$$(i) \sqrt{n}\sigma_{n,\psi}^{-1}(\psi_n(\mathbb{F}_n^*) - \psi_n(\mathbb{F}_n)) \rightsquigarrow N(0, 1);$$

$$(ii) \sqrt{n}\sigma_{n,\zeta}^{-1}(\zeta_n(\mathbb{F}_n^*) - \zeta_n(\mathbb{F}_n)) \rightsquigarrow N(0, 1).$$

The two main theorems lead us to the following result on the coverage of Bayesian credible intervals for  $\gamma$ .

**Corollary 2** (Coverage). *Let  $Q_{\beta/2}$  and  $Q_{1-\beta/2}$  be the  $\beta/2$  and  $(1 - \beta/2)$ -posterior quantiles of  $\gamma$ . If the conditions in Theorem 4 are satisfied, then*

$$P_{F_0}(Q_{\beta/2} \leq \gamma_0 \leq Q_{1-\beta/2}) \rightarrow 1 - \beta. \quad (2.21)$$

## 2.4 Numerical Results

In this section, we study the accuracy of the Bayesian bootstrap (BB) estimator based on the functionals  $\psi_n$  and  $\zeta_n$  and the coverage of the BB credible interval for various true distributions in the case when  $\gamma \neq 0$ . We also present results from an implementation of the modified Bayes factor developed for the Gumbel MDA.

For distributions in the Fréchet MDA ( $\gamma > 0$ ), the specific simulation steps, implemented in R, are outlined below.

- Set a random sample size  $n$ .
- Set  $B$  as the number of random samples to be generated from a distribution in the Fréchet MDA. Perform the following loop  $B$  times.
  - Obtain a random sample  $X_1, \dots, X_n$  from  $F_0$ , where  $F_0$  is one of the following: Fréchet, Pareto, Student t, Snedecor F.
  - Set  $p_n = n^{-\tau}$ .
  - Evaluate  $\hat{a}_n := X_{[n(1-p_n)]}$ .
  - Set  $v_n = n^{-\xi}$ .
  - Calculate  $\hat{b}_n = \hat{a}_n * (1 + v_n)$
  - Set  $C$  as the number of iterations to be performed (for the Bayesian bootstrap). Perform the following loop  $C$  times.
    - \* Generate a random sample  $E_1, \dots, E_n$  from a standard exponential distribution (independently from  $X_1, \dots, X_n$ ).
    - \* Generate a Dirichlet distribution from the standard exponential random sample.
    - \* Apply (3) to determine  $\psi_n(\mathbb{F}_n^\#)$  using  $\hat{a}_n$  for  $a_n$  and  $\hat{b}_n$  for  $b_n$ .

The corresponding simulation steps for distributions in the Reverse Weibull MDA ( $\gamma < 0$ ), are similar to the above, but with the following differences:

- The hyperparameter  $\epsilon_n$  is set to  $n^{-\xi'}$ .
- $F_0$  in this case would be uniform, beta, or triangular.
- The parameter  $\omega$  is estimated with  $\omega = X_{n:n}$ .

We have used  $B = 10,000$ , and  $C = 100$  as the number of random samples and number of Bayesian bootstrap iterations, respectively. In addition, we set  $\tau = 1/4$ ,  $\xi = 1/2$ , and  $\xi' = 1/5$ .

Given random sample data, each Bayesian bootstrap simulation yields a Dirichlet distribution of  $\gamma_0$ , and the sample average of the  $[2\psi_n(\mathbb{F}_n^\#) - 1]^{-1}$  or  $-[2\zeta_n(\mathbb{F}_n^\#) + 1]^{-1}$  values across the  $C$  simulations was taken as the Bayes estimator for  $\gamma_0$ .

We calculated coverage, the proportion of the  $B$  credible intervals (from the  $B$  random samples) that contain the true value  $\gamma_0$ . Since we calculate 95% credible intervals using Monte Carlo methods, the closer the coverage ratio is to 95%, the better.

In the following tables, we present three metrics for various distributions under the Fréchet and Reverse Weibull MDAs, for select parameter values. We show the coverage, mean credible interval length, and the bias associated with the Bayes estimator. For comparison, we calculate similar metrics using the Hill/Negative Hill and Pickands estimators. For the implementation of these estimators, the integer floor of  $\log(n)$  is chosen as the  $k(n)$  that satisfies  $k(n) \rightarrow \infty$  and  $k(n)/n \rightarrow 0$ . We show results for the Bayes (BA), Pickands (PI), and Hill/Negative Hill (HI) for various tests.

Table 2.1: Selected EVI inference simulation results for Pareto distribution with parameters  $(t, \alpha)$ .

$(t, \alpha)$	$n$	Coverage			Length			Bias		
		BA	PI	HI	BA	PI	HI	BA	PI	HI
(3,2)	500	0.95	0.94	0.88	1.06	3.19	0.80	-2.34	-0.07	-0.26
(3,2)	5000	0.96	0.94	0.90	0.58	2.75	0.70	-1.05	-0.08	-0.12
(3,2)	50 000	0.96	0.94	0.90	0.48	2.45	0.62	-1.03	-0.81	0.13
(4,3)	500	0.93	0.95	0.88	0.49	3.11	0.53	-0.94	-0.06	-0.17
(4,3)	5000	0.95	0.95	0.90	0.31	2.68	0.46	-0.72	-0.07	-0.08
(4,3)	50 000	0.97	0.95	0.90	0.26	2.39	0.41	-0.68	-0.78	0.09

The Pareto distribution exceeds the 95% threshold for a large class of shape parameters. The mean credible interval is also shorter than the corresponding confidence intervals as  $n$  increases.

As the degrees of freedom increase, the coverage of the t-distribution decreases. This reflects the loss of "heavy-tailedness" of the t distribution as it converges to the lighter-tailed normal distribution, which is not in the Fréchet MDA.

Table 2.2: Selected EVI inference simulation results for Student’s  $t$ -distribution with  $\nu$  degrees of freedom.

$\nu$	$n$	Coverage			Length			Bias		
		BA	PI	HI	BA	PI	HI	BA	PI	HI
1	500	0.91	0.92	0.88	9.27	3.51	1.61	-1.93	-1.04	0.17
1	5000	0.95	0.93	0.90	3.14	3.03	1.39	-2.18	-0.55	0.69
1	50 000	0.92	0.93	0.90	1.58	2.70	1.24	-2.14	1.28	0.02
2	500	0.70	0.94	0.89	6.19	3.18	0.82	-1.21	0.86	-0.08
2	5000	0.83	0.94	0.90	1.28	2.75	0.70	-0.90	0.11	0.15
2	50 000	0.83	0.94	0.91	0.68	2.45	0.62	-1.03	-0.68	0.21

Table 2.3: Selected EVI inference simulation results for Fréchet Distribution with parameters (loc, scale, shape).

(loc, scale, shape)	$n$	Coverage			Length			Bias		
		BA	PI	HI	BA	PI	HI	BA	PI	HI
(0,1,2)	500	0.95	0.94	0.88	1.39	3.19	0.80	-5.36	-0.08	-0.26
(0,1,2)	5000	0.95	0.94	0.90	0.65	2.75	0.70	-1.08	-0.08	-0.12
(0,1,2)	50 000	0.95	0.94	0.90	0.50	2.45	0.62	-1.05	-0.81	0.13
(0,1,3)	500	0.90	0.95	0.88	0.60	3.11	0.53	-1.11	-0.07	-0.17
(0,1,3)	5000	0.95	0.95	0.90	0.34	2.68	0.46	-0.74	-0.08	-0.08
(0,1,3)	50 000	0.96	0.95	0.90	0.27	2.39	0.41	-0.69	-0.78	0.09

Similar to the Pareto distribution, the Fréchet distribution meets the 95% threshold for a large class of shape parameters. The mean credible interval is also generally shorter than the corresponding confidence intervals under the Pickands and Hill estimators.

We observe similarities between the  $t$ -distribution and F-distribution in terms of coverage. As  $\nu_2$  increases, the F-distribution converges to the lighter-tailed chi-squared (gamma) distribution, which is not in the Fréchet MDA.

Next, we present simulation results using the functional  $\zeta_n$  for distributions in the reverse Weibull MDA. Here, we show the metric results for random sample sizes  $n = 500$ ,  $n = 1,000$  or  $n = 5,000$  for the uniform, beta, and triangular distributions.

In general, Bayesian coverage results are not as close to the threshold 95% as those for the Fréchet MDA distributions. For the distributions in the reverse Weibull MDA, we used the Negative Hill estimator (Falk 1995) instead of the Hill estimator, which was defined only for  $\gamma > 0$ . Generally, the simulation results under the Negative Hill estimator, which was intended to be applicable when  $\gamma < -1/2$ , are comparatively worse relative to either the

Table 2.4: Selected EVI inference simulation results for Snedecor’s  $F$ -distribution with  $\nu_1, \nu_2$  degrees of freedom.

$(\nu_1, \nu_2)$	$n$	Coverage			Length			Bias		
		BA	PI	HI	BA	PI	HI	BA	PI	HI
(1,2)	500	0.90	0.92	0.89	10.08	3.52	1.64	-2.79	-0.09	-0.02
(1,2)	5000	0.96	0.93	0.89	3.93	3.02	1.38	-2.97	2.04	-0.02
(1,2)	50 000	0.88	0.93	0.90	1.81	2.70	1.24	-2.39	0.42	0.16
(2,2)	500	0.93	0.92	0.88	8.01	3.51	1.62	-1.98	0.07	0.11
(2,2)	5000	0.95	0.92	0.89	3.45	3.02	1.38	-1.76	0.11	0.29
(2,2)	50 000	0.90	0.93	0.90	1.72	2.70	1.24	-1.77	0.15	-0.14

Table 2.5: Selected EVI inference simulation results for uniform distribution on  $[a, b]$ ,  $a < b$ .

$[a, b]$	$n$	Coverage			Length			Bias		
		BA	PI	HI	BA	PI	HI	BA	PI	HI
[0, 0.5]	500	0.85	0.98	0.08	0.49	2.97	0.48	0.03	-0.07	1.03
[0, 0.5]	1000	0.88	0.98	0.05	0.41	2.97	0.54	0.02	-0.06	1.16
[0, 0.5]	5000	0.89	0.97	0.05	0.25	2.55	0.37	0.03	-0.06	1.02
[0, 0.75]	500	0.84	0.98	0.10	0.59	2.97	0.47	0.04	-0.07	0.96
[0, 0.75]	1000	0.83	0.98	0.06	0.48	2.97	0.40	0.06	-0.06	1.09
[0, 0.75]	5000	0.79	0.97	0.06	0.25	2.55	0.37	0.04	-0.06	0.97

Bayes estimator or the Pickands estimator.

For the uniform distribution, the confidence interval associated with the Pickands estimator appears to achieve the best coverage among the three estimators. However, it does so by using significantly longer intervals than the Bayesian and Hill estimator-based intervals, which are similar in length. Furthermore, the Bayes estimator has the lowest bias, followed closely by the Pickands estimator.

We observe that there is a range of parameter values for which the Bayesian bootstrap technique works well in terms of coverage. For example,  $\alpha, \beta$  parameters in the beta distribution that allow for higher values to be sampled (densities that are not sloping down near 1) yield coverage results that are closer to the target credibility level.

Of all the distributions considered in this paper, the triangular distribution appears to be the most challenging distribution to achieve the target credibility level. The triangular density, with a linear drop-off from the mode to the maximum, does not allow sufficient sampling of extreme values near the maximum. This, in turn, affected the simulation results.

Table 2.6: Selected EVI inference simulation results for beta distribution with parameters  $(\alpha, \beta)$ .

$(\alpha, \beta)$	$n$	Coverage			Length			Bias		
		BA	PI	HI	BA	PI	HI	BA	PI	HI
(1.5,1.0)	500	0.85	0.98	0.09	0.66	2.98	0.48	-0.05	-0.07	0.98
(1.5,1.0)	1000	0.88	0.98	0.06	0.52	2.98	0.51	-0.02	-0.07	1.10
(1.5,1.0)	5000	0.87	0.97	0.05	0.29	2.54	0.36	-0.02	-0.05	0.99
(1.2,1.1)	500	0.81	0.98	0.12	0.57	2.96	0.45	0.03	-0.08	0.84
(1.2,1.1)	1000	0.80	0.98	0.08	0.43	2.95	0.47	0.04	-0.06	0.94
(1.2,1.1)	5000	0.82	0.97	0.08	0.26	2.53	0.34	0.01	-0.05	0.85

Table 2.7: Selected EVI inference simulation results for triangular distribution with parameters  $a < b < c$ .

$(a, b, c)$	$n$	Coverage			Length			Bias		
		BA	PI	HI	BA	PI	HI	BA	PI	HI
(0.0,0.1,1.0)	500	0.84	0.98	0.30	0.38	2.92	0.41	-0.04	-0.09	0.33
(0.0,0.1,1.0)	1000	0.83	0.98	0.24	0.30	2.93	0.38	-0.02	-0.08	0.39
(0.0,0.1,1.0)	5000	0.86	0.97	0.29	0.24	2.50	0.33	-0.01	-0.07	0.32
(0,1,2)	500	0.77	0.98	0.37	0.69	2.92	0.46	-0.06	-0.09	0.26
(0,1,2)	1000	0.75	0.98	0.31	0.44	2.93	0.41	-0.02	-0.08	0.32
(0,1,2)	5000	0.79	0.97	0.34	0.33	2.50	0.37	0.01	-0.07	0.27

In this last subsection, we show the results of the modified Bayes factor testing for the distributions in the Gumbel MDA. Recall that we calculate a modified Bayes factor between two models  $M_0$  and  $M_1$ , which are both underlined by a Dirichlet distribution. Model  $M_1$  also reflects an additional constraint that is treated as an approximate equivalent of the limiting condition for distributions in the Gumbel MDA. A Stan code file containing model  $M_1$  was called in R to perform Hamiltonian Monte Carlo (HMC) sampling. The high-level steps performed within the R or Stan codes are as follows:

- Generate  $B$  sets of ordered statistics from  $n$  simulated observations, each from the normal, logistic, gamma, Weibull, Fréchet, Pareto, Student's t, Snedecor's F, uniform, triangular, and beta.
- Loop through various combinations of hyperparameters  $(\epsilon, \alpha, r)$  as described in (2.8) and (2.9). Within each loop, perform the following:
  - Using the compiled constrained Stan model, draw samples for the constrained

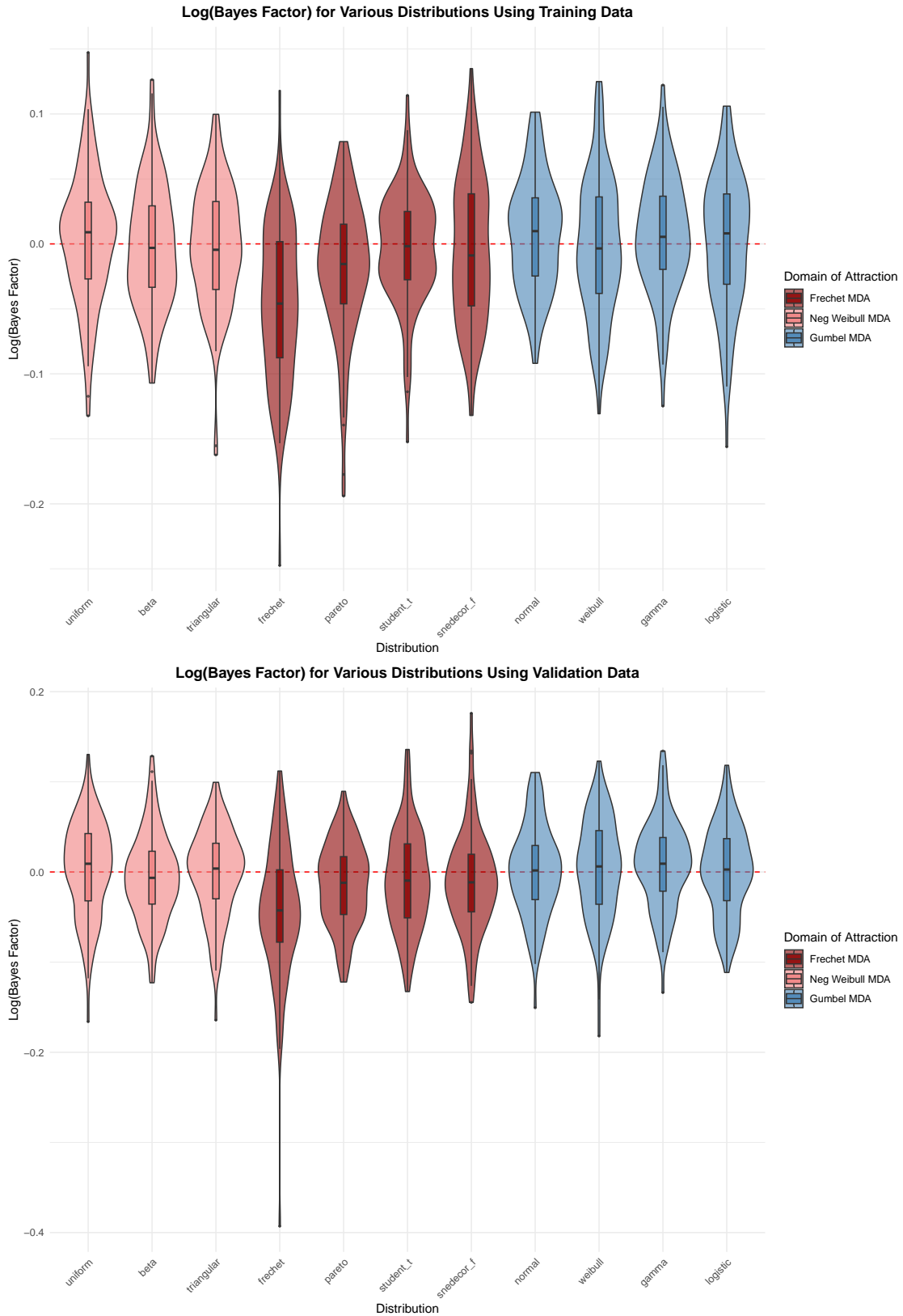
weights satisfying (2.8).

- Compute the Bayes factor by dividing the harmonic mean of the constrained Dirichlet likelihood by the analytic marginal likelihood in (2.9).
  - Calculate a score based on the classification accuracy and mean of the logarithm of the modified Bayes factor ( $\log(\text{BF})$ ), applying penalties to discourage solutions that rely on consistently weak evidence that barely crosses the zero line.
  - Identify the hyperparameter combination that has the best score.
- Regenerate new sets of ordered statistics for the 11 distributions.
  - Using the winning hyperparameter combination, calculate the  $\log(\text{BF})$  for the newly generated data.
  - Plot the logarithm of the Bayes factors for the new sample replications using a box plot or a violin plot.

The box or violin plots show the  $\log(\text{BF})$  values for the various distributions considered in this paper. We use the power  $\alpha_n = 1 - n^{-\alpha}$  and run  $B = 100$  replications with sample size  $n = 50$ .

The target distributions are those belonging to the Gumbel MDA. As explained in the high-level steps, a run using training data was initiated to obtain optimal hyperparameters. The result of this run is shown as the top plot in Figure 2.1. Here, we observe that the mean  $\log(\text{BF})$  values are mostly above 0 for the target distributions, while mean  $\log(\text{BF})$  values are mostly negative for the non-target distributions. The  $\log(\text{BF})$  being positive, or equivalently the modified Bayes factor being greater than 1, can be interpreted as providing evidence supporting the hypothesis that the distribution  $F$  belongs to a Gumbel MDA. After the training data run, a succeeding run utilizing the optimal hyperparameters and an independent validation data was processed. The results of this run is shown as the bottom plot in Figure 2.1. The results of the validation data run suggests that the hyperparameters obtained in the training step were robust enough to produce largely similar results obtained from the training data run.

Figure 2.1: Modified Bayes factors for various distributions using training and validation data,  $n = 50$ ,  $B = 100$ .



## 2.5 Proofs

*Proof of Proposition 1.*

**Case 1: Pareto.** We have  $\bar{F}(x) = x^{-\alpha}$  for  $x \geq x_m$  for some  $x_m \in \mathbb{R}$  and  $\alpha > 0$ . Then since  $\gamma = 1/\alpha$ ,

$$\frac{\bar{F}(a_n z)}{\bar{F}(a_n)} - z^{-\alpha} = \frac{a_n^{-\alpha} z^{-\alpha}}{a_n^{-\alpha}} - z^{-\alpha} = 0$$

. Thus  $m_n = 0$ .

**Case 2: Fréchet.** We have  $\bar{F}(x) = 1 - \exp(-x^{-\alpha})$  for  $x > 0$  and  $\alpha > 0$ . Using  $e^x = 1 + x + O(x^2)$ , we obtain using  $\gamma = 1/\alpha$ ,

$$\left| \frac{\bar{F}(a_n z)}{\bar{F}(a_n)} - z^{-\alpha} \right| = \left| \frac{a_n^{-\alpha} z^{-\alpha} + O(a_n^{-2\alpha})}{a_n^{-\alpha} + O(a_n^{-2\alpha})} - z^{-\alpha} \right| = O(a_n^{-\alpha}).$$

Thus  $m_n := a_n^{-\alpha} \rightarrow 0$  as  $n \rightarrow \infty$ .

**Case 3: Student's t.** We write the density of the t-distribution as

$$f(t) = K_1 (1 + t^2/\nu)^{-(\nu+1)/2} = K_2 t^{-\nu+1} \{1 + K_3 t^{-2} + O(t^{-4})\}$$

for some constants  $K_1$ ,  $K_2$ , and  $K_3$ . Hence

$$\bar{F}(x) = \int_x^\infty K_2 t^{-(\nu+1)} \{1 + K_3 t^{-2} + O(t^{-4})\} dt = K_4 x^{-\nu} + K_5 x^{-\nu-2} + O(x^{-\nu-4})$$

for some constants  $K_4$  and  $K_5$ . This implies

$$\left| \frac{\bar{F}(a_n z)}{\bar{F}(a_n)} - z^{-\nu} \right| = \left| \frac{K_4 a_n^{-\nu} z^{-\nu} + K_5 a_n^{-\nu} z^{-\nu} + O(a_n^{-\nu-4} z^{-\nu-4})}{K_4 a_n^{-\nu} + K_5 a_n^{-\nu-2} + O(a_n^{-\nu-4})} - z^{-\nu} \right| = O(a_n^{-4}).$$

Thus  $m_n := a_n^{-4} \rightarrow 0$  as  $n \rightarrow \infty$ .

**Case 4: Snedecor's F.** Similar to the t-distribution, we write the density as

$$f(t) = K_1 \frac{t^{\nu_1/2-1}}{(\nu_1 t + \nu_2)^{(\nu_1+\nu_2)/2}} = K_2 t^{-\nu_2/2-1} + K_3 t^{-\nu_2/2-2} + O(t^{-\nu_2/2-3})$$

for some constants  $K_1$ ,  $K_2$ , and  $K_3$ . Hence

$$\bar{F}(x) = \int_x^\infty f(t) dt = K_4 x^{-\nu_2/2} + K_5 x^{-\nu_2/2-1} + O(x^{-\nu_2/2-2})$$

for some constants  $K_4$  and  $K_5$ . Thus

$$\left| \frac{\bar{F}(a_n z)}{\bar{F}(a_n)} - z^{-\nu_2/2} \right| = \left| \frac{K_4(a_n z)^{-\nu_2/2} + K_5(a_n z)^{-\nu_2/2-1} + O(a_n^{-\nu_2/2-2})}{K_4 a_n^{-\nu_2/2} + K_5 a_n^{-\nu_2/2-1} + O(a_n^{-\nu_2/2-2})} - z^{-\nu_2/2} \right|$$

is  $O(a_n^{-\nu_2/2-1})$ .

**Case 5: Uniform.** We have for  $a \leq x \leq \omega$ ,  $\bar{F}(x) = (\omega - x)/(\omega - a)$ . Hence

$$\left| \frac{\bar{F}(\omega - \epsilon_n z)}{\bar{F}(\omega - \epsilon_n)} - z \right| = \left| \frac{\epsilon_n z}{\epsilon_n} - z \right| = 0,$$

so that  $m_n = 0$ .

**Case 6: Beta.** Write  $K$  for the normalizing constant of the beta distribution with parameters  $\alpha$  and  $\beta$ . We prove for  $\alpha > 1$ ; the case  $\alpha < 1$  is similar. We have for  $0 \leq x \leq \omega = 1$  and  $x$  near  $\omega$ , we have

$$\begin{aligned} K' x^{\alpha-1} (1-x)^\beta &\leq \int_x^1 K x^{\alpha-1} (1-t)^{\beta-1} dt \\ &\leq \bar{F}(x) \\ &= \int_x^\omega K t^{\alpha-1} (1-t)^{\beta-1} dt \leq K'(1-x)^\beta, \end{aligned}$$

where  $K' = K/\beta$ , leading to the estimate

$$z^\beta \{(1 - \epsilon_n z)^{-(\alpha-1)} - 1\} \frac{\bar{F}(1 - \epsilon_n z)}{\bar{F}(1 - \epsilon_n)} - z^\beta \leq z^\beta \{(1 - \epsilon_n)^{-(\alpha-1)} - 1\} \quad (2.22)$$

By the binomial expansion near 0, the factors with  $z^\beta$  on both sides of (2.22) are of the order of  $\epsilon_n$ . Thus, the condition holds with  $\gamma = -1/\beta$  and  $m_n = \epsilon_n$ .

**Case 7: Triangular.** For  $x$  close to  $\omega = c$ ,  $\bar{F}(x) = K(\omega - x)^2$  for some constant  $K$ , so  $\bar{F}(\omega - \epsilon_n z)/\bar{F}(\omega - \epsilon_n) = z^2$  exactly. Hence,  $\gamma = -1/2$  and  $m_n = 0$ .  $\square$

*Proof of Proposition 2.* By assumption, there exists  $K > 0$  and  $N_1 > 0$  such that if  $n > N_1$  then for all  $z > 0$ ,

$$z^{-1/\gamma} - Km_n \leq \frac{\bar{F}(a_n z)}{\bar{F}(a_n)} \leq z^{-1/\gamma} + Km_n.$$

Integrating over  $[1, 1 + v_n]$  we obtain

$$\begin{aligned} \int_1^{1+v_n} [z^{-1/\gamma} - Km_n] dz &\leq \int_1^{1+v_n} \frac{\bar{F}(a_n z)}{z\bar{F}(a_n)} dz \leq \int_1^{1+v_n} [z^{-1/\gamma} + Km_n] dz \\ -\gamma[1 - (1 + v_n)^{-1/\gamma}] - Km_n v_n &\leq \int_1^{1+v_n} \frac{\bar{F}(a_n z)}{z\bar{F}(a_n)} dz \leq -\gamma[1 - (1 + v_n)^{-1/\gamma}] + Km_n v_n \end{aligned}$$

Applying the generalized binomial theorem, we obtain

$$\frac{1}{2}(1 + 1/\gamma) + O(v_n) - K \frac{m_n}{v_n} \leq \psi_n(F) \leq \frac{1}{2}(1 + 1/\gamma) + O(v_n) + K \frac{m_n}{v_n}.$$

Hence

$$\psi_n(F) = \frac{1}{2}(1 + 1/\gamma) + O(v_n + m_n/v_n).$$

The case for the Reverse Weibull MDA can be proved similarly.  $\square$

*Proof of Lemma 1.* We compute the derivative of  $\psi_n$  as follows:

$$\begin{aligned} \psi'_{n,F}(H) &= \lim_{t \rightarrow 0} \frac{\frac{1}{v_n} \left\{ 1 - \frac{1}{v_n} \frac{\int_{a_n}^{b_n} x^{-1} [\bar{F}(x) - tH(x)] dx}{\bar{F}(a_n) - tH(a_n)} \right\} - \frac{1}{v_n} \left\{ 1 - \frac{1}{v_n} \frac{\int_{a_n}^{b_n} x^{-1} \bar{F}(x) dx}{\bar{F}(a_n)} \right\}}{t} \\ &= \frac{d}{dt} \left\{ \frac{1}{v_n} - \frac{1}{v_n^2} \frac{\int_{a_n}^{b_n} x^{-1} [\bar{F}(x) - tH(x)] dx}{\bar{F}(a_n) - tH(a_n)} \right\} \Bigg|_{t=0} \\ &= -\frac{1}{v_n^2} \frac{-\bar{F}(a_n) \int_{a_n}^{b_n} x^{-1} H(x) dx + H(a_n) \int_{a_n}^{b_n} x^{-1} \bar{F}(x) dx}{[\bar{F}(a_n)]^2} \end{aligned}$$

The derivative of  $\zeta_n$  can be calculated similarly.  $\square$

*Proof of Lemma 2.* From the first derivative calculation for  $\psi_n$  we have

$$\phi'_{n,F,H,\psi}(t) = -\frac{1}{v_n^2} \frac{-\bar{F}(a_n) \int_{a_n}^{b_n} x^{-1} H(x) dx + H(a_n) \int_{a_n}^{b_n} x^{-1} \bar{F}(x) dx}{[\bar{F}(a_n) - tH(a_n)]^2}.$$

Hence

$$\phi''_{n,F,H,\psi}(t) = -\frac{1}{v_n^2} \frac{-2\bar{F}(a_n)H(a_n) \int_{a_n}^{b_n} x^{-1} H(x) dx + 2[H(a_n)]^2 \int_{a_n}^{b_n} x^{-1} \bar{F}(x) dx}{[\bar{F}(a_n) - tH(a_n)]^3}.$$

The second derivative for  $\zeta_n$  can be calculated similarly.  $\square$

*Proof of Lemma 3.* To calculate the variance  $\sigma_{n,\psi}^2$ , we first note the following properties of Kiefer processes:

- $E[K_1^2(F_0(a_n), n)] = nF_0(a_n)\bar{F}_0(a_n)$ ;
- $E\left[\int_{a_n}^{b_n} x^{-1}K_1(F_0(a_n), n)K_1(F_0(x), n) dx\right] = nF_0(a_n)\int_{a_n}^{b_n} x^{-1}\bar{F}_0(x) dx$ ;
- $E\left[\int_{a_n}^{b_n} x^{-1}K_1(F_0(x), n) dx\right]^2 = \int_{a_n}^{b_n}\int_{a_n}^{b_n}(xy)^{-1}[F_0(x)\wedge F_0(y) - F_0(x)F_0(y)] dx dy$ .

We also note that since  $\psi'_{n,F_0}$  is linear in its argument and that the mean of the normal random variables underlying a Kiefer process is 0,

$$E\left[\psi'_{n,F_0}\left(n^{-1/2}K_1(F_0(\cdot), n)\right)\right] = 0.$$

Then we have  $\psi'_{n,F_0}\left(n^{-1/2}K_1(F_0(\cdot), n)\right)$  is normally distributed with variance

$$\begin{aligned}\sigma_{n,\psi}^2 &= E[\psi'_{n,F_0}(n^{-1/2}K_1(F_0(\cdot), n))^2] \\ &= n^{-1}E\left\{v_n^{-4}[\bar{F}_0(a_n)]^{-4}\left[[\bar{F}_0(a_n)]^2\left(\int_{a_n}^{b_n} x^{-1}K_1(F_0(x), n) dx\right)^2\right. \right. \\ &\quad \left. \left.- 2\bar{F}_0(a_n)K_1(F_0(a_n), n)\int_{a_n}^{b_n} x^{-1}\bar{F}_0(x) dx\int_{a_n}^{b_n} x^{-1}K_1(\bar{F}_0(x), n) dx\right. \right. \\ &\quad \left. \left. + K_1^2(F_0(a_n), n)\left(\int_{a_n}^{b_n} x^{-1}\bar{F}_0(x) dx\right)^2\right]\right\} \\ &= v_n^{-4}[\bar{F}_0(a_n)]^{-2}\int_{a_n}^{b_n}\int_{a_n}^{b_n}(xy)^{-1}[F_0(x)\wedge F_0(y) - F_0(x)F_0(y)] dx dy \\ &\quad - v_n^{-4}[\bar{F}_0(a_n)]\left(\int_{a_n}^{b_n} x^{-1}\bar{F}_0(x) dx\right)^2 F_0(a_n).\end{aligned}$$

For large  $n$ , let  $\bar{F}_0(a_n) \asymp p_n$ . Also for large  $x$  and  $y$ ,

$$F_0(x)\wedge F_0(y) - F_0(x)F_0(y) \asymp (1-p_n) - (1-p_n)^2 \sim p_n$$

since  $F_0(a_n) \rightarrow 1$ .

The first term in the variance is approximately proportional to  $v_n^{-4}p_n^{-2}(p_na_n^{-2})(v_n^2a_n^2) = v_n^{-2}p_n^{-1}$ .

The second term in the variance is approximately proportional to  $v_n^{-4}p_n^{-3}[(p_na_n^{-1})(v_na_n)]^2 = v_n^{-2}p_n^{-1}$ . Thus  $\sigma_{n,\psi}$  is approximately proportional to  $v_n^{-1}p_n^{-1/2}$ .

A similar proof applies to  $\zeta_n$ . □

*Proof of Theorem 4.* First, from the earlier Taylor approximation, we have

$$\begin{aligned} & \frac{\sqrt{n}}{\sigma_{n,\psi}} [\psi_n(\mathbb{F}_n) - \psi_n(F_0)] \\ &= \frac{1}{\sigma_{n,\psi}} \psi'_{n,F_0}(n^{-1/2} K_1(F_0(\cdot), n)) + \frac{1}{\sigma_{n,\psi}} \psi'_{n,F_0}(\kappa_n) + \frac{1}{2\sigma_{n,\psi}\sqrt{n}} \phi''_{n,F_0,\mathbb{J}_n,\psi}(\eta_n). \end{aligned}$$

Since the first term on the right is normal for every  $n$ , it suffices to show that the last two terms tend to 0 almost surely. Recalling that  $\sup\{|\kappa_n(x)| : x \in \mathbb{R}\} = O(n^{-1/2} \log^2 n)$  a.s. as  $n \rightarrow \infty$ , we have

$$\begin{aligned} & \frac{1}{\sigma_{n,\psi}} \psi'_{n,F_0}(\kappa_n) \\ &= \frac{v_n^{-2}}{\sigma_{n,\psi}} \left\{ \frac{\bar{F}_0(a_n) \int_{a_n}^{b_n} x^{-1} \kappa_n(x) dx - \kappa_n(a_n) \int_{a_n}^{b_n} x^{-1} \bar{F}_0(x) dx}{[\bar{F}_0(a_n)]^2} \right\} \\ &\asymp v_n \sqrt{p_n} v_n^{-2} [(p_n a_n^{-1})(n^{-1/2} \log^2 n)(v_n a_n) + (n^{-1/2} \log^2 n)(p_n a_n^{-1})(v_n a_n)] p_n^{-2} \\ &\asymp \frac{\log^2 n}{\sqrt{n p_n}} \rightarrow 0 \end{aligned}$$

a.s. as  $n \rightarrow \infty$  if  $p_n \asymp n^{-\tau}$  and  $0 < \tau < 1$ .

Recalling that  $\sup\{|\mathbb{J}_n(x)| : x \in \mathbb{R}\} = O(\sqrt{\log \log n})$  a.s. as  $n \rightarrow \infty$ , we have

$$\begin{aligned} & \frac{1}{2\sigma_{n,\psi}\sqrt{n}} \phi''_{n,F_0,\mathbb{J}_n,\psi}(\eta_n) \\ &\asymp \frac{v_n \sqrt{p_n}}{\sqrt{n}} \left( \frac{1}{v_n^2} \right) \left[ p_n \sqrt{\log \log n} \frac{\sqrt{\log \log n}}{a_n} v_n a_n + (\log \log n) \frac{p_n}{a_n} v_n a_n \right] \frac{1}{p_n^3} \\ &\asymp \frac{\log \log n}{\sqrt{n p_n}^{3/2}} \rightarrow 0 \end{aligned}$$

a.s. as  $n \rightarrow \infty$  if  $p_n \asymp n^{-\tau}$  and  $0 < \tau < 1/3$ .

In addition to consistency, we want to ensure that bias is controlled, taking into account the previously calculated standard deviation. That is, we also want to show that

$$\frac{\sqrt{n}}{\sigma_{n,\psi}} \left[ \psi_n(F_0) - \frac{1 + 1/\gamma_0}{2} \right] \rightarrow 0.$$

Using Proposition 2, we then have

$$\frac{\sqrt{n}}{\sigma_{n,\psi}} \left[ \psi_n(F_0) - \frac{1 + 1/\gamma_0}{2} \right] \asymp \sqrt{n} (v_n \sqrt{p_n}) \left( v_n + \frac{m_n}{v_n} \right) \asymp n^{1/2-\tau/2-2\xi} + n^{1/2-\tau/2} m_n,$$

which tends to 0 as  $n \rightarrow \infty$  if  $m_n = o(n^{\tau/2-1/2})$  and  $1/2 - \tau/2 - 2\xi < 0$ . Combining this last inequality with the restriction  $\tau < 1/3$  leads to  $\xi > 1/6$ . The  $\zeta_n$  case can be proved similarly.  $\square$

*Proof of Theorem 5.* Since  $\mathbb{J}_n(x) := \sqrt{n}(\mathbb{F}_n(x) - F_0(x))$  and  $\mathbb{J}_n^\#(x) := \sqrt{n}(\mathbb{F}_n^\#(x) - F_0(x))$ , we have

$$\mathbb{F}_n^\# = \mathbb{F}_n + n^{-1/2}\mathbb{J}_n = F_0 + n^{-1/2}(\mathbb{J}_n + \mathbb{J}_n^\#).$$

Applying a second Taylor expansion of  $\psi_n$  around  $F_0$  we obtain

$$\frac{\sqrt{n}}{\sigma_{n,\psi}}[\psi_n(\mathbb{F}_n^\#) - \psi_n(F_0)] = \frac{1}{\sigma_{n,\psi}}\psi'_{n,F_0}(\mathbb{J}_n + \mathbb{J}_n^\#) + \frac{\phi''_{n,F_0,\mathbb{J}_n+\mathbb{J}_n^\#,\psi}(\eta_n)}{2\sigma_{n,\psi}\sqrt{n}}.$$

Recalling that

$$\frac{\sqrt{n}}{\sigma_{n,\psi}}[\psi_n(\mathbb{F}_n) - \psi_n(F_0)] = \frac{1}{\sigma_{n,\psi}}\psi'_{n,F_0}(\mathbb{J}_n) + \frac{\phi''_{n,F_0,\mathbb{J}_n,\psi}(\eta_n)}{2\sigma_{n,\psi}\sqrt{n}},$$

we thus obtain

$$\begin{aligned} \frac{\sqrt{n}}{\sigma_{n,\psi}}[\psi_n(\mathbb{F}_n^\#) - \psi_n(\mathbb{F}_n)] &= \frac{1}{\sigma_{n,\psi}}\psi'_{n,F_0}(\mathbb{J}_n^\#) + \frac{\phi''_{n,F_0,\mathbb{J}_n+\mathbb{J}_n^\#,\psi}(\eta_n)}{2\sigma_{n,\psi}\sqrt{n}} - \frac{\phi''_{n,F_0,\mathbb{J}_n,\psi}(\eta_n)}{2\sigma_{n,\psi}\sqrt{n}} \\ &= \frac{1}{\sigma_{n,\psi}}\psi'_{n,F_0}(n^{-1/2}K_2(F_0(\cdot), n)) + \frac{1}{\sigma_{n,\psi}}\psi'_{n,F_0}(\delta_n) \\ &\quad + \frac{\phi''_{n,F_0,\mathbb{J}_n+\mathbb{J}_n^\#,\psi}(\xi_n)}{2\sigma_{n,\psi}\sqrt{n}} - \frac{\phi''_{n,F_0,\mathbb{J}_n,\psi}(\eta_n)}{2\sigma_{n,\psi}\sqrt{n}}. \end{aligned}$$

We proceed similarly as before; we show that the first term is standard normal and that the remaining terms are stochastically bounded. The first term is indeed standard normal, and the last term was shown to go to 0 a.s. as  $n \rightarrow \infty$ . It remains to show that the second and third terms also go to 0 almost surely.

For the second term, using the earlier argument for  $\sigma_{n,\psi}^{-1}\psi'_{n,F_0}(\kappa_n)$  and noting that  $\sup\{|\delta_n(x)| : x \in \mathbb{R}\} = O(n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4})$  a.s., we obtain

$$\frac{1}{\sigma_n}\psi'_{n,F_0}(\delta_n) = O\left(\frac{n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}}{p_n^{1/2}}\right) \rightarrow 0,$$

a.s. if  $\tau < 1/2$ .

For the third term, using the earlier argument for  $\frac{\phi''_{n,F_0,\mathbb{J}_n,\psi}(\eta_n)}{2\sigma_{n,\psi}\sqrt{n}}$  we obtain

$$\frac{\phi''_{n,F_0,\mathbb{J}_n,\psi}(\eta_n)}{2\sigma_{n,\psi}\sqrt{n}} = O\left(\frac{\log \log n}{\sqrt{np_n}^{3/2}}\right) \rightarrow 0,$$

a.s. if  $0 < \tau < 1/3$ .

The Reverse Weibull MDA case can be proved in a similar manner. □

*Proof of Corollary 2.* The event  $\gamma_0 \in [Q_{\beta/2}, Q_{1-\beta/2}]$  holds if and only if

$$\beta/2 \leq \Pi(\psi_n(\mathbb{F}_n^*) \leq \gamma_0 | D_n) \leq 1 - \beta/2.$$

By Theorem 4, the probability, under the true distribution  $F_0$  tends to  $1 - \beta$  if the assumed conditions hold. The same justification holds for  $\zeta_n$ . □

## CHAPTER

### 3

# BAYESIAN INFERENCE ON THE WEIBULL TAIL COEFFICIENT

## 3.1 Introduction

In this chapter, we focus on the inference of the Weibull tail coefficient (WTC) for some unknown distribution  $F$ , formally defined as the value  $\theta > 0$  such that

$$\lim_{z \rightarrow \infty} \frac{\log(1/\bar{F}(\zeta z))}{\log(1/\bar{F}(z))} = \zeta^{1/\theta} \quad \text{for all } \zeta > 0. \quad (3.1)$$

In this case,  $F$  is said to have a Weibull-type tail. The normal, gamma, logistic, and Weibull distributions all have Weibull-type tails. Distributions with Weibull-type tails form an important class within the Gumbel MDA. The significance of the WTC is explained in detail in Chapter 4, where it plays a primary role in estimating extreme quantiles and conditional tail expectations.

The underlying approach in this chapter is similar to the approach taken in the previous chapter for the non-zero EVI  $\gamma$  case. We start by constructing a sequence of functionals that approximate the WTC  $\theta$  in terms of the unknown distribution  $F$ . These functionals are also

defined in terms of integrals with integration limits that approach infinity at different rates. A Dirichlet process prior applied to  $F$  induces a posterior of the functional of  $F$ . Instead of working directly with the DP prior, we use its noninformative limit, the Bayesian bootstrap, which enables straightforward MCMC sampling from the approximate posterior distribution of  $\theta$ .

We compare the Bayesian approach with two well-known estimators using simulated data from the normal, gamma, logistic, and Weibull distributions, including some variations of the Weibull. An estimator was proposed in Beirlant et al. (1995) based on the mean residual function  $E(X - x|X > x)$ . A Hill-type estimator was designed by Girard (2004). We use coverage and lengths of confidence/credibility intervals as comparison metrics.

The chapter is organized as follows. We formally describe the setup and the methodology in the next section. In Section 3.3, we derive convergence results. Section 3.4 presents results from a simulation study across various settings, comparing the proposed Bayesian procedure with other methods in the literature. The proofs are given in Section 3.5.

## 3.2 Setup and Methodology

From here on, we assume that  $F$  is continuous and strictly increasing. Because  $F$  is continuous, the cumulative hazard function  $H$  of  $F$  is given by  $H(z) := -\log \bar{F}(z)$ . Hence the quantile function can be expressed as  $q(\tau) := \bar{F}^{-1}(\tau) = H^{-1}(\log(1/\tau))$  for all  $\tau \in (0, 1)$ . By Resnick (1987), if (3.1), then there exists a slowly-varying function  $l(\cdot)$  such that  $H^{-1}(z) = z^\theta l(z)$  for  $z > 0$ . Our objective is to perform Bayesian inference on  $\theta$  with an asymptotic frequentist justification.

To infer about the WTC  $\theta$ , we shall represent  $\theta$  approximately as a sequence of functionals  $\psi_n(F)$ . Then, a reasonable estimator is obtained by plugging in the empirical distribution in place of  $F$ . For the Bayesian approach,  $F$  is given a prior distribution, and the inference is based on the induced posterior distribution of  $\psi_n(F)$ .

To motivate a sequence of functionals approximating  $\theta$  on which the proposed Bayesian method is based, we observe that from the definition of WTC (3.1) that for any  $a_n \rightarrow \infty$ ,

$$\psi_n^0(F) = \int_0^1 \frac{\log(1/\bar{F}(\zeta a_n))}{\log(1/\bar{F}(a_n))} d\zeta \rightarrow \frac{\theta}{1 + \theta}, \quad (3.2)$$

by the dominated convergence theorem, since the integrand is nonnegative and bounded by one. Hence the plug-in estimator  $\psi_n^0(\mathbb{F}_n)$  and the induced posterior of  $\psi_n^0(F)$  are consistent for estimating  $\theta/(1 + \theta)$ . However, the convergence in (3.2) may be arbitrarily slow, not allowing for obtaining a convergence rate, an asymptotic confidence interval, or validating a Bayesian

credible interval regarding asymptotic frequentist coverage.

To alleviate the problem, we restrict  $\zeta$  to a small interval near one. For some small  $\epsilon > 0$ , we integrate both sides of (3.1) with respect to  $\zeta$  over  $(1 - \epsilon, 1)$  to obtain for  $z \rightarrow \infty$  and  $\epsilon \rightarrow 0$ ,

$$\int_{1-\epsilon}^1 \frac{\log(1/\bar{F}_0(\zeta z))}{\log(1/\bar{F}_0(z))} d\zeta \approx \frac{\theta}{\theta + 1} [1 - (1 - \epsilon)^{1+1/\theta}] = \epsilon - \frac{\epsilon^2}{2\theta} + O(\epsilon^3),$$

using the generalized binomial expansion. This leads to the relation

$$\epsilon^{-1} \left( 1 - \epsilon^{-1} \int_{1-\epsilon}^1 \frac{\log(1/\bar{F}(\zeta z))}{\log(1/\bar{F}(z))} d\zeta \right) = \frac{1}{2\theta} + O(\epsilon).$$

Let  $\epsilon_n \downarrow 0$  and  $a_n \uparrow \infty$  be deterministic real sequences. By substituting  $a_n$  for  $z$  and  $x$  for  $\zeta z$  on the integral above, the functional

$$\psi_n(F) = \epsilon_n^{-1} \left\{ 1 - \frac{1}{\epsilon_n a_n \log(1/\bar{F}(a_n))} \int_{(1-\epsilon_n)a_n}^{a_n} \log(1/\bar{F}(x)) dx \right\} \quad (3.3)$$

should approximate  $(2\theta)^{-1}$ , that is,  $(2\psi_n(F))^{-1}$  should approximate  $\theta$ , within  $O(\epsilon_n)$ . A precise result is given by Proposition 6.

For the convergence theory to work, the choices of  $\epsilon_n$  and  $a_n$  must satisfy specific requirements to control the remainder terms in the expansions to lead to the desired asymptotic normality results. We choose  $a_n$  such that  $p_n = \bar{F}(a_n) \asymp n^{-\tau}$  for some  $0 < \tau < 1/2$  and  $\epsilon_n = n^{-\xi}$  for some  $\xi > 0$ . All Weibull-type tailed distributions are infinitely supported to the right and have exponentially decaying tails, implying that  $a_n$  increases logarithmically. Hence, we require that  $a_n \epsilon_n \rightarrow 0$ . Some relation connecting  $\tau$  and  $\xi$  will be imposed to ensure convergence for common Weibull-type tailed distributions (Weibull, normal, logistic, gamma). We also experimented with different choices in our simulation study. Since  $F$  is unknown, we proxy the choice of  $a_n$  by using the empirical distribution function; choose  $a_n$  as the empirical quantile  $q_n(p_n)$ . Then, by the law of the iterated logarithm for the empirical process, we find that

$$\bar{F}(q_n(p_n)) = \bar{\mathbb{F}}_n(q_n(p_n)) + O(\sqrt{(\log \log n)/n}) \asymp p_n$$

almost surely, since  $\bar{\mathbb{F}}_n(q_n(p_n)) = p_n + O(n^{-1})$ . Hence, with probability one, the implementable choice  $a_n = q_n(p_n)$  satisfies the requirement.

Consider a Dirichlet process prior  $F \sim \text{DP}(MG)$ , where  $G$  is a cumulative probability distribution function and  $M > 0$  is the precision parameter. Then the posterior process for  $F$  is the distribution of  $\mathbb{F}_n^* \sim \text{DP}(MG + n\mathbb{F}_n)$  and its BB approximation is given by

$\mathbb{F}_n^\# \sim \text{DP}(n\mathbb{F}_n)$ . With the above choices for  $a_n$  and  $\epsilon_n$  to define the functional  $\psi_n$ , the posterior distribution for  $\eta = (2\theta)^{-1}$  is the distribution of  $\psi_n(\mathbb{F}_n^*)$ , and the BB distribution is the distribution of  $\psi_n(\mathbb{F}_n^\#)$ . The associated plug-in empirical estimator  $\psi_n(\mathbb{F}_n)$  centers the limiting posterior and BB distributions. By sampling from the distributions of  $\psi_n(\mathbb{F}_n^*)$  and  $\psi_n(\mathbb{F}_n^\#)$ , we may obtain Bayesian (and BB) estimators as the posterior mean or median, and also posterior equal-tail credible intervals. The following result describes the computation of the BB distribution for  $\theta$ , whose proof is evident from the definitions of the BB process and the functional  $\psi_n$ .

**Proposition 4.** *Writing  $\mathbb{F}_n^\# = \sum_{i=1}^n W_i \delta_{X_{i:n}}$  where  $X_{1:n} < \dots < X_{n:n}$  are the order statistics of the data  $D_n = (X_1, \dots, X_n)$ , and  $\sum_{i=1}^n W_i = 1$  with  $W_i > 0$  for  $i = 1, \dots, n$ , we may express*

- (i)  $\bar{\mathbb{F}}_n^\#(x) = \sum_{i=j+1}^n W_i$  if  $X_{j:n} \leq x < X_{j+1:n}$  for some  $j = j(x)$ ;
- (ii)  $\int_{X_{j:n}}^{X_{j+1:n}} \log(1/\bar{\mathbb{F}}_n^\#(x)) dx = [\log(1/\sum_{i=j+1}^n W_i)](X_{j+1:n} - X_{j:n})$ ;
- (iii)  $\psi_n(\mathbb{F}_n^\#) = \epsilon_n^{-1} \{1 - (N_1 + N_2 + N_3)/(D\epsilon_n)\}$ , where

$$\begin{aligned} N_1 &= [\log(1/\sum_{i=r}^n W_i)][X_{r:n} - (1 - \epsilon_n)a_n], \\ N_2 &= \sum_{j=r}^{s-1} [\log(1/\sum_{i=j+1}^n W_i)](X_{j+1:n} - X_{j:n}), \\ N_3 &= \log[(1/\sum_{i=s+1}^n W_i)](a_n - X_{s:n}), \\ D &= a_n \log(1/(\sum_{i=s+1}^n W_i)), \end{aligned}$$

and  $r$  and  $s$  are such that  $X_{r-1:n} < (1 - \epsilon_n)a_n \leq X_{r:n}$  and  $X_{s:n} \leq a_n < X_{s+1:n}$ .

In the next section, we shall obtain the convergence rate of the estimators and the limiting coverage and length of the credible intervals.

### 3.3 Convergence Results

We shall call a distribution  $F$  with WTC  $\theta$  to be  $(a_n, \epsilon_n, m_n)$ -regular if

$$\sup_{\zeta \in [1-\epsilon_n, 1]} \left| \frac{\log(1/\bar{F}(\zeta a_n))}{\log(1/\bar{F}(a_n))} - \zeta^{1/\theta} \right| = O(m_n), \quad (3.4)$$

where  $m_n \rightarrow 0$ .

Many standard distributions with Weibull-type tails satisfy regularity such that  $m_n$  can be computed explicitly. The following result provides the values of the WTC and  $m_n$  for some prominent examples.

**Proposition 5.** *The following distributions have regular Weibull tails with  $\theta$  and  $m_n$  as described:*

- *Weibull with shape parameter  $\alpha$ :  $\theta = 1/\alpha$ ,  $m_n = 0$ ;*
- *Logistic:  $\theta = 1$ ,  $m_n = \max(p_n \epsilon_n, p_n^2/a_n)$ ;*
- *Normal:  $\theta = 1/2$ ,  $m_n = a_n^{-2} \epsilon_n \log a_n + a_n^{-4}$ ;*
- *Gamma with shape parameter  $\alpha \neq 1$ :  $\theta = 1$ ,*

$$m_n = \max\{a_n^{-1} \epsilon_n \log a_n, a_n^{-(\lfloor \alpha \rfloor + 1)}\}.$$

**Proposition 6.** *Let  $a_n \rightarrow \infty$ ,  $\epsilon_n \rightarrow 0$ , and  $\psi_n$  be defined by (3.3). If  $F$  is  $(a_n, \epsilon_n, m_n)$  regular with  $m_n = o(\epsilon_n)$ , then*

$$\psi_n(F) = (2\theta)^{-1} + O(\max\{\epsilon_n, m_n/\epsilon_n\}) \rightarrow (2\theta)^{-1}. \quad (3.5)$$

Let the true distribution function of the observations  $X_1, \dots, X_n$  be denoted by  $F_0$ . We assume that  $F_0$  has Weibull-type tails with WTC  $\theta_0$ .

**Corollary 3** (Consistency). *Let  $a_n \rightarrow \infty$ ,  $\epsilon_n \rightarrow 0$ , and  $\psi_n$  be defined by (3.3). If  $F_0$  is  $(a_n, \epsilon_n, m_n)$ -regular with  $m_n = o(\epsilon_n)$ , then*

- (i)  *$\psi_n(\mathbb{F}_n)$  is a consistent estimator for  $(2\theta)^{-1}$  at  $\theta = \theta_0$ ;*
- (ii) *the induced posterior distribution of  $\psi_n(F)$  is consistent for  $(2\theta)^{-1}$  at  $\theta = \theta_0$ .*

By the consistency of the empirical distribution and the Dirichlet posterior distribution, the result is immediately obtained from Proposition 6. It is also apparent that any choice of  $\epsilon_n \rightarrow 0$  gives consistency in the Weibull and normal cases, while consistency holds in the logistic case provided  $p_n = O(\epsilon_n)$ , that is  $\tau \geq \xi$ . For the gamma case, consistency can be obtained only by choosing  $\epsilon_n \rightarrow 0$  logarithmically slowly.

The main theoretical results of this paper are

- $\psi_n(\mathbb{F}_n) - \psi_n(F_0)$  is asymptotically normal when suitably scaled;
- conditional on the data,  $\psi_n(\mathbb{F}_n^*) - \psi_n(\mathbb{F}_n)$  is asymptotically normal when suitably scaled;
- for  $0 < \beta < 1$ , the limiting coverage of a  $(1 - \beta)$  credible interval for  $\theta$  given data converges to  $1 - \beta$ .

To address these problems, we shall apply the functional delta method to the functional  $F \mapsto \psi_n(F)$ . This requires the calculation of a functional derivative of  $\psi_n(\cdot)$ . However, as  $\psi_n$  is a sequence of functionals, weak convergence theory is not sufficient to derive these results. We shall apply the strong approximation theory. An additional advantage is that we can work with the Gateaux (directional) derivative instead of the more complicated concept of Hadamard differentiability. We begin with an expression for the Gateaux derivative of the functional  $\psi_n$ .

**Lemma 4.** *For any càdlàg function  $H$  on  $\mathbb{R}$ , the Gateaux derivative of  $\psi_n$  at  $F_0$  in the direction of  $H$  is given by*

$$\begin{aligned} \psi'_{n,F_0}(H) = \frac{1}{\epsilon_n^2 a_n} \left\{ \frac{H(a_n)}{\bar{F}_0(a_n) [\log(1/\bar{F}_0(a_n))]^2} \int_{(1-\epsilon_n)a_n}^{a_n} \log(1/\bar{F}_0(x)) dx \right. \\ \left. + \frac{1}{\log(1/\bar{F}_0(a_n))} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{H(x)}{\bar{F}_0(x)} dx \right\}. \end{aligned} \quad (3.6)$$

The remainder term in the Taylor series expansion involves the function's second derivative. Let  $\phi_{n,F_0,H}(t) = \psi_n(F_0 + tH)$ , so that  $\phi_{n,F_0,H}(0) = \psi_n(F_0)$ . Then  $\phi'_{n,F_0,H}(0) = \psi'_{n,F_0}(H)$  and  $\phi''_{n,F_0,H}(0) = \psi''_{n,F_0}(H)$ . The formula for the latter is given below.

**Lemma 5.** *For any càdlàg function  $H$  on  $\mathbb{R}$ , the second derivative of  $\phi_{n,F_0,H}$  at  $t = \nu$  is given by*

$$\begin{aligned} \phi''_{n,F_0,H}(\nu) = -\frac{1}{\epsilon_n^2 a_n} \left\{ -\frac{1}{\log[\bar{F}_0(a_n) - \nu H(a_n)]} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{H^2(x)}{[\bar{F}_0(x) - \nu H(x)]^2} dx \right. \\ + \frac{H(a_n)}{[\bar{F}_0(a_n) - \nu H(a_n)] \{\log[\bar{F}_0(a_n) - \nu H(a_n)]\}^2} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{H^2(x)}{[\bar{F}_0(x) - \nu H(x)]^2} dx \\ - \frac{2H^2(a_n)}{[\bar{F}_0(a_n) - \nu H(a_n)]^2 \{\log[\bar{F}_0(a_n) - \nu H(a_n)]\}^3} \int_{(1-\epsilon_n)a_n}^{a_n} \log \frac{1}{\bar{F}_0(x) - \nu H(x)} dx \\ - \frac{H^2(a_n)}{[\bar{F}_0(a_n) - \nu H(a_n)]^2 \{\log[\bar{F}_0(a_n) - \nu H(a_n)]\}^2} \int_{(1-\epsilon_n)a_n}^{a_n} \log \frac{1}{\bar{F}_0(x) - \nu H(x)} dx \\ \left. - \frac{H(a_n)}{[\bar{F}_0(a_n) - \nu H(a_n)] \{\log[\bar{F}_0(a_n) - \nu H(a_n)]\}^2} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{H(x)}{\bar{F}_0(x) - \nu H(x)} dx \right\}. \end{aligned}$$

Substituting (1.4) in the second order Taylor expansion for  $\phi_{n,F_0,\mathbb{J}_n}$  leads to

$$\begin{aligned} \psi_n(\mathbb{F}_n) - \psi_n(F_0) &= n^{-1/2} \psi'_{n,F_0}(n^{-1/2} K_1(F_0(\cdot), n) + \kappa_n) + \frac{1}{2n} \phi''_{n,F_0,\mathbb{J}_n}(\nu) \\ &= n^{-1/2} \psi'_{n,F_0}(n^{-1/2} K_1(F_0(\cdot), n)) + n^{-1/2} \psi'_{n,F_0}(\kappa_n) + \frac{1}{2n} \phi''_{n,F_0,\mathbb{J}_n}(\nu). \end{aligned} \quad (3.7)$$

Here, the first term represents the linearization term, which provides asymptotic normality, while the other two terms are remainder terms shown to be negligible relative to the first term. The following result provides a formula for the variance of the linear term.

**Lemma 6.** *The variance  $\sigma_n^2$  of  $\psi'_{n,F_0}(n^{-1/2}K_1(F_0(\cdot), n))$  is given by*

$$\begin{aligned} & \frac{F_0(a_n)}{\epsilon_n^4 a_n^2 \bar{F}_0(a_n) [\log(1/\bar{F}_0(a_n))]^4} \left[ \int_{(1-\epsilon_n)a_n}^{a_n} \log(1/\bar{F}_0(x)) dx \right]^2 \\ & + \frac{2}{\epsilon_n^4 a_n^2 [\log(1/\bar{F}_0(a_n))]^3} \int_{(1-\epsilon_n)a_n}^{a_n} \log(1/\bar{F}_0(x)) dx \int_{(1-\epsilon_n)a_n}^{a_n} \frac{F_0(x)}{\bar{F}_0(x)} dx \\ & + \frac{1}{\epsilon_n^4 a_n^2 [\log(1/\bar{F}_0(a_n))]^2} \int_{(1-\epsilon_n)a_n}^{a_n} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{F_0(x) \wedge F_0(y) - F_0(x)F_0(y)}{\bar{F}_0(x)\bar{F}_0(y)} dx dy \end{aligned}$$

and is of the order  $\epsilon_n^{-2} p_n^{-1} (\log n)^{-2} \asymp n^{(\tau+2\xi)} (\log n)^{-2}$ .

The following result establishes the asymptotic normality of the plug-in estimator.

**Theorem 6** (Asymptotic normality of plug-in estimator). *If  $\tau < 1/5$ ,  $\tau/2 + 2\xi \geq 1/2$ , then*

$$\frac{\sqrt{n}}{\sigma_n} (\psi_n(\mathbb{F}_n) - \psi_n(F_0)) \rightsquigarrow N(0, 1).$$

If, in addition, and  $F_0$  is  $(a_n, \epsilon_n, m_n)$ -regular with  $m_n = o(n^{-(1/2-\tau)}/\log n)$ , then also

$$\frac{\sqrt{n}}{\sigma_n} (\psi_n(\mathbb{F}_n) - (2\theta_0)^{-1}) \rightsquigarrow N(0, 1).$$

We now study the asymptotic behavior of the Dirichlet process posterior distribution (and, as a special case, the Bayesian bootstrap distribution) of the induced functional.

**Theorem 7** (Bernstein-von Mises theorem). *If  $\tau < 1/5$ , then conditional on the data  $D_n$ , almost surely  $[F_0]$ ,*

$$\frac{\sqrt{n}}{\sigma_n} (\psi_n(\mathbb{F}_n^*) - \psi_n(\mathbb{F}_n)) \rightsquigarrow N(0, 1).$$

Combining the conclusions of the last two theorems, it follows that if  $\tau < 1/5$  and  $\tau/2 + 2\xi \geq 1/2$  and  $F_0$  is  $(a_n, \epsilon_n, m_n)$ -regular with  $m_n = o(n^{-(1/2-\tau)}/\log n)$ , then the posterior distribution contracts at the rate  $\sigma_n/\sqrt{n}$ , which is asymptotically  $n^{-(1/2-\xi-\tau/2)}/\log n$ . Moreover, these two theorems lead us to the following result on the coverage of Bayesian credible intervals for  $\theta$ .

**Corollary 4** (Coverage). *Let  $Q_{\beta/2}$  and  $Q_{1-\beta/2}$  be the  $\beta/2$  and  $1 - \beta/2$  posterior quantiles of  $\theta$ . Then  $\mathbb{P}_{F_0}(Q_{\beta/2} \leq \theta_0 \leq Q_{1-\beta/2}) \rightarrow 1 - \beta$ .*

**Remark 1.** The conditions  $\tau < 1/5$  and  $\tau/2 + 2\xi \geq 1/2$  are essential for obtaining asymptotically correct coverage for a Bayesian credible interval. No further condition is needed for the Weibull distribution, as  $m_n = 0$ . The nearly best rate is obtained for  $\tau \gtrsim 0$  and  $\xi \gtrsim 1/4$ , giving the  $n^{-1/4}$  rate of convergence up to a logarithmic factor; here and below, by  $\tau \gtrsim 0$  we mean that  $p_n = O(1/l_n)$  and by  $\xi \gtrsim 1/4$ , we mean  $\epsilon_n = O(n^{-1/4}/l_n)$ , for a logarithmically increasing sequence  $l_n$ . For the logistic case, the additional assumption on the bias reduces to  $p_n^2 \epsilon_n = o(n^{-1/2}/\log n)$  and  $p_n^3 = o(n^{-1/2})$ , so that  $2\tau + \xi \gtrsim 1/2$  and  $\tau \gtrsim 1/6$ . The fastest convergence rate is obtained by choosing  $\tau \gtrsim 1/6$  and  $\xi \gtrsim 1/3$ , giving the rate  $n^{-1/12}$  up to a logarithmic factor.

## 3.4 Numerical Results

In this section, we study the accuracy of the Bayesian bootstrap (BB) estimator and the coverage of the BB credible interval for various true distributions, and compare the performance with several non-Bayesian estimators and confidence intervals from the literature.

The specific simulation steps, implemented in R, are outlined below.

- Set a random sample size  $n$ .
- Set  $B$  as the number of random samples to be generated from a distribution with Weibull-type tail. Perform the following loop  $B$  times.
  - Obtain a random sample  $X_1, \dots, X_n$  from  $F_0$ , where  $F_0$  is one of the following: logistic or Weibull (including the extended and generalized power Weibull variants).
  - Compute  $p_n = n^{-\tau}/\log n$ .
  - Evaluate  $\hat{a}_n := X_{\lfloor n(1-p_n) \rfloor}$ .
  - Set the value of  $\epsilon_n = n^{-\xi}/\log n$ .
  - Set  $C$  as the number of iterations to be performed (for the Bayesian bootstrap). Perform the following loop  $C$  times.
    - \* Generate a random sample  $E_1, \dots, E_n$  from a standard exponential distribution (independently from  $X_1, \dots, X_n$ ).
    - \* Generate a Dirichlet distribution from the standard exponential random sample.

\* Apply Proposition 4 to determine  $\psi_n(\mathbb{F}_n^\#)$  using  $\hat{a}_n$  for  $a_n$ .

We have used  $B = 100,000$ ,  $C = 100$  and  $n = 500$  as the number of random samples, number of Bayesian bootstrap iterations, and the random sample size, respectively. In addition, we set  $\tau = 1/6$  and  $\xi = 1/4$ . Note that we have made adjustments to  $p_n$  and  $\epsilon_n$  with  $\log n$  terms in the denominator which appear to improve the results and the runtime of the R code.

Define  $\omega_n := 1/[2\psi_n(\mathbb{F}_n^\#)]$ . Given random sample data, each Bayesian bootstrap simulation yields a Dirichlet distribution of  $\theta_0$ , and the sample average of the  $\omega_n$  values across the  $C$  simulations was taken as the Bayes estimator for  $\theta_0$ .

We compare the Bayes estimator with the following frequentist estimators listed in (Girard 2004):

- Girard Hill-type (Girard 2004) estimator

$$\theta_n^{\text{GH}} = \frac{\sum_{i=1}^{k_n-1} (\log(X_{n-i+1:n}) - \log(X_{n-k_n+1:n}))}{\sum_{i=1}^{k_n-1} (\log \log(n/i) - \log \log(n/k_n))},$$

- Beirlant (Beirlant et al. 1995) estimator

$$\theta_n^{\text{BE}} = \frac{\log(n/k_n)}{X_{n-k_n+1:n}} \frac{1}{k_n - 1} \sum_{i=1}^{k_n-1} (X_{n-i+1:n} - X_{n-k_n+1:n}).$$

For  $\beta > 0$ , define  $z_{\beta/2}$  as the quantile such that  $P(Z > z_{\beta/2}) = \beta/2$  where  $Z$  is standard normal. A  $(1 - \beta)$  credible interval is formed by calculating the  $\beta/2$  and  $1 - \beta/2$  quantiles from the  $\omega_n$  values across the  $C$  simulations. The  $(1 - \beta)$  confidence intervals associated with the frequentist estimators are given as follows:

- Girard Hill-type confidence interval

$$\left( \frac{\theta_n^{\text{GH}}}{1 + z_{\beta/2}/\sqrt{k_n}}, \frac{\theta_n^{\text{GH}}}{1 - z_{\beta/2}/\sqrt{k_n}} \right)$$

- Beirlant confidence interval

$$\left( \frac{\theta_n^{\text{BE}}}{1 + z_{\beta/2}/\sqrt{k_n}}, \frac{\theta_n^{\text{BE}}}{1 - z_{\beta/2}/\sqrt{k_n}} \right)$$

The choice of  $k_n$  is important, as each choice entails a bias-variance trade-off. There's no definitive guidance on an optimal choice of  $k_n$  beyond the fact that it should satisfy the

conditions of Theorem 2 in Girard (2004). We use  $k_n = \log n$  for the frequentist estimators as it satisfies the conditions of Theorem 2.

We calculated two metrics to compare our Bayesian approach with the Beirlant and Girard-Hill frequentist approaches. First, we calculated a coverage ratio, which is the proportion of the  $B$  credible/confidence intervals (from the  $B$  random samples) that contain the true  $\theta_0$  value. Since we calculated 95% credible/confidence intervals, the closer the coverage ratio is to 95%, the better. The second metric is the median length of the  $B$  credible/confidence intervals. If the first metric values are all close, the second metric should determine the best method, as the shortest credible/confidence interval would be deemed the better or best approach. As we shall see, the credible intervals are significantly smaller than the Beirlant and Girard-Hill confidence intervals in most of the cases considered.

In the following tables, we present the coverage ratios and median lengths of credible or confidence intervals in the columns marked as BA, BE, and GH corresponding to the Bayes, Beirlant, and Girard-Hill estimators, respectively.

Table 3.1: Selected WTC inference simulation results for a Weibull distribution with shape  $\alpha$ .

$\alpha$	Coverage			Length		
	BA	BE	GH	BA	BE	GH
1/2	0.89	0.84	0.93	11.36	9.57	9.52
1	0.96	0.95	0.93	3.26	4.09	4.76
3	0.97	0.98	0.93	0.57	1.23	1.59
6	0.96	0.99	0.93	0.18	0.60	0.79
10	0.95	0.99	0.93	0.08	0.36	0.48

We make the following observations based on the table of Weibull distribution results. First, we note that the coverage ratios, except for  $\alpha = 1/2$  for Bayes and Beirlant estimators, appear to be reasonably close to 95%. Due to the nature of the Girard-Hill estimator and confidence interval, the GH coverage ratio is showing no dependence on  $\alpha$ . The median length of the credible/confidence interval decreases with  $\alpha$  for each of the estimators, with the Bayes estimator giving the shortest median interval lengths for  $\alpha \geq 1$ .

We also consider two variants of the Weibull distribution. First, we consider the extended Weibull distribution,  $EW(\alpha, \beta)$  with  $\alpha > 0$  and  $\beta \in \mathbb{R}$  and survival function  $\bar{F}(x) = r(x) \exp(-x^\alpha)$  where  $r(\cdot) \in RV_\infty(\beta)$ , the class of regularly varying functions of order  $\beta$  at infinity, such as  $x^\beta$ . He et al. (2020) has shown that  $\theta = 1/\alpha$  in this case. We use a version

of the extended Weibull distribution as described in Almalki and Nadarajah (2014) and implemented as the `rExW` function within the `RelDists` library in R. This function has a cumulative distribution function given by

$$\bar{F}(x) = \frac{1 - e^{-\mu x^\alpha}}{1 - (1 - \nu)e^{-\mu x^\alpha}}, \quad x > 0,$$

where  $\mu$ ,  $\alpha$ , and  $\nu$  are positive parameters.

Table 3.2: Selected WTC inference simulation results for an extended Weibull distribution with  $\mu = 1$ ,  $\nu = 0.9$  and shape  $\sigma$ .

$\alpha$	Coverage			Length		
	BA	BE	GH	BA	BE	GH
1/2	0.89	0.82	0.91	11.85	9.74	9.71
1	0.96	0.94	0.91	3.37	4.16	4.86
3	0.98	0.98	0.91	0.60	1.26	1.62
6	0.96	0.98	0.91	0.19	0.61	0.81
10	0.94	0.99	0.91	0.08	0.36	0.49

For the extended Weibull distribution, setting  $\nu = 1$  would revert to the ordinary Weibull distribution. We instead set  $\nu = 0.9$  for the calculations to assess comparability of results for the ordinary Weibull. As seen in the Extended Weibull results table, similar conclusions can be drawn for the Extended Weibull that were made for the ordinary Weibull. In particular, for  $\alpha \geq 1$ , the Bayes approach appear to produce almost 95% coverage with the median credible interval length significantly shorter than either confidence interval length.

A second Weibull distribution variant we consider is the generalized power Weibull distribution (Almalki and Nadarajah 2014) with three positive parameters  $\sigma$ ,  $\mu$ , and  $\nu$  having survival function

$$\bar{F}(x) = \exp[1 - (1 + \sigma x^\mu)^{1/\nu}], \quad x > 0.$$

It is straightforward to verify that the WTC for this distribution is  $\nu/\mu$ .

Similar to the extended Weibull, we set  $\nu = 0.9$  to assess comparability of results with the ordinary Weibull. Again, the conclusions made earlier for the ordinary Weibull can be made for the generalized power Weibull.

The results for the standard logistic show that only the Bayesian interval reaches the 95% coverage level, and it has the shortest median interval length.

Table 3.3: Selected WTC inference simulation results for a generalized power Weibull distribution with  $\nu = 0.9$ ,  $\sigma = 1$ , and shape  $\sigma$ .

$\alpha$	Coverage			Length		
	BA	BE	GH	BA	BE	GH
1/2	0.90	0.83	0.90	10.51	8.70	8.92
1	0.97	0.94	0.90	2.93	3.77	4.46
3	0.98	0.98	0.90	0.52	1.15	1.49
6	0.95	0.98	0.90	0.17	0.56	0.74
10	0.90	0.99	0.90	0.07	0.33	0.45

Table 3.4: Selected WTC inference simulation results for the standard logistic distribution.

Coverage			Length		
BA	BE	GH	BA	BE	GH
0.96	0.94	0.92	3.80	4.11	4.80

As noted earlier, the normal and gamma distributions do not fully satisfy the Proposition 6 conditions and thus the Bayes estimator based on the functional in (3.3) may not consistently estimate  $\theta$ , although the Bayes estimator based on the functional in (3.2) does. This, in turn, may affect the coverage of the Bayesian interval. Nevertheless, the following table, which presents results for normal observations, shows that the coverage for the Bayesian interval is almost 95% and its length is also shorter than that of its competitors.

Table 3.5: Selected WTC inference simulation results for the standard normal distribution.

Coverage			Length		
BA	BE	GH	BA	BE	GH
0.94	0.81	0.66	2.80	2.80	3.42

For the gamma distribution, however, the coverage of the Bayesian interval deteriorates as the shape parameter  $\alpha$  exceeds 1 and further increases.

Table 3.6: Selected WTC inference simulation results for a gamma distribution with scale  $\theta = 1$  and shape  $\alpha$ .

$\alpha$	Coverage			Length		
	BA	BE	GH	BA	BE	GH
1/2	0.94	0.83	0.80	6.06	5.11	5.79
1	0.96	0.94	0.92	3.20	4.08	4.74
3	0.86	0.98	0.99	1.79	2.77	3.37
6	0.66	0.96	0.98	1.09	2.11	2.63
10	0.44	0.90	0.96	0.72	1.70	2.16

### 3.5 Proofs

*Proof of Proposition 5.*

**Case 1: Weibull.** Since  $\bar{F}(x) = \exp\{-x^\alpha\}$ , for all  $\zeta > 0$ ,  $a > 0$ ,

$$\frac{\log(1/\bar{F}(\zeta a_n))}{\log(1/\bar{F}(a_n))} - \zeta^\alpha = 0,$$

and hence  $\theta = 1/\alpha$  and  $m_n = 0$ .

**Case 2: Logistic.** In this case,  $\log(1/\bar{F}(x)) = x + \log(1 + e^{-x})$ , and hence

$$\begin{aligned} \frac{\log(1/\bar{F}_0(\zeta a_n))}{\log(1/\bar{F}_0(a_n))} - \zeta &= \frac{\zeta a_n + \log(1 + e^{-\zeta a_n})}{a_n + \log(1 + e^{-a_n})} - \zeta \\ &= \frac{e^{-\zeta a_n} + O(e^{-2\zeta a_n}) - \zeta e^{-a_n} - O(e^{-2a_n})}{a_n + \log(1 + e^{-a_n})}. \end{aligned}$$

Therefore, using  $a_n \epsilon_n \rightarrow 0$ , uniformly for any  $1 - \epsilon_n \leq \zeta \leq 1$ , the expression in the numerator reduces to

$$\begin{aligned} e^{-\zeta a_n} - \zeta e^{-a_n} + O(e^{-2(1-\epsilon_n)a_n}) &= e^{-a_n}(e^{(1-\zeta)a_n} - \zeta) + O(e^{-2a_n}) \\ &\sim p_n(1 - \zeta)[1 + a_n + O((1 - \zeta)a_n^2)] + O(p_n^2) \\ &= O(p_n \epsilon_n a_n + p_n \epsilon_n^2 a_n^2 + p_n^2). \end{aligned}$$

Hence  $\theta = 1$  and  $m_n = p_n(\epsilon_n a_n + p_n)/a_n \asymp \max(p_n \epsilon_n, p_n^2/a_n)$ .

**Case 3: Normal.** If  $F$  is the standard normal, then from a well-known estimate

$$\frac{1}{x\sqrt{1+1/x^2}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \leq \bar{F}(x) \leq x^{-1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

it follows that for large  $a > 0$  and  $0 < \zeta < 1$ ,

$$\begin{aligned} & \frac{\frac{(\zeta a)^2}{2} + \log(\zeta a) + \frac{\log(2\pi)}{2} - \zeta^2 \log a - \frac{\zeta^2 \log(2\pi)}{2} - \frac{\zeta^2 a^2}{2} - \frac{\zeta^2 \log(1 + \frac{1}{a^2})}{2}}{\log a + \frac{\log(2\pi)}{2} + \frac{a^2}{2} + \frac{\zeta^2 \log(1 + \frac{1}{a^2})}{2}} \\ & \leq \frac{\log(1/\bar{F}(\zeta a))}{\log(1/\bar{F}(a))} - \zeta^2 \\ & \leq \frac{\frac{(\zeta a)^2}{2} + \log(\zeta a) + \frac{1}{2} \log \left[ 1 + \frac{1}{\zeta^2 a^2} \right] + \frac{\log(2\pi)}{2} - \frac{\zeta^2 a^2}{2} - \zeta^2 \log a - \frac{\zeta^2 \log(2\pi)}{2}}{\log a + \frac{\log(2\pi)}{2} + \frac{a^2}{2}}. \end{aligned}$$

The numerator of the upper bound for large  $a$  has itself an upper bound

$$\begin{aligned} & \log \zeta + (1 - \zeta^2) \log a + (1 - \zeta^2) \frac{\log(2\pi)}{2} + \frac{1}{2} \frac{1}{\zeta^2 a^2} + O\left(\frac{1}{\zeta^4 a^4}\right) \\ & \leq \log 1 + \log a - (1 - \epsilon)^2 \log a + [1 - (1 - \epsilon)^2] \log(2\pi)/2 \\ & = O(\epsilon \log a + a^{-2}) \end{aligned}$$

The numerator of the lower bound is handled similarly and is also  $O(\epsilon \log a + a^{-2})$ . Hence  $\theta = 1/2$  and  $m_n = (\epsilon_n \log a_n)/a_n^2 + 1/a_n^4$ .

**Case 4: Gamma.** Defining  $k = \lfloor \alpha \rfloor$  and using integration by parts, we have for sufficiently large  $x$

$$\begin{aligned} & \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)} \left[ 1 + \frac{\alpha-1}{x} + \dots + \frac{(\alpha-1) \cdots (\alpha-k-1)}{x^{k+1}} \right] \\ & \leq \bar{F}(x) \\ & \leq \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)} \left[ 1 + \frac{\alpha-1}{x} + \dots + \frac{(\alpha-1) \cdots (\alpha-k)}{x^k} \right]. \end{aligned}$$

It is easy to verify that, if the lower bound is denoted by  $h_1(x)$  and the upper bound by  $h_2(x)$ , then  $(\log h_1(x))/(\log h_2(x)) = 1 + O(x^{-(k+1)})$ , and hence it suffices to replace  $\bar{F}(x)$  by  $f(x) = x^{\alpha-1} e^{-x}/\Gamma(\alpha)$  up to an error  $O(a_n^{-(k+1)})$  in  $m_n$ . Now  $\log(1/f(x)) = x - (\alpha-1) \log x + \log \Gamma(\alpha)$ ,

so that unless  $\alpha = 1$ , uniformly for  $1 - \epsilon_n \leq \zeta \leq 1$ ,

$$\begin{aligned} \frac{\log(1/f(\zeta a_n))}{\log(1/f(a_n))} - \zeta &= \frac{-(\alpha - 1) \log \zeta - (\alpha - 1)(1 - \zeta) \log a_n + (1 - \zeta) \log \Gamma(\alpha)}{a_n - (\alpha - 1) \log a_n + \log \Gamma(\alpha)} \\ &= O(a_n^{-1} \epsilon_n \log a_n). \end{aligned}$$

This shows that  $\theta = 1$  and  $m_n = \max(a_n^{-1} \epsilon_n \log a_n, a_n^{-(\lfloor \alpha \rfloor + 1)})$ .  $\square$

*Proof of Proposition 6.* By the assumption, there exists  $K > 0$  and  $N_1 > 0$  such that if  $n > N_1$  then for all  $\zeta$ ,

$$\zeta^{1/\theta} - Km_n \leq \frac{\log(1/\bar{F}(\zeta a_n))}{\log(1/\bar{F}(a_n))} \leq \zeta^{1/\theta} + Km_n.$$

Integrating over  $[1 - \epsilon_n, 1]$  we obtain

$$\int_{1-\epsilon_n}^1 [\zeta^{1/\theta} - Km_n] d\zeta \leq \int_{1-\epsilon_n}^1 \frac{\log(1/\bar{F}(\zeta a_n))}{\log(1/\bar{F}(a_n))} d\zeta \leq \int_{1-\epsilon_n}^1 [\zeta^{1/\theta} + Km_n] d\zeta$$

which implies

$$\begin{aligned} \epsilon_n^{-2} \left\{ \epsilon_n - \frac{\theta}{1 + \theta} [1 - (1 - \epsilon_n)^{(1+\theta)/\theta}] - Km_n \epsilon_n \right\} \\ \leq \psi_n(F) \\ \leq \epsilon_n^{-2} \left\{ \epsilon_n - \frac{\theta}{1 + \theta} [1 - (1 - \epsilon_n)^{(1+\theta)/\theta}] + Km_n \epsilon_n \right\}. \end{aligned}$$

Expanding into the infinite generalized binomial series

$$(1 - \epsilon_n)^{(1+\theta_0)/\theta_0} = 1 + \frac{1 + \theta_0}{\theta_0} (-\epsilon_n) + \frac{1}{2} \frac{1 + \theta_0}{\theta_0} \frac{1}{\theta_0} \epsilon_n^2 + \dots,$$

the last inequality leads to  $\psi_n(F) = (2\theta)^{-1} + O(\epsilon_n + m_n/\epsilon_n) \rightarrow (2\theta)^{-1}$  as  $n \rightarrow \infty$ .  $\square$

*Proof of Lemma 4.* We compute

$$\begin{aligned}
\psi'_{n,F_0}(H) &= \lim_{t \rightarrow 0} \frac{1}{t} [\psi_n(F_0 + tH) - \psi_n(F_0)] \\
&= \lim_{t \rightarrow 0} \frac{1}{t} \left\{ \frac{1}{\epsilon_n} \left[ 1 - \frac{1}{\epsilon_n a_n \log \frac{1}{\bar{F}_0(a_n) - tH(a_n)}} \int_{(1-\epsilon_n)a_n}^{a_n} \log \frac{1}{\bar{F}_0(x) - tH(x)} dx \right] \right. \\
&\quad \left. - \frac{1}{\epsilon_n} \left[ 1 - \frac{1}{\epsilon_n a_n \log \frac{1}{\bar{F}_0(a_n)}} \int_{(1-\epsilon_n)a_n}^{a_n} \log \frac{1}{\bar{F}_0(x)} dx \right] \right\} \\
&= \left\{ \frac{1}{\epsilon_n} \left[ 1 - \frac{1}{\epsilon_n a_n \log \frac{1}{\bar{F}_0(a_n) - tH(a_n)}} \int_{(1-\epsilon_n)a_n}^{a_n} \log \frac{1}{\bar{F}_0(x) - tH(x)} dx \right] \right\}' \Big|_{t=0} \\
&= \frac{-1}{\epsilon_n^2 a_n} \left\{ \frac{1}{\log \frac{1}{[\bar{F}_0(a_n) - tH(a_n)]}} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{H(x)}{\bar{F}_0(x) - tH(x)} dx \right. \\
&\quad \left. - \frac{H(a_n)}{[\bar{F}_0(a_n) - tH(a_n)] \log^2 [\bar{F}_0(a_n) - tH(a_n)]} \int_{(1-\epsilon_n)a_n}^{a_n} \log \frac{1}{\bar{F}_0(x) - tH(x)} dx \right\} \Big|_{t=0} \\
&= \frac{1}{\epsilon_n^2 a_n} \left\{ \frac{H(a_n)}{\bar{F}_0(a_n) [\log \bar{F}_0(a_n)]^2} \int_{(1-\epsilon_n)a_n}^{a_n} \log(1/\bar{F}_0(x)) dx \right. \\
&\quad \left. + \frac{1}{\log \bar{F}_0(a_n)} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{H(x)}{\bar{F}_0(x)} dx \right\}.
\end{aligned}$$

□

*Proof of Lemma 5.* From the above calculation of the derivative of  $\psi_n$  we have

$$\begin{aligned}
\phi'_{n,F_0,H}(\nu) &= \frac{-1}{\epsilon_n^2 a_n} \left\{ \frac{1}{\log \frac{1}{[\bar{F}_0(a_n) - \nu H(a_n)]}} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{H(x)}{\bar{F}_0(x) - \nu H(x)} dx \right. \\
&\quad \left. - \frac{H(a_n)}{[\bar{F}_0(a_n) - \nu H(a_n)] \log^2 [\bar{F}_0(a_n) - \nu H(a_n)]} \int_{(1-\epsilon_n)a_n}^{a_n} \log \frac{1}{\bar{F}_0(x) - \nu H(x)} dx \right\}.
\end{aligned}$$

Differentiating both sides of this equation with respect to  $\nu$  using the product rule yields

$$\begin{aligned} \phi''_{n,F_0,H}(\nu) = & -\frac{1}{\epsilon_n^2 a_n} \left\{ \frac{1}{\log \frac{1}{\bar{F}_0(a_n) - \nu H(a_n)}} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{H^2(x)}{[\bar{F}_0(x) - \nu H(x)]^2} dx \right. \\ & + \frac{H(a_n)}{[\bar{F}_0(a_n) - \nu H(a_n)] \{\log[\bar{F}_0(a_n) - \nu H(a_n)]\}^2} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{H^2(x)}{[\bar{F}_0(x) - \nu H(x)]^2} dx \\ & - \frac{H^2(a_n)}{[\bar{F}_0(a_n) - \nu H(a_n)]^2 \{\log[\bar{F}_0(a_n) - \nu H(a_n)]\}^2} \int_{(1-\epsilon_n)a_n}^{a_n} \log \frac{1}{\bar{F}_0(x) - \nu H(x)} dx \\ & - \frac{2H^2(a_n)}{[\bar{F}_0(a_n) - \nu H(a_n)]^2 \{\log[\bar{F}_0(a_n) - \nu H(a_n)]\}^3} \int_{(1-\epsilon_n)a_n}^{a_n} \log \frac{1}{\bar{F}_0(x) - \nu H(x)} dx \\ & \left. - \frac{H(a_n)}{[\bar{F}_0(a_n) - \nu H(a_n)] \{\log[\bar{F}_0(a_n) - \nu H(a_n)]\}^2} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{H(x)}{\bar{F}_0(x) - \nu H(x)} dx \right\} \end{aligned}$$

□

*Proof of Lemma 6.* To calculate the variance  $\sigma_n^2$ , we first note the following properties of Kiefer processes:

$$\begin{aligned} \mathbb{E}[K_1^2(F_0(a_n), n)] &= nF_0(a_n)\bar{F}_0(a_n); \\ \mathbb{E} \left[ \int_{(1-\epsilon_n)a_n}^{a_n} \frac{K_1(F_0(a_n), n)K_1(F_0(x), n)}{\bar{F}_0(x)} dx \right] &= nF_0(a_n) \int_{(1-\epsilon_n)a_n}^{a_n} \frac{F_0(x)}{\bar{F}_0(x)} dx; \\ \mathbb{E} \left[ \int_{(1-\epsilon_n)a_n}^{a_n} \frac{K_1(F_0(x), n)}{\bar{F}_0(x)} dx \right]^2 &= \int_{(1-\epsilon_n)a_n}^{a_n} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{F_0(x) \wedge F_0(y) - F_0(x)F_0(y)}{\bar{F}_0(x)\bar{F}_0(y)} dx dy. \end{aligned}$$

We also note that since  $\psi'_{n,F_0}$  is linear in its argument and that the mean of the normal random variables underlying a Kiefer process is 0,

$$\mathbb{E} \left[ \psi'_{n,F_0} \left( n^{-1/2} K_1(F_0(\cdot), n) \right) \right] = 0.$$

Then we have  $\psi'_{n,F_0}(n^{-1/2}K_1(F_0(\cdot), n))$  is normally distributed with variance

$$\begin{aligned}
\sigma_n^2 &= \mathbb{E} \left[ \psi'_{n,F_0}(n^{-1/2}K_1(F_0(\cdot), n))^2 \right] \\
&= \frac{1}{n} \mathbb{E} \left[ \frac{\left[ \int_{(1-\epsilon_n)a_n}^{a_n} \log \frac{1}{F_0(x)} dx \right]^2}{\epsilon_n^4 a_n^2 [\bar{F}_0(a_n)]^2 [\log \bar{F}_0(a_n)]^4} K_1^2(F_0(a_n), n) \right] \\
&\quad + \frac{1}{n} \mathbb{E} \left[ \frac{1}{\epsilon_n^4 a_n^2 \left[ \log \frac{1}{\bar{F}_0(a_n)} \right]^2} \left[ \int_{(1-\epsilon_n)a_n}^{a_n} \frac{K_1(F_0(x), n)}{\bar{F}_0(x)} dx \right]^2 \right] \\
&\quad - \mathbb{E} \left[ \frac{2 \left( \int_{(1-\epsilon_n)a_n}^{a_n} \log \frac{1}{\bar{F}_0(x)} dx \right) K_1(F_0(a_n), n)}{n \epsilon_n^4 a_n^2 [\bar{F}_0(a_n)] \left[ \log \frac{1}{\bar{F}_0(a_n)} \right] [\log \bar{F}_0(a_n)]^2} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{K_1(F_0(x), n)}{\bar{F}_0(x)} dx \right] \\
&= \frac{1}{n} \left[ \frac{n F_0(a_n) \bar{F}_0(a_n)}{\epsilon_n^4 a_n^2 [\bar{F}_0(a_n)]^2 [\log \bar{F}_0(a_n)]^4} \left[ \int_{(1-\epsilon_n)a_n}^{a_n} \log \frac{1}{\bar{F}_0(x)} dx \right]^2 \right. \\
&\quad + \left[ \frac{2n \bar{F}_0(a_n)}{n \epsilon_n^4 a_n^2 [\bar{F}_0(a_n)] [\log \bar{F}_0(a_n)]^3} \int_{(1-\epsilon_n)a_n}^{a_n} \log \frac{1}{\bar{F}_0(x)} dx \int_{(1-\epsilon_n)a_n}^{a_n} \frac{F_0(x)}{\bar{F}_0(x)} dx \right. \\
&\quad + \left. \frac{1}{\epsilon_n^4 a_n^2 \left[ \log \frac{1}{\bar{F}_0(a_n)} \right]^2} \int_{(1-\epsilon_n)a_n}^{a_n} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{F_0(x) \wedge F_0(y) - F_0(x)F_0(y)}{\bar{F}_0(x)\bar{F}_0(y)} dx dy \right. \\
&= \frac{F_0(a_n)}{\epsilon_n^4 a_n^2 \bar{F}_0(a_n) [\log \bar{F}_0(a_n)]^4} \left[ \int_{(1-\epsilon_n)a_n}^{a_n} \log \bar{F}_0(x) dx \right]^2 \\
&\quad + \frac{2}{\epsilon_n^4 a_n^2 [\log \bar{F}_0(a_n)]^3} \int_{(1-\epsilon_n)a_n}^{a_n} \log(1/\bar{F}_0(x)) dx \int_{(1-\epsilon_n)a_n}^{a_n} \frac{F_0(x)}{\bar{F}_0(x)} dx \\
&\quad + \frac{1}{\epsilon_n^4 a_n^2 \left[ \log \bar{F}_0(a_n) \right]^2} \int_{(1-\epsilon_n)a_n}^{a_n} \int_{(1-\epsilon_n)a_n}^{a_n} \frac{F_0(x) \wedge F_0(y) - F_0(x)F_0(y)}{\bar{F}_0(x)\bar{F}_0(y)} dx dy.
\end{aligned}$$

Note that

- $F_0(a_n) \asymp 1$ ,  $\bar{F}_0(a_n) \asymp p_n \asymp n^{-\tau}$ ,  $\log(1/\bar{F}_0(a_n)) \asymp \log n$ ;
- uniformly for all  $(1-\epsilon_n)a_n \leq x \leq a_n$ ,  $F_0(x) \asymp 1$ ,  $\bar{F}_0(x) \asymp p_n \asymp n^{-\tau}$ ,  $\log(1/\bar{F}_0(x)) \asymp \log n$ ;
- uniformly for all  $(1-\epsilon_n)a_n \leq x, y \leq a_n$ ,  $F_0(x) \wedge F_0(y) - F_0(x)F_0(y) \asymp (1-p_n) - (1-p_n)^2 \asymp p_n$ .

Therefore,

(i) the order of the first term is

$$[\epsilon_n^4 a_n^2 p_n (\log n)^4]^{-1} (\log n)^2 a_n^2 \epsilon_n^2 \asymp \epsilon_n^{-2} p_n^{-1} (\log n)^{-2};$$

(ii) the order of the second term is

$$[\epsilon_n^4 a_n^2 (\log n)^3]^{-1} (\log n) (a_n \epsilon_n) (a_n \epsilon_n p_n^{-1}) \asymp \epsilon_n^{-2} p_n^{-1} (\log n)^{-2};$$

(iii) the order of the third term is

$$[\epsilon_n^4 a_n^2 (\log n)^2]^{-1} (a_n \epsilon_n)^2 (p_n p_n^{-2}) \asymp \epsilon_n^{-2} p_n^{-1} (\log n)^{-2}.$$

Hence  $\sigma_n^2 \asymp \epsilon_n^{-2} p_n^{-1} (\log n)^{-2} \asymp n^{2\xi+\tau} (\log n)^{-2}$ , so that  $\sigma_n \asymp n^{\xi+\tau/2} / \log n$ .  $\square$

*Proof of Theorem 6.* Using the strong representation that for some  $0 < \nu_n < 1$  depending on the sample,

$$\frac{\sqrt{n}}{\sigma_n} [\psi_n(\mathbb{F}_n) - \psi_n(F_0)] = \frac{1}{\sigma_n} \psi'_{n,F_0}(n^{-1/2} K_1(F_0(\cdot), n)) + \frac{\psi'_{n,F_0}(\kappa_n)}{\sigma_n} + \frac{\phi''_{n,F_0,\mathbb{J}_n}(\nu_n)}{2\sigma_n \sqrt{n}}, \quad (3.8)$$

we see that the leading term on the right side is distributed as the standard normal by the last lemma, so verifying that the last two terms tend to zero suffices.

Using the uniform bound  $O(n^{-1/2} \log^2 n)$  for the strong approximation error  $\kappa_n$  in Lemma 4, we obtain the bound for  $|\psi'_{n,F_0}(\kappa_n)|$  of the order

$$a_n^{-1} \epsilon_n^{-2} \left[ \frac{n^{-1/2} \log^2 n}{p_n \log^2 n} a_n \epsilon_n \log n + \frac{n^{-1/2} \log^2 n}{p_n \log n} a_n \epsilon_n \right].$$

Hence, the first derivative term in (3.8) is of the order

$$\epsilon_n p_n^{1/2} \log n \times n^{-1/2} \epsilon_n^{-1} p_n^{-1} \log n = (n p_n)^{-1/2} \log^2 n \asymp n^{-(1-\tau)/2} \log^2 n \rightarrow 0.$$

We next estimate the second derivative term in (3.8) using the expression from Lemma 5 and plugging in  $n^{-1/2} \mathbb{J}_n$  for  $H$ . Noting that  $\bar{F}_0(x) - \nu_n n^{-1/2} \mathbb{J}_n(x)$  is of the order  $p_n$  uniformly in  $[(1 - \epsilon_n) a_n, a_n]$  since  $p_n \asymp n^{-\tau} \gg n^{-1/2} \sqrt{\log \log n}$ , we obtain that  $\phi''_{n,F_0,\mathbb{J}_n}(\nu_n)$  is bounded

by the order

$$\begin{aligned}
& \epsilon_n^{-2} a_n^{-1} \left\{ \frac{1}{\log n} \frac{\log \log n}{p_n} a_n \epsilon_n + \frac{\sqrt{\log \log n}}{p_n (\log n)^2} a_n \epsilon_n \frac{\log \log n}{p_n^2} \right. \\
& \quad \left. + \frac{\log \log n}{p_n^2 \log^2 n} a_n \epsilon_n \log n + \frac{\sqrt{\log \log n}}{p_n (\log n)^2} a_n \epsilon_n \sqrt{\log \log n} \right\} \\
& \asymp \epsilon_n^{-1} \log \log n [p_n^{-1} (\log n)^{-1} + p_n^{-3} (\log n)^{-2} + p_n^{-2} (\log n)^{-1} + p_n^{-1} (\log n)^{-2}] \\
& \asymp \epsilon_n^{-1} p_n^{-3} (\log n)^{-2} \log \log n.
\end{aligned}$$

Hence, the second derivative term in (3.8) is of the order

$$\begin{aligned}
n^{-1/2} \epsilon_n p_n^{1/2} (\log n) \epsilon_n^{-1} p_n^{-3} (\log n)^{-2} \log \log n &= n^{-1/2} p_n^{-5/2} (\log n)^{-1} \log \log n \\
&\asymp n^{-(1/2-5\tau/2)} (\log n)^{-1} \log \log n \\
&\rightarrow 0
\end{aligned}$$

by the assumed condition  $\tau < 1/5$ .

Finally, we need to control the deterministic bias  $\sqrt{n}(\psi_n(F_0) - (2\theta_0)^{-1})/\sigma_n$ . By Proposition 6, this is of the order

$$\begin{aligned}
n^{1/2} \epsilon_n p_n^{1/2} \log n \max(\epsilon_n, m_n/\epsilon_n) &\asymp n^{1/2-\tau/2} \max(m_n, \epsilon_n^2) \log n \\
&\asymp \max(n^{1/2-\tau/2-2\xi}, n^{1/2-\tau/2} m_n) \log n \\
&\rightarrow 0
\end{aligned}$$

by the assumption  $\tau/2 + 2\xi > 1/2$  and  $m_n = o(n^{-(1/2-\tau/2)}/\log n)$ . □

*Proof of Theorem 7.* Observe that by the definition of  $\mathbb{J}_n^*$  and  $\mathbb{J}_n$ ,

$$\mathbb{F}_n^* = \mathbb{F}_n + n^{-1/2} \mathbb{J}_n^* = F_0 + n^{-1/2} (\mathbb{J}_n + \mathbb{J}_n^*). \tag{3.9}$$

By a second-order Taylor expansion of  $\psi_n$  around  $F_0$ , and using the relations(1.4)–(1.10), for

some random  $0 \leq \xi_n^* \leq 1$  and two independent Kiefer processes  $K_1$  and  $K_2$ ,

$$\begin{aligned}
& \frac{\sqrt{n}}{\sigma_n} [\psi_n(\mathbb{F}_n) - \psi_n(F_0)] \\
&= \frac{1}{\sigma_n} \psi'_{n,F_0}(n^{-1/2}(K_1(F_0(\cdot), n)) + \psi'_{n,F_0}(n^{-1/2}(K_2(F_0(\cdot), n))) \\
&\quad + \frac{\psi'_{n,F_0}(\gamma_n)}{\sigma_n} + \frac{\phi''_{n,F_0, \mathbb{J}_n + \mathbb{J}_n^*}(\xi_n^*)}{2\sigma_n \sqrt{n}} \\
&= \frac{1}{\sigma_n} \psi'_{n,F_0}(n^{-1/2}(K_1(F_0(\cdot), n)) + \psi'_{n,F_0}(n^{-1/2}(K_2(F_0(\cdot), n))) \\
&\quad + O(\epsilon_n^{-1} p_n^{-1} \ell_n / \log n) \times \epsilon_n p_n^{1/2} (\log n) + O(n^{-1/2} p_n^{-2} (\log n)^{-1} \log \log n) \\
&= \frac{1}{\sigma_n} \psi'_{n,F_0}(n^{-1/2}(K_1(F_0(\cdot), n)) + \psi'_{n,F_0}(n^{-1/2}(K_2(F_0(\cdot), n))) \\
&\quad + O(n^{-(1/4-\tau/2)} (\log n)^{1/2} (\log \log n)^{1/4}) \\
&\quad + O(n^{2\tau-1/2} (\log n)^{-1} \log \log n) \tag{3.10}
\end{aligned}$$

as in (3.8), where  $\sup\{|\gamma_n(x)| : x \in \mathbb{R}\} = O(\ell_n)$  a.s., and  $\ell_n$  stands for  $n^{-1/4} (\log n)^{1/2} (\log \log n)^{1/4}$  here and throughout the rest.

Subtracting (3.8) from (3.10), we obtain, given the samples, almost surely for all posterior realizations,

$$\begin{aligned}
\frac{\sqrt{n}}{\sigma_n} [\psi_n(\mathbb{F}_n^*) - \psi_n(\mathbb{F}_n)] &= \frac{1}{\sigma_n} \psi'_n(n^{-1/2} K_2(F_0(\cdot), n)) \\
&\quad + O(n^{-(1/4-\tau/2)} (\log n)^{1/2} (\log \log n)^{1/4}) + O(n^{2\tau-1/2}).
\end{aligned}$$

The leading term has a standard normal distribution conditional on the samples, while the error term converges to zero almost surely under the assumed condition  $\tau < 1/4$ .  $\square$

*Proof of Corollary 4.* The event  $\theta_0 \in [Q_{\beta/2}, Q_{1-\beta/2}]$  holds if and only if

$$\beta/2 \leq \Pi(\psi_n(\mathbb{F}_n^*) \leq \theta_0 | D_n) \leq 1 - \beta/2.$$

By Theorem 6, the probability, under the true distribution  $F_0$  tends to  $1 - \beta$  if the assumed conditions hold.  $\square$

## CHAPTER

### 4

# EXTREME QUANTILE ESTIMATION FOR WEIBULL-TYPE DISTRIBUTIONS

## 4.1 Introduction

The goal of this chapter is to utilize the Weibull tail coefficient inferred in the previous chapter to further infer extreme quantiles and conditional tail expectations, which are useful measures in a large class of real-world applications, such as finance and insurance, among others. In this chapter, we work with regression data in which the response is  $Y$  and the covariate is  $\mathbf{X}$ . We're interested in the extreme quantile, using the intermediate quantile as a tool for inference. We assume the distribution of  $\log Y$  given  $\mathbf{X} = \mathbf{x}$  has a Weibull-type tail with WTC  $\theta$  for every  $\mathbf{x}$ . For the intermediate level quantile, we assume the  $\tau$ -th quantile of  $\log Y$  given  $\mathbf{X}$  is  $\mathbf{X}^T \boldsymbol{\beta}_\tau$  for some regression coefficient  $\boldsymbol{\beta}_\tau$ . Using these assumptions, we make an inference on the extreme quantile. To achieve the desired inference, we first introduce the notions of intermediate and extreme quantile levels. A method for estimating extreme quantiles for regression data based on minimizing an objective function was proposed in He et al. (2020), which we will use as a starting point for our Bayesian approach. Their approach involves estimating  $\boldsymbol{\beta}_\tau$ , which then yields an estimate of the intermediate quantile for any predictor value  $x$  via regression. To obtain an estimate of the extreme quantile, a

transformation map is applied to the intermediate quantile, using an estimate of the Weibull tail coefficient as the exponent of that factor. Our proposed method differs in that we apply quantile regression to  $\log Y$  rather than to  $Y$  directly.

In parallel, the Bayesian approach infers the unknown parameter  $\beta_\tau$  by sampling from the posterior distribution of  $\beta_\tau$ . As shown in Yu and Moyeed (2001), the combination of the noninformative uniform prior and the asymmetric Laplace likelihood leads to an asymptotically valid asymmetric Laplace posterior even if  $Y$  itself may not have an asymmetric Laplace distribution. We use the method in the previous chapter to infer the Weibull tail coefficient  $\theta$ . The inferences on  $\beta_\tau$  and  $\theta$  enable a similar calculation on the extreme quantile via a scaling factor on the intermediate quantile. The similarity of the frequentist and Bayesian approaches allows for direct comparison of estimates.

Additionally, we extend the application from extreme quantiles to conditional tail expectations, with the latter being coherent risk measures that are used in actuarial metrics, for example. We use simulated data for comparison of frequentist vs Bayesian approaches, using the root mean square error and coverage as comparison metrics. Lastly, we apply the proposed methodology on a real-life dataset.

## 4.2 Setup and Methodology

We call  $\tau_n$  as an intermediate quantile level if  $\tau_n \rightarrow 0$  and  $n\tau_n \rightarrow \infty$  as  $n \rightarrow \infty$ , and  $\nu_n$  as an extreme quantile level if  $\nu_n \rightarrow 0$  and  $\log \nu_n / \log \tau_n \rightarrow \kappa \in (1, \infty)$  as  $n \rightarrow \infty$ .

Let  $\mathbf{X}$  be a  $p$ -dimensional covariate with first element equal to 1 and  $Y$  be the dependent variable with data denoted  $(\mathbf{X}_i, Y_i), i = 1, \dots, n$ . Recall that the continuous conditional distribution of  $Y$  given  $\mathbf{X} = \mathbf{x}$  is denoted  $F_Y(y|\mathbf{x})$ , the corresponding survival function is denoted  $\bar{F}_Y(y|\mathbf{x}) = 1 - F_Y(y|\mathbf{x})$ , and the  $(1 - \tau)$ th conditional quantile of  $Y$  given  $\mathbf{X} = \mathbf{x}$  is denoted  $q_Y(\tau|\mathbf{x}) := \inf\{y : \bar{F}_Y(y|\mathbf{x}) \leq \tau\}$ .

### 4.2.1 He-Wang-Tong (HWT) Approach

For the baseline setting, we follow the setup proposed in He et al. (2020) to calculate the extreme quantile estimator and an asymptotic confidence interval around the true extreme quantile.

We first note that a consistent estimator for the linear quantile regression model

$$q_Y(\tau|\mathbf{x}) = \mathbf{x}^T \beta_\tau \tag{4.1}$$

for some quantile level  $\tau \in (0, 1)$  is given by

$$\hat{\boldsymbol{\beta}}_\tau = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i^T \boldsymbol{\beta}). \quad (4.2)$$

This quantity can be calculated using the `qr` function in the `quantreg` library in R. This allows the intermediate quantile to be estimated by

$$\hat{q}_Y(\tau|\mathbf{x}) := \mathbf{x}^T \hat{\boldsymbol{\beta}}_\tau. \quad (4.3)$$

Multiple estimators for the WTC were proposed in He et al. (2020). To facilitate comparison we use the constants weights estimator given by

$$\theta_{n,P}^c(\mathbf{x}) = \frac{\log(1/\tau_n)}{(J-1)\log(1/r)} \left\{ \log \hat{q}_Y(r^{J-1}\tau_n|\mathbf{x}) - \log \hat{q}_Y(\tau_n|\mathbf{x}) \right\}, \quad (4.4)$$

where  $J$  is a positive integer that can be treated as number of additional quantile levels for purposes in calculating this estimator for  $\theta$ , and  $r \in (0, 1)$ . We will use the shorthand  $\hat{\theta}_n$  for  $\theta_{n,P}^c(\mathbf{x})$  thereafter.

The extreme quantile estimator is given by

$$\hat{q}_n(\nu_n|\mathbf{x}) := \hat{q}_n(\tau_n|\mathbf{x}) \left( \frac{\log \nu_n}{\log \tau_n} \right)^{\hat{\theta}_n}. \quad (4.5)$$

We define  $\kappa_n = \log \nu_n / \log \tau_n$ .

By Theorem 2 in He et al. (2020), the asymptotic normal variance of  $\sqrt{n\tau_n}(\hat{\theta}_n - \theta)$  is

$$\sigma_\theta^2 = (\log r)^{-2} W^T \Sigma_{q(\mathbf{x})} W, \quad (4.6)$$

where  $W = (w_0 - w_1, \dots, w_{J-1} - w_J)^T$  with  $w_0 = w_J = 0$ ,  $w_j = 1/(J-1)$  for  $j = 1, \dots, J-1$ , and  $\Sigma_{q(\mathbf{x})}$  is calculated in step as follows:

- $s_j = r^{j-1}$  for  $j = 1, \dots, J$ ;
- $H(\mathbf{x}) = [K(\mu_{\mathbf{X}}/K(\mathbf{x}))]^\theta$ , with  $\mu_{\mathbf{X}} = \mathbb{E}(\mathbf{X})$ ;
- $\mathcal{Q}_{\mathbf{X}} = \mathbb{E}(\mathbf{X}\mathbf{X}^T)$ ;
- $\mathcal{Q}_H = \mathbb{E}((H(\mathbf{X}))^{-1}\mathbf{X}\mathbf{X}^T)$ ;
- $\Omega_1 = \mathcal{Q}_H^{-1}\mathcal{Q}_{\mathbf{X}}\mathcal{Q}_H^{-1}$ ;

- $(\Sigma_{q(\mathbf{x})})_{j,j'} = \theta^2(\mathbf{x}^T \Omega_1 \mathbf{x}) H^{-2}(\mathbf{x}) (\max(s_j, s_{j'}))^{-1}$  for  $j, j' = 1, \dots, J$ .

Theorem 4 in He et al. (2020) established asymptotic normality of the extreme estimator to the true extreme quantile, leading to the following approximate confidence interval for the extreme quantile:

$$\left( \frac{\hat{q}_Y(\nu_n|\mathbf{x})}{1+r_n}, \frac{\hat{q}_Y(\nu_n|\mathbf{x})}{1-r_n} \right), \quad (4.7)$$

where

$$r_n = \frac{z_{\alpha/2}}{\sqrt{n\tau_n}} \sigma_\theta \log \kappa_n. \quad (4.8)$$

## 4.2.2 Bayesian Approach

We now look at (4.1) from a Bayesian perspective. Suppose  $\sigma$  is a fixed scale parameter. If  $\pi(\boldsymbol{\beta}_\tau)$  is the prior on  $\boldsymbol{\beta}_\tau$ , then given data the posterior of  $\boldsymbol{\beta}_\tau$  can be written as

$$p(\boldsymbol{\beta}_\tau|\text{data}) \propto \pi(\boldsymbol{\beta}_\tau) \exp \left\{ - \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\tau) / \sigma \right\}. \quad (4.9)$$

By Yu and Moyeed (2001), the choice of the improper uniform prior on  $\boldsymbol{\beta}_\tau$  results in a proper posterior distribution. Hence, it is reasonable to consider the working likelihood

$$L(\boldsymbol{\beta}_\tau|\text{data}) = \frac{\tau^n (1-\tau)^n}{\sigma^n} \exp \left\{ - \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\tau) / \sigma \right\}. \quad (4.10)$$

This means that even if  $\boldsymbol{\beta}_\tau$  does not follow an asymmetric Laplace distribution (i.e. misspecified) in reality, we can still sample from the posterior of  $\boldsymbol{\beta}_\tau$  by assuming it follows ALD, and Sriram et al. (2013) demonstrated posterior consistency of the model parameters under certain sufficient conditions with the above likelihood setup.

According to Kotz et al. (2001), if  $U$  has an asymmetric Laplace distribution, then  $U$  has the following scale mixture of normals representation:

$$U = \sigma(\delta V + \epsilon \sqrt{V} Z), \quad (4.11)$$

where  $\delta = (1 - 2\tau)/(1 - \tau)$ ,  $\epsilon^2 = 2/[\tau(1 - \tau)]$ ,  $V$  and  $Z$  are respectively independent standard normal and standard exponential random variables. Here,  $\sigma$  could be fixed or a parameter for which a prior distribution could be assigned. Kozumi and Kobayashi (2011) utilized this

representation and proposed a Gibbs sampling algorithm to estimate the posterior. This algorithm underlies the parameter calculations in the `bayesQR` package in R—see Benoit and Van Den Poel (2017).

However, as noted by Yang et al. (2016), the posterior variance must be adjusted to obtain asymptotically valid interval estimates. Furthermore, this paper suggests that the variance adjustments make the results asymptotically invariant to the choice of fixed  $\sigma$ .

### 4.2.3 Conditional Tail Expectation

We also seek to infer the conditional tail expectation (CTE), possibly by expressing it in terms of extreme quantiles and other terms. First, as we saw earlier in Chapter 3, if  $W$  has a Weibull-type distribution, then for large  $t$ ,

$$\frac{\log \bar{F}(\zeta t)}{\log \bar{F}(t)} \approx \zeta^{1/\theta} \text{ implies } \bar{F}(\zeta t) \approx \bar{F}(t)\zeta^{1/\theta}.$$

Hence, the corresponding mean excess function can be approximated as

$$\begin{aligned} \text{E}[W - t | W > t] &= \int_t^\infty \frac{\bar{F}(z)}{\bar{F}(t)} dz \\ &\approx t \int_1^\infty \bar{F}(t)\zeta^{1/\theta-1} d\zeta \\ &= t \int_1^\infty \bar{F}(t)u^{-1}\theta u^{\theta-1} du \\ &= \frac{t\theta}{\bar{F}(t)} \int_1^\infty \exp\left\{-u \log \frac{1}{\bar{F}(t)}\right\} u^{\theta-1} du \\ &= \left(\log^2 \frac{1}{\bar{F}(t)}\right)^{-\theta} \frac{t\theta}{\bar{F}(t)} \int_{-\log \bar{F}(t)}^\infty e^{-v} v^{\theta-1} dv \\ &= \left(\log^2 \frac{1}{\bar{F}(t)}\right)^{-\theta} \frac{t\theta}{\bar{F}(t)} \Gamma\left(\theta; \log \frac{1}{\bar{F}(t)}\right) \end{aligned}$$

where  $\Gamma(\cdot; \cdot)$  is the upper incomplete gamma function. Taking  $t = q_Y(\nu_n | \mathbf{x})$  we have  $\bar{F}(t) = \nu_n$  and thus

$$\text{CTE}(\nu_n) \approx \left(\log^2 \frac{1}{\nu_n}\right)^{-\theta} \frac{q_Y(\nu_n | \mathbf{x})\theta}{\nu_n} \Gamma\left(\theta; \log \frac{1}{\nu_n}\right) + q_Y(\nu_n | \mathbf{x}). \quad (4.12)$$

Thus to infer  $\text{CTE}(\nu_n)$ , we use the HWT and Bayesian inference procedures for both  $\theta$  and  $q_Y(\nu_n | \mathbf{x})$ .

## 4.3 Numerical Results

In this section, we study the accuracy of the HWT and Bayesian quantile and CTE estimators. Generally, the estimation process takes the following high-level steps:

1. Infer  $\beta_\tau$  from the data;
2. Estimate the intermediate quantile in (4.3) using the inferred  $\beta_\tau$ ;
3. Compute  $\theta$  from the data and the inferred  $\beta_\tau$ ;
4. Calculate the extreme quantile in (4.5) and a corresponding confidence/credible region using the estimated intermediate quantile and the computed  $\theta$ ;
5. Approximate the conditional tail expectation in (4.12) and a corresponding confidence/credible region using the calculated extreme quantile and computed  $\theta$ .

The HWT approach was described in the previous section. The Bayesian inference would be performed using an Adaptive MCMC sampling with  $\beta_\tau$  corrections, further described below. The `bayesQR` package in R would be used to validate the posterior mean of  $\beta_\tau$ .

### 4.3.1 Adaptive Metropolis-Hastings (AMH) Method

Haario et al. (1999) introduced the adaptive proposal (AP) method, modifying the Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970) by using information from the simulation to adapt the proposal distribution to the target distribution. In the AP algorithm, the Gaussian proposal distribution has the current state as its mean and covariance based on a fixed number of previous states. Haario et al. (2001) proposed an adaptive Metropolis (AM) algorithm that enhanced the AP algorithm by having the covariance of the proposed distribution be calculated using all previous states. Roberts and Rosenthal (2009); Andrieu and Thoms (2008) described other adaptive MCMC algorithms in the literature. We will use the AM algorithm for the inference on  $\beta_\tau$ . If  $d$  is the dimension of  $\beta_\tau$ , Roberts et al. (1997); Gelman et al. (1996) have shown that the optimal symmetric jumping kernel is about  $2.4/\sqrt{d}$  times the scale of the target distribution, and this is used in our R implementation.

As mentioned previously, a correction has to be performed in order to obtain asymptotically valid credible intervals for  $\beta_\tau$ . The correction steps are as follows:

1. Using the MCMC samples from the uncorrected approach, compute the posterior mean vector  $\tilde{\beta}_\tau$  and the posterior dispersion matrix  $M$ ;
2. Compute  $V = M^{-1}$  and its positive definite square root  $V^{1/2}$ ;

3. Calculate the estimated asymptotic covariance matrix of the MLE given by

$$\Sigma = \left[ \tau^2 \sum_{i:Y_i > \tilde{\beta}_\tau^T X_i} X_i X_i^T + (1 - \tau)^2 \sum_{i:Y_i < \tilde{\beta}_\tau^T X_i} X_i X_i^T \right] / n; \quad (4.13)$$

4. Obtain the positive definite square root  $\Sigma^{1/2}$  of  $\Sigma$ ;
5. Compute the correction factor matrix  $C = V^{-1}\Sigma^{1/2}$ ;
6. For every previously sampled value of  $\beta_\tau$ , correct it to  $\beta_\tau^* = \tilde{\beta}_\tau + C(\beta_\tau - \tilde{\beta}_\tau)$ ;
7. Use the corrected posterior samples to construct credible sets.

We refer to the procedure described above as AMH.

### 4.3.2 bayesQR (BQR) Method

As an alternative Bayesian procedure, which we call BQR, we would utilize the `bayesQR` function in the `bayesQR` library in R to obtain  $\beta_\tau$  samples. A simple function call, analogous to the familiar `lm` function, allows  $\beta_\tau$  samples to be directly obtained. Due to some limitations inherent in the `bayesQR` function, we would only use this method to validate the posterior mean of  $\beta_\tau$  as obtained in the AMH method.

### 4.3.3 Simulated Dataset Results

We considered two data-generating models for which the true extreme quantile and conditional tail expectation can be calculated. This then allowed the comparison of estimates between the HWT and Bayesian approaches relative to the truth. For our simulations, we focused on  $p = 2$ , which allowed quantile plots to be visualized in two dimensions.

Data Model 1: The true underlying model is  $Y = Z \exp(\mathbf{X}^T \boldsymbol{\beta})$  where  $\boldsymbol{\beta}$  is an unknown parameter, and  $Z$  is a random variable with a Weibull tail distribution. We consider  $Z$  to have the following distributions in turn: truncated (positive) normal, truncated (positive) logistic, gamma, and Weibull. In addition, we consider three different Weibull shapes,  $k = 2/3, 1, 2$ , as the WTC  $\theta$  varies with shape.

Data Model 2: Let  $\xi_1, \xi_2 : [0, 1] \rightarrow [0, 1]$  be monotonically increasing, smooth, and onto functions, and define  $\xi = (\xi_2, \xi_1 - \xi_2)$ . Here, the true underlying model is  $Y = \{-\log[\mathbf{X}^T \xi]\}^\theta$ . We consider three different function pairs, labeled "Poly", "Trig", and "Beta", respectively:

- Poly:  $\xi_1(u) = u^2, \quad \xi_2(u) = u^3$ ;

- Trig:  $\xi_1(u) = \sin(\pi u)/2$ ,  $\xi_2(u) = 2 \arcsin(u)/\pi$ ;
- Beta:  $\xi_1(u) = \text{pbeta}(u, \text{shape1}=1, \text{shape2}=3)$ ,  $\xi_2(u) = \text{pbeta}(u, \text{shape1}=2, \text{shape2}=2)$ .

For both models, we simulated  $B = 1000$  number of data samples and 4000 MCMC draws were taken with a 60% burn-in under the Bayesian method. Sample sizes of  $n = 1000, 5000, 10000$  were used in the simulations. For the  $\psi$  functional parameters used in the inference for  $\theta$ , we used  $\tau = 1/6$  and  $\xi = 1/4$  as we used in the previous chapter.

We analyzed results under two sets of quantile levels:

- Quantile set A: Intermediate:  $\tau_n = 3n^{-1/3}$  and Extreme:  $\nu_n = 2n^{-2/3}$ ;
- Quantile set B: Intermediate:  $\tau_n = 10 \log(\log n)/n$  and Extreme:  $\nu_n = \log n/n^{1.01}$ .

Since no specific guidance was provided in He et al. (2020) on a suitable choice for the hyperparameters  $r$  and  $J$ , we chose  $r = 0.75$  and  $J = 2$  for purposes of comparison.

Because we're working with regression data, we propose analogs to the concepts of coverage and interval length we used in the previous chapters. We define the average proportion of containment (APC) as portion of the true quantile or CTE plot that lies within the credible region for a given data sample. For the interval-length counterpart, we compute the average length of the credible intervals around the plotting points for each data sample. We call this the average containment length (ACL) metric. For consistency, we use the same APC and ACL terminologies for  $\theta$  even though this parameter does not necessarily vary by  $x$ .

Generally, we present a set of three figures for each run, which is defined by the combination of data model and quantile level option:

1. Intermediate/extreme quantile and CTE plots for one data sample;
2. RMSE, bias, and RMSE efficiency box plots on extreme quantiles and CTE, across all replications;
3. RMSE and bias box plots on WTC across all replications.

In addition, a run produced output summary tables of ACL and APC for WTC, extreme quantile, and CTE under HWT and Bayesian methods, using the median and interquartile range (IQR) as summary measures. Under the Bayesian method, corrected and uncorrected ACL and APC summaries were shown.

It is instructive to observe the quantile and CTE plots for possible patterns under one data sample, and so we first analyzed the results for one sample under Data Model 1 - Gamma using quantile set A in an AMH run. In the quantile and CTE plots for a data sample in Figure 4.1, we observed that the true, HWT, and AMH intermediate quantile plots (broken

lines) were very close and nearly indistinguishable. This is true across all runs we have tested. However, the deviations from the true extreme quantile plot were more apparent. We also detected wider credible regions around the AMH extreme quantile plot than around the HWT extreme quantile plot. The second row shows the comparison of the true, HWT, and AMH CTE plots as well as the corresponding credible/confidence regions. Generally, the AMH extreme quantile and CTE plots were higher than their HWT counterparts. The reason for this is that while the intermediate estimated quantile plots were practically the same as the true intermediate quantile plot, the  $\theta$  estimates, which were used in Equation (4.5), were higher under the Bayesian approach.

The extreme quantile RMSE plot in Figure 4.2 showed that while the HWT RMSEs were less dispersed compared to the AMH RMSEs, the bulk of the RMSEs under the Bayesian method were lower and closer to zero, particularly at higher sample sizes. This observation was validated by the relative RMSE efficiency box plot on the right. The relative RMSE efficiency is the ratio of the HWT RMSE over the AMH RMSE for each data sample. Thus a ratio greater than 1 favored the Bayesian approach, which was clearly the case for higher sample sizes. The Bayesian bias was also more dispersed than the HWT bias, but the latter plot sat entirely on the negative side, while the bulk of the AMH bias included 0 and appeared to be shrinking as sample size increased. Similar observations could be made on the CTE row.

In Figure 4.3, we also compared the estimation of the WTC between HWT and AMH. Both RMSE and WTC box plots supported the effectiveness of the AMH approach compared to HWT. The WTC underestimation by the  $\theta$  estimator under HWT manifested as negative bias, which in turn explained the underestimation of the extreme quantile by the corresponding HWT estimator. In Table 4.1, we compared the coverage (APC) and interval lengths (ACL) metrics under the HWT and Bayes approaches. We used the median and the interquartile range (IQR) as summary measures. The HWT confidence intervals, which tended to be relatively shorter, did not cover the true  $\theta = 1$  value. This was not an issue under the Bayes approach however, achieving coverage of 1.00, though requiring longer median interval lengths to do so.

While the plots in Figure 4.1 suggested some observations on APL and APC for one data sample, a more robust approach would be to show these metrics over the  $B$  samples, as can be seen in Table 4.2. Like the previous table for WTC coverage and length, we used the median and IQR summary measures in order to avoid outliers. In this table we also showed the impact of the correction on the posterior variance of  $\beta$ . We noticed that there were no salient differences between the uncorrected and corrected plots for either extreme quantiles and CTE. We also observed that the HWT tended to produce really narrow confidence

intervals that prevented the true value to be contained particularly at higher sample sizes. The Bayes approach contained the APC each time, though at the expense of a significantly higher median ACL. Similar conclusions can be drawn on the CTE plots in this table.

The results discussed thus far used quantile set A, which could be viewed as "less extreme" than quantile set B, a variation of the quantile set used in He et al. (2020). Comparing Figure 4.1 and Figure 4.4, we immediately saw that the confidence and credible bands were much narrower under quantile set B. In Figure 4.5 and Figure 4.6, we arrived at the same conclusions as under quantile set A with respect to RMSE, bias, and RMSE efficiency for extreme quantiles, CTE, and WTC. The APC/ACL tables in Table 4.3 and Table 4.4 also suggested perfect coverage of  $\theta$  and the true quantile and CTE plots with sufficiently wide median ACL. One could also perceive a significant decrease in ACL in both median and IQR under quantile set B. Finally, the posterior variance correction still did not yield any material difference in the results.

Now we look at results from Data Model 2 - Beta under quantile set A in an AMH run, specifically the beta pair model with  $\theta = 0.5$ . The true quantile model here is not linear in  $x$ , unlike those considered under Data Model 1, and this can be clearly observed in Figure 4.7. For this particular data sample, the HWT extreme quantile and CTE plots were lower than their AMH counterparts, as previously observed in the gamma data model results, and this held across all runs, again due to the inferred  $\theta$  under AMH being larger than that under HWT. In Figure 4.8, we make the same observations as in the gamma case in Data Model 1, where (1) the bias plots under HWT were mostly negative, while the plots under AMH contained 0; and (2) the bulk of the RMSE efficiency plots were trending to be above 1, favoring the Bayes RMSE results. The RMSE plot in Figure 4.9 suggested improving RMSE under the AMH approach, while the opposite is true for HWT. For the bias plots, the HWT bias box plot showed a downward trend away from 0, while the Bayes bias box plot appeared to include 0 still. Furthermore, the bias was clearly negative under HWT, but this was not the case under AMH. Finally, we observed once more that the narrow median ACL interval lengths to cover the true  $\theta$ , extreme quantile, or CTE were not sufficient, which clearly contrasted the results under the Bayesian approach.

For completeness, the remaining run results for the other data models under the two quantile sets considered are shown in the Appendices. In general, the conclusions from the other results mirror the conclusions described above for gamma and beta.

Figure 4.1: Extreme quantile and CTE plots for a sample simulation from DM 1 - Gamma, with  $\theta = 1.0$  using the HWT and AMH methods under quantile set A.

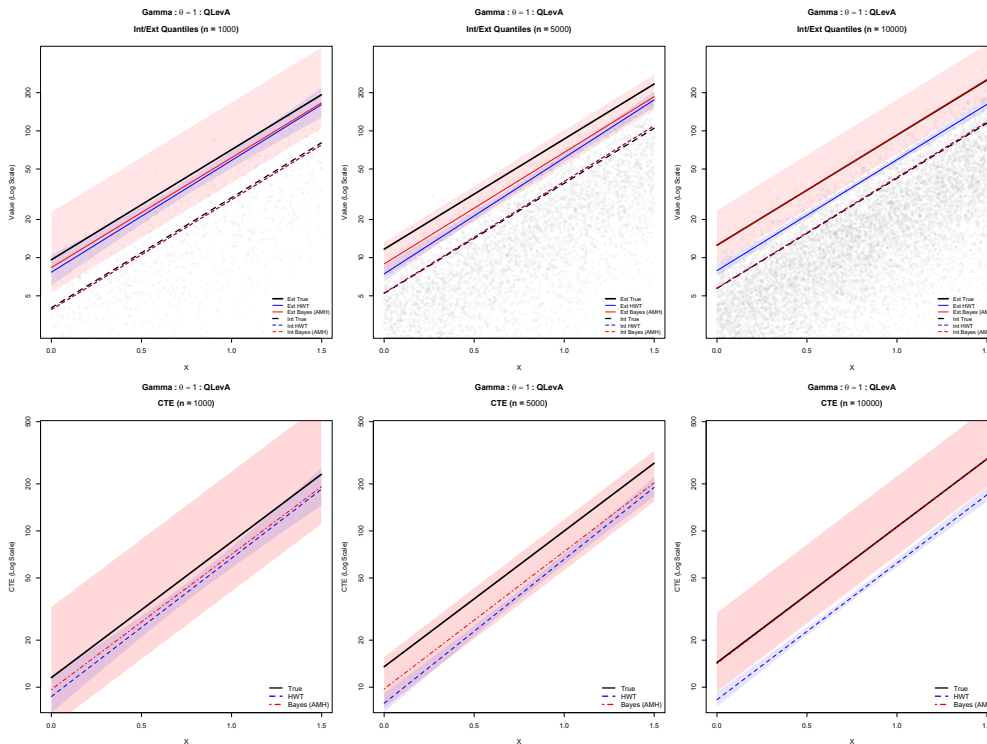


Table 4.1: Coverage and length results on WTC for simulations from DM 1 - Gamma, with  $\theta = 1.0$  using the HWT and AMH methods under quantile set A.

<b>n</b>	<b>Method</b>	<b>APC Median [IQR]</b>	<b>ACL Median [IQR]</b>
1000	HWT	0.000 [0.000]	0.313 [0.003]
	Bayes	1.000 [0.000]	1.766 [2.043]
5000	HWT	0.000 [0.000]	0.183 [0.001]
	Bayes	1.000 [0.000]	1.151 [0.900]
10000	HWT	0.000 [0.000]	0.145 [0.000]
	Bayes	1.000 [0.000]	0.998 [0.705]

Figure 4.2: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Gamma, with  $\theta = 1.0$  using the HWT and AMH methods under quantile set A.

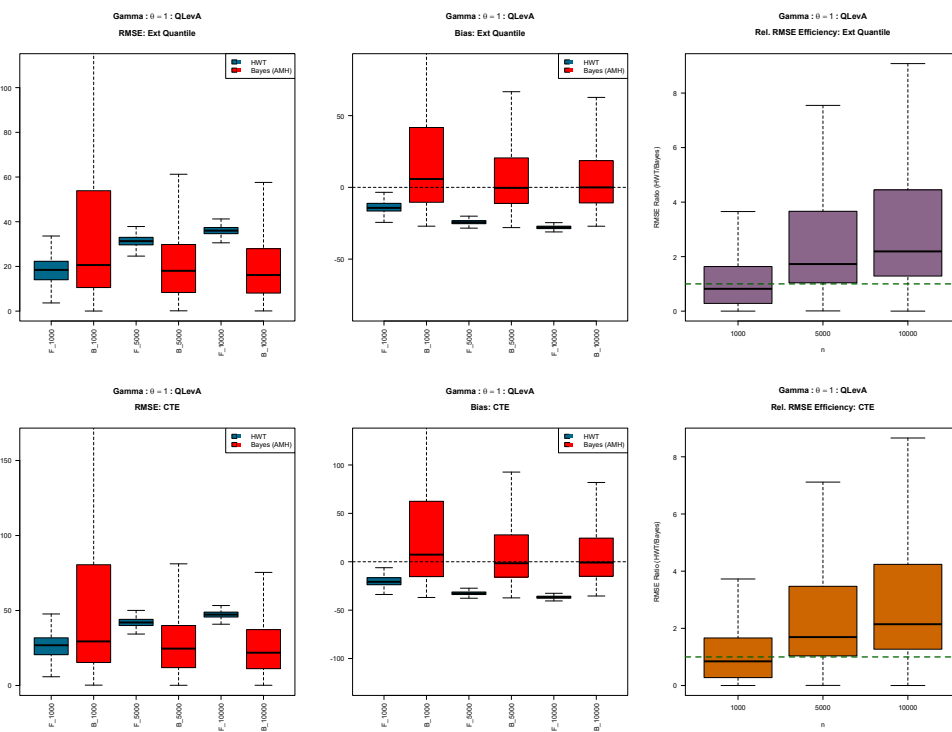


Figure 4.3: RMSE and bias box plots on WTC for simulations from DM 1 - Gamma, with  $\theta = 1.0$  using the HWT and AMH methods under quantile set A.

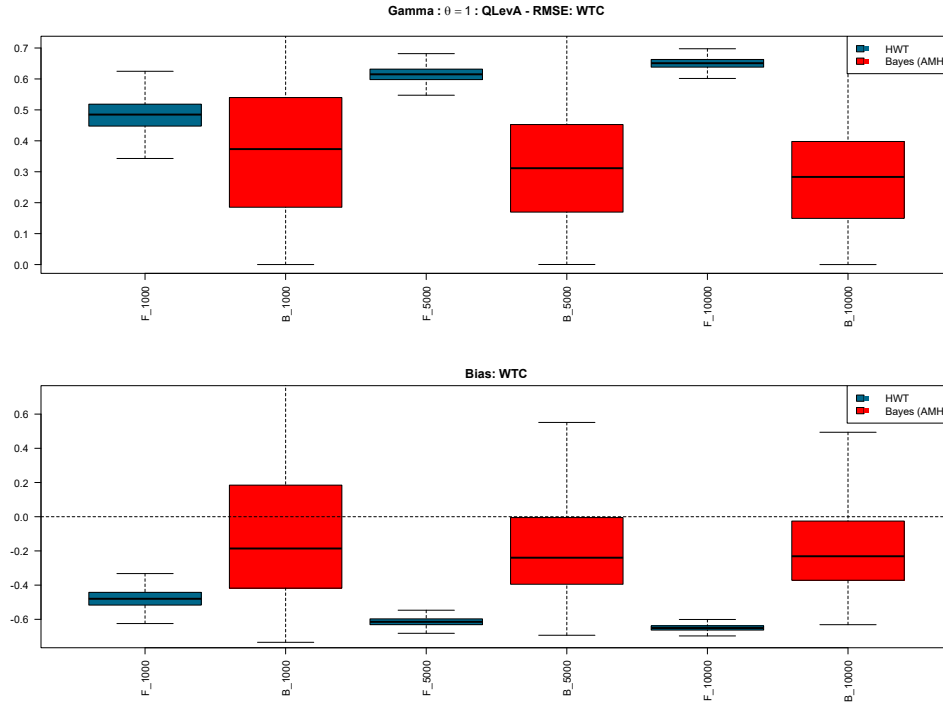


Figure 4.4: Extreme quantile and CTE plots for a sample simulation from DM 1 - Gamma, with  $\theta = 1.0$  using the HWT and AMH methods under quantile set B.

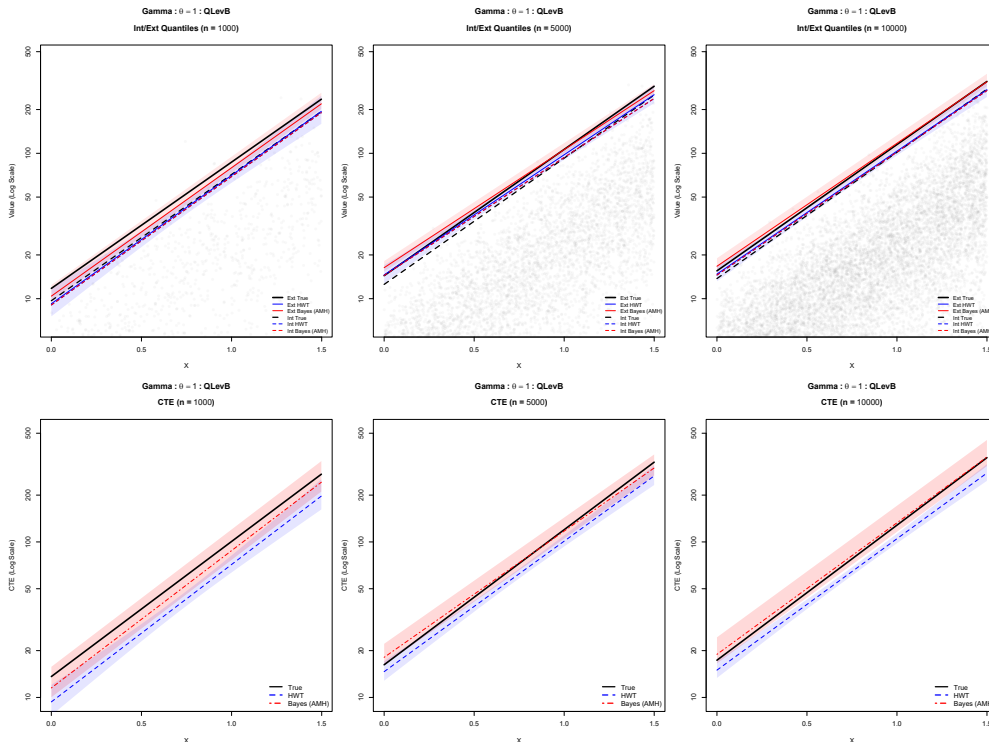


Table 4.2: Coverage and length results on extreme quantile and CTE for simulations under DM 1 - Gamma, with  $\theta = 1.0$  using the HWT and AMH methods under quantile set A.

<b>n</b>	<b>Metric</b>	<b>Method</b>	<b>APC Median [IQR]</b>	<b>ACL Median [IQR]</b>
1000	Quantile	HWT	0.230 [0.230]	19.247 [2.373]
		Bayes Unc	1.000 [0.000]	258.118 [1399.051]
		Bayes Corr	1.000 [0.000]	258.143 [1399.190]
	CTE	HWT	0.125 [0.225]	21.539 [3.123]
		Bayes Unc	1.000 [0.000]	454.868 [3931.150]
		Bayes Corr	1.000 [0.000]	454.865 [3931.337]
5000	Quantile	HWT	0.000 [0.000]	10.369 [0.481]
		Bayes Unc	1.000 [0.000]	115.310 [204.349]
		Bayes Corr	1.000 [0.000]	115.325 [204.208]
	CTE	HWT	0.000 [0.000]	11.103 [0.581]
		Bayes Unc	1.000 [0.000]	166.570 [348.312]
		Bayes Corr	1.000 [0.000]	166.526 [348.168]
10000	Quantile	HWT	0.000 [0.000]	8.213 [0.293]
		Bayes Unc	1.000 [0.000]	96.587 [138.272]
		Bayes Corr	1.000 [0.000]	96.569 [138.309]
	CTE	HWT	0.000 [0.000]	8.688 [0.353]
		Bayes Unc	1.000 [0.000]	134.531 [219.550]
		Bayes Corr	1.000 [0.000]	134.513 [219.619]

Table 4.3: Coverage and length results on WTC for simulations from DM 1 - Gamma, with  $\theta = 1.0$  using the HWT and AMH methods under quantile set B.

<b>n</b>	<b>Method</b>	<b>APC Median [IQR]</b>	<b>ACL Median [IQR]</b>
1000	HWT	0.130 [0.135]	1.233 [0.011]
	Bayes	1.000 [0.000]	1.701 [1.980]
5000	HWT	0.045 [0.100]	1.171 [0.004]
	Bayes	1.000 [0.000]	1.131 [0.857]
10000	HWT	0.015 [0.070]	1.151 [0.003]
	Bayes	1.000 [0.000]	1.020 [0.664]

Figure 4.5: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Gamma, with  $\theta = 1.0$  using the HWT and AMH methods under quantile set B.

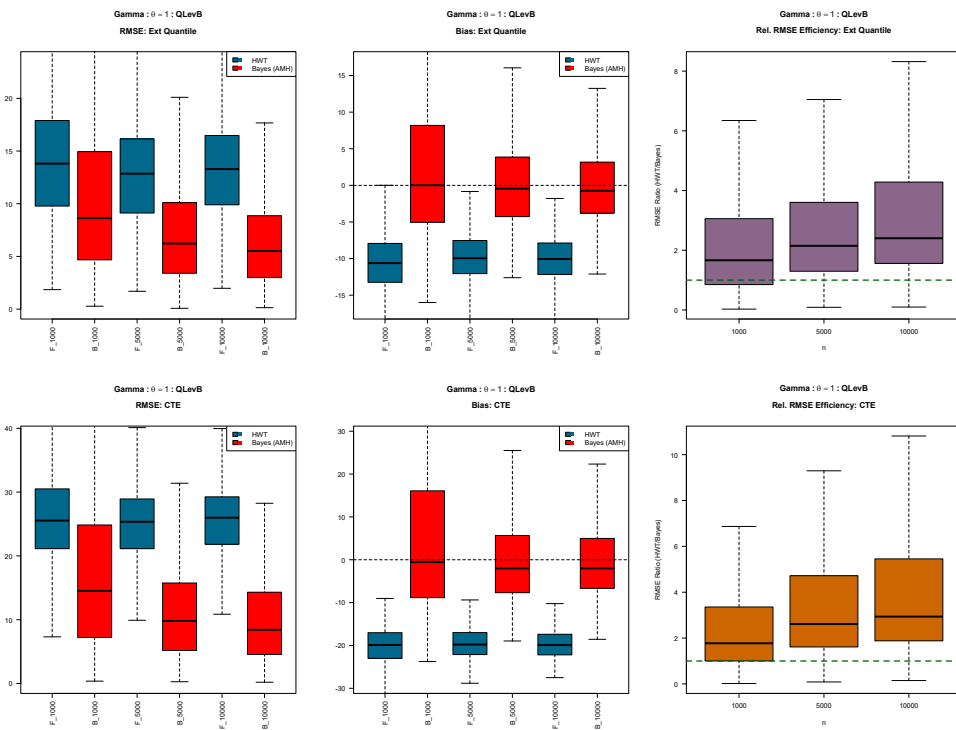


Figure 4.6: RMSE and bias box plots on WTC for simulations from DM 1 - Gamma, with  $\theta = 1.0$  using the HWT and AMH methods under quantile set B.

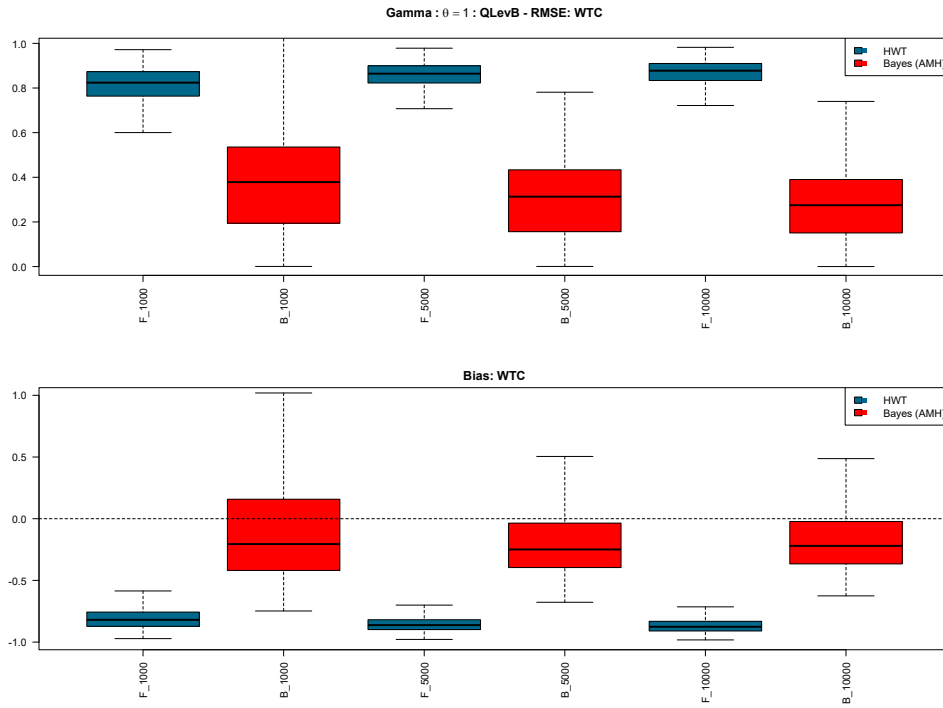


Figure 4.7: Extreme quantile and CTE plots for a sample simulation from DM 1 - Beta, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set A.

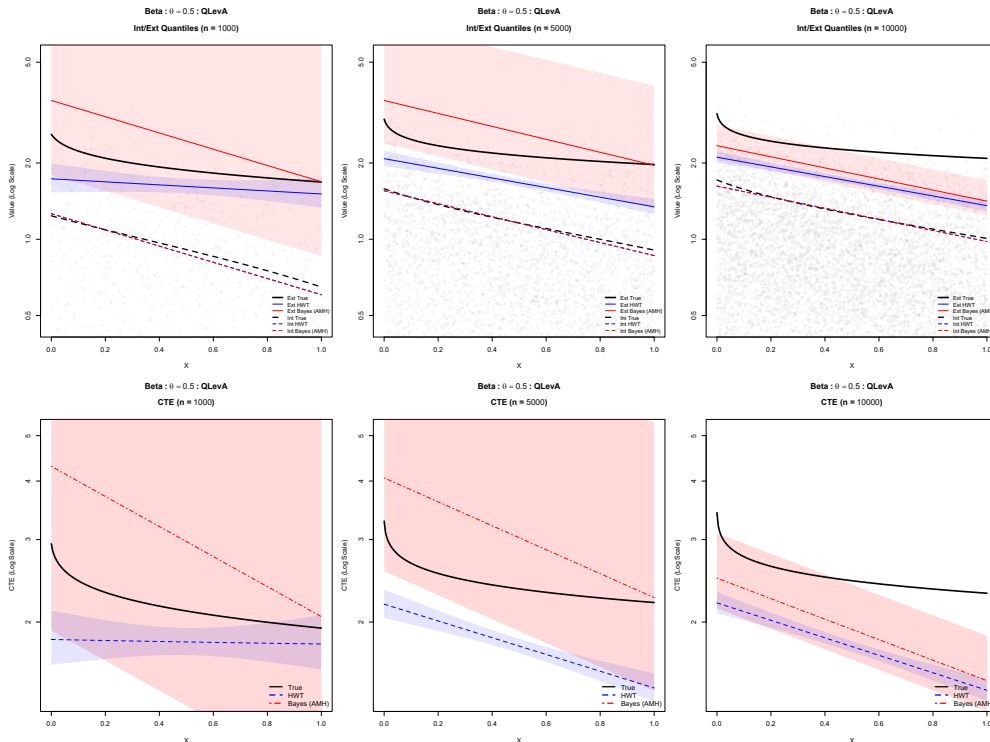


Table 4.4: Coverage and length results on extreme quantile and CTE for simulations under DM 1 - Gamma, with  $\theta = 1.0$  using the HWT and AMH methods under quantile set B.

<b>n</b>	<b>Metric</b>	<b>Method</b>	<b>APC Median [IQR]</b>	<b>ACL Median [IQR]</b>
1000	Quantile	HWT	0.445 [0.435]	21.329 [1.921]
		Bayes Unc	1.000 [0.140]	34.013 [53.716]
		Bayes Corr	1.000 [0.140]	33.991 [53.721]
	CTE	HWT	0.090 [0.201]	22.038 [2.528]
		Bayes Unc	1.000 [0.000]	77.309 [177.113]
		Bayes Corr	1.000 [0.000]	77.329 [177.124]
5000	Quantile	HWT	0.325 [0.275]	17.497 [1.020]
		Bayes Unc	1.000 [0.276]	18.258 [15.474]
		Bayes Corr	1.000 [0.275]	18.248 [15.451]
	CTE	HWT	0.000 [0.075]	17.855 [1.296]
		Bayes Unc	1.000 [0.000]	39.661 [39.361]
		Bayes Corr	1.000 [0.000]	39.582 [39.495]
10000	Quantile	HWT	0.280 [0.265]	16.226 [0.836]
		Bayes Unc	1.000 [0.320]	15.068 [11.025]
		Bayes Corr	1.000 [0.321]	15.106 [11.071]
	CTE	HWT	0.000 [0.027]	16.476 [1.046]
		Bayes Unc	1.000 [0.000]	33.096 [28.062]
		Bayes Corr	1.000 [0.000]	33.151 [28.050]

Table 4.5: Coverage and length results on WTC for simulations from DM 1 - Beta, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set A.

<b>n</b>	<b>Method</b>	<b>APC Median [IQR]</b>	<b>ACL Median [IQR]</b>
1000	HWT	0.520 [0.645]	0.157 [0.001]
	Bayes	1.000 [0.000]	1.207 [1.154]
5000	HWT	0.000 [0.051]	0.092 [0.000]
	Bayes	1.000 [0.000]	0.781 [0.558]
10000	HWT	0.000 [0.000]	0.073 [0.000]
	Bayes	1.000 [0.000]	0.647 [0.361]

Figure 4.8: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Beta, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set A.

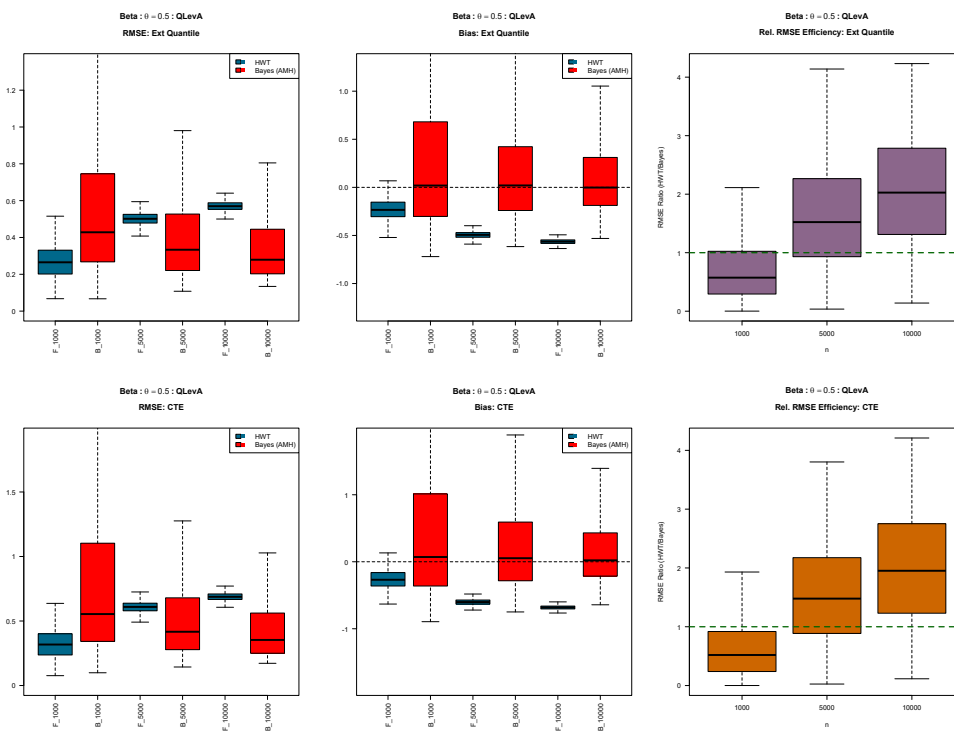


Figure 4.9: RMSE and bias box plots on WTC for simulations from DM 1 - Beta, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set A.

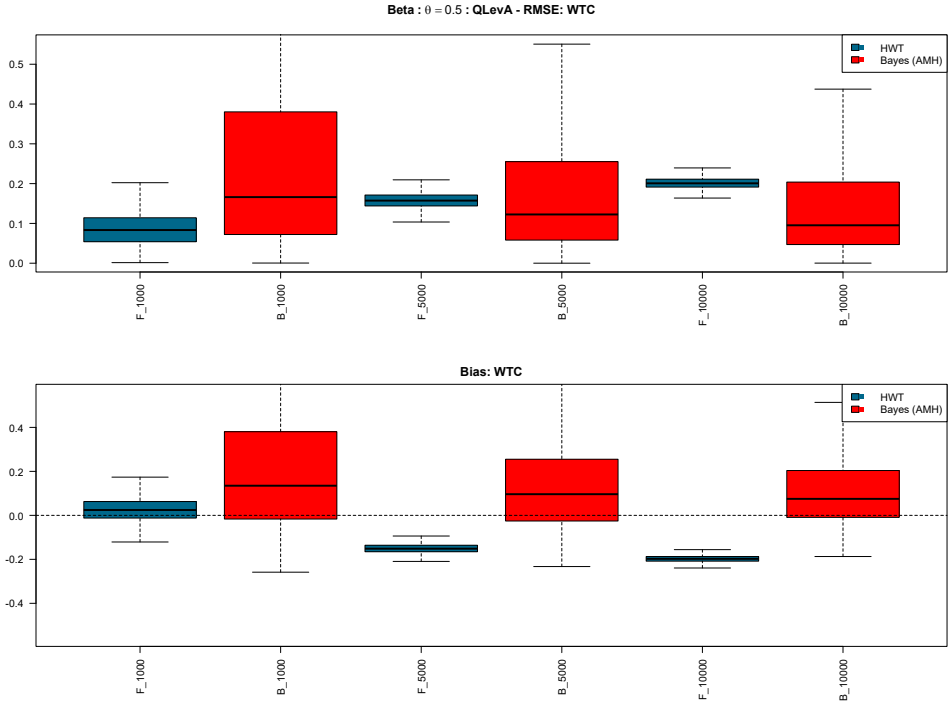


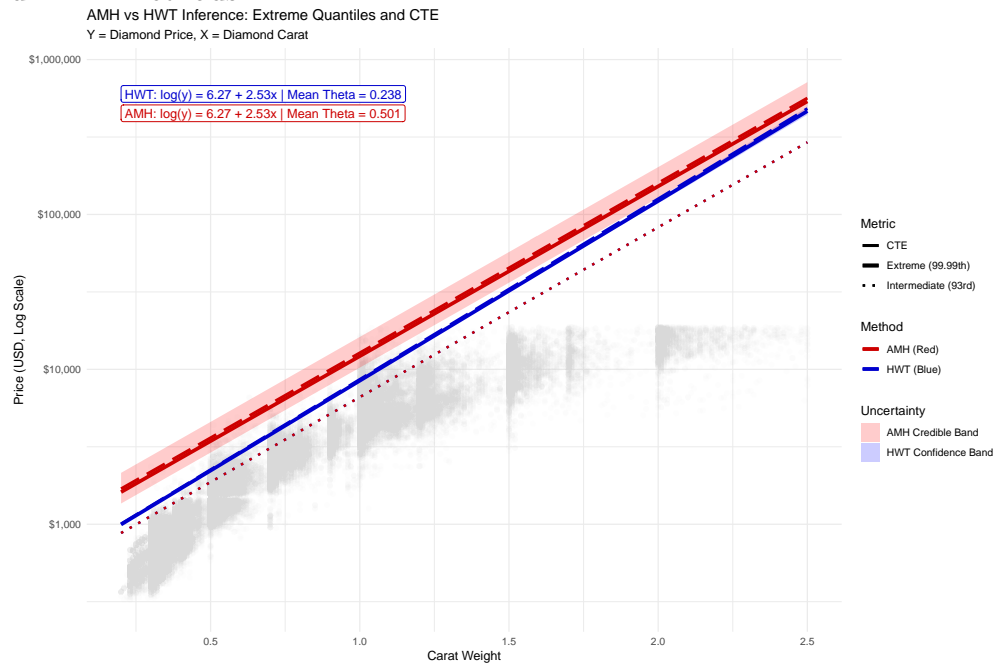
Table 4.6: Coverage and length results on extreme quantile and CTE for simulations under DM 1 - Beta, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set A.

<b>n</b>	<b>Metric</b>	<b>Method</b>	<b>APC Median [IQR]</b>	<b>ACL Median [IQR]</b>
1000	Quantile	HWT	0.300 [0.501]	0.316 [0.028]
		Bayes Unc	1.000 [0.000]	3.928 [9.932]
		Bayes Corr	1.000 [0.000]	3.932 [9.944]
	CTE	HWT	0.295 [0.480]	0.354 [0.037]
		Bayes Unc	1.000 [0.000]	6.003 [19.159]
		Bayes Corr	1.000 [0.000]	6.007 [19.177]
5000	Quantile	HWT	0.000 [0.000]	0.165 [0.005]
		Bayes Unc	1.000 [0.000]	2.095 [2.648]
		Bayes Corr	1.000 [0.000]	2.094 [2.647]
	CTE	HWT	0.000 [0.000]	0.175 [0.006]
		Bayes Unc	1.000 [0.000]	2.838 [3.973]
		Bayes Corr	1.000 [0.000]	2.837 [3.973]
10000	Quantile	HWT	0.000 [0.000]	0.129 [0.003]
		Bayes Unc	1.000 [0.000]	1.668 [1.505]
		Bayes Corr	1.000 [0.000]	1.667 [1.505]
	CTE	HWT	0.000 [0.000]	0.135 [0.003]
		Bayes Unc	1.000 [0.000]	2.187 [2.149]
		Bayes Corr	1.000 [0.000]	2.188 [2.149]

### 4.3.4 Real Life Dataset Results

Here we apply the HWT and AMH methodologies to diamonds data found in the `diamonds` dataset in the `ggplot2` library in R. We take  $X$  as the diamond carat weight and regress against the diamond price which we take as the response  $Y$  variable. We limit the analysis to data points with carat weight less than or equal to 2.5. The plot in Figure 4.10 shows the extreme quantile and CTE plots under an intermediate quantile level of 93% and an extreme quantile level of 99.99%. As observed previously, the intermediate quantile plots are overlapping and the HWT extreme quantile and CTE plots are lower than their AMH counterparts due to average HWT  $\theta$  estimate being lower than the mean AWT  $\theta$ . Due to the somewhat high extreme quantile level, the mean excess associated with this level is very small, which manifested in the close extreme quantile and CTE plots.

Figure 4.10: Extreme quantile and CTE of diamond price as function of carat under the HWT and AMH methods.



## CHAPTER

### 5

# CONCLUSION

## 5.1 Summary

In Chapter 2, we presented a nonparametric Bayesian approach that allowed for the inference of the extreme value index  $\gamma \neq 0$ . We constructed sequences of functionals that approximated  $\gamma$ . On the theoretical side, we established a Bernstein-von Mises theorem that provided a posterior contraction rate and assured asymptotic frequentist coverage. The accompanying calculations were implemented using Bayesian bootstrap simulations using fixed hyperparameters  $\tau$ ,  $\xi$ , and  $\xi'$  for various distributions. Comparison of the Bayes estimator against the Pickands and Hill/Negative Hill estimators were performed under coverage, credible/confidence interval length, and bias metrics. These metric comparisons were done for distributions under select parameters so that the true EVI can be determined. The Bayes estimator was found to be strong under certain distributions compared to others and more effective under certain parameter subspaces (e.g. parameters that make the tail heavier).

For the case  $\gamma = 0$ , we proposed a modified Bayes factor that allowed the comparison of model  $M_0$  denoting unrestricted cdf  $F$  and model  $M_1$  denoting  $F$  in the Gumbel MDA. A limit characterization on distributions with zero EVI allowed a region that can be defined and associated with model  $M_1$ . This enabled the comparison of the modified marginal likelihood in  $M_0$  and the modified empirical likelihood for  $M_1$ , with sampling for the latter performed

using Hamiltonian Monte Carlo. An optimal hyperparameter set that generally associated positive (negative) logarithm of the Bayes factor with Gumbel (non-Gumbel) MDA was obtained using training data and validated with independent data, indicating viability of the technique.

A similar nonparametric Bayesian approach was discussed in Chapter 3 to infer the Weibull tail coefficient  $\theta$ . The WTC is defined for distributions in the Gumbel MDA, which consists of the well-known lighter-tailed distributions such as the normal, gamma, logistic, and Weibull. A sequence of functionals was also constructed that approximated  $\theta$ , and a Bernstein-von Mises theorem was established. The resulting Bayes estimator was compared to Beirlant and Girard-Hill estimators in terms of coverage and confidence/credible interval lengths. Bayesian bootstrap simulations were performed for various distributions that are in the Gumbel MDA. Perhaps unsurprisingly, the Bayes estimator was demonstrated to perform very well in estimating  $\theta$  for the Weibull distribution (along with its tested variants) against the other estimators in terms of both coverage and interval lengths. Bayes estimation for the logistic WTC of 1 and normal WTC of 0.5 also performed comparably well. The estimation for the gamma distribution was not as successful however, owing to additional constraints not satisfied on top of regularity, which affected coverage to some degree.

In Chapter 4, we built upon the Bayesian inference approach to the WTC described in Chapter 3 to estimate extreme quantiles and conditional tail expectations. Here we considered a regression data setting, assuming that the logarithm of the response variable  $Y$  given covariate  $\mathbf{X} = \mathbf{x}$  has a Weibull-type tail with WTC  $\theta$  for every  $\mathbf{x}$ , and that the  $\tau$ -th intermediate quantile of  $\log Y$  given  $\mathbf{X}$  is  $\mathbf{X}^T \boldsymbol{\beta}_\tau$  for some coefficient  $\boldsymbol{\beta}_\tau$ . We mirrored the framework described in He et al. (2020) and utilized their approach, named HWT in Chapter 4, for comparison. The  $\boldsymbol{\beta}_\tau$  coefficient was estimated using `qr` in the `quantreg` library in R for the HWT approach, while under Bayes,  $\boldsymbol{\beta}_\tau$  was primarily inferred using an adaptive Metropolis-Hastings (AMH) approach, and the built-in package `bayesQR` in R was used as a secondary check to the posterior mean calculations. In general, the estimators for  $\boldsymbol{\beta}_\tau$  under the three calculations are very close, leading to intermediate quantile plots that largely overlap. Under the Bayes AMH approach, we applied an adjustment to  $\boldsymbol{\beta}_\tau$  for the sole purpose of correcting the asymptotic posterior variance. Utilizing the extreme quantile estimator in Equation (4.5), we see that the inference on the WTC  $\theta$  played a big role in the estimation and largely explained the differences among the extreme quantile estimates. Metrics for comparison included RMSE and bias box plots as well as regression counterparts of coverage vs interval length summarized using median and IQR in tables. A recurring theme observed in the calculations is the apparent underestimation of  $\theta$  using the HWT approach relative to Bayes estimator. In general, the simulated test results provide evidence of viability of the

proposed Bayesian approach. We also extended the analysis to include the conditional tail expectation by introducing an approximation for the CTE expressed this quantity in terms of the extreme quantile and WTC. Finally, we applied the HWT and AMH approaches to a real-life dataset.

## 5.2 Further Research

For Chapters 2 and 3, with their similar methodologies, the calculation of  $a_n$  and  $\epsilon_n$  to be data-driven could be an interesting extension to the calculations related to the  $\psi$  functional. The functional sequence itself could perhaps be revised that improves convergence rates and minimize regularity conditions. A theoretical justification of the modified Bayes factor for testing  $\gamma = 0$  could also be explored.

As hinted at Chapter 4, while the numerical results prove promising especially in comparison with the HWT approach, further research can be made on the theoretical underpinnings of the proposed methodology.

The theory of extreme value analysis is very well-developed in the iid case, and to a lesser extent the dependent case. In this vein, one can also study extreme quantile inference on dependent observations coming from an ergodic stationary Markov chain. Another direction is extending the univariate setting to multivariate observations and joint extreme quantiles, say for example, studying both hurricane wind speed and the associated damage dollar amount.

## REFERENCES

- Almalki, S. and Nadarajah, S. (2014). Modifications of the weibull distribution: A review. *Reliability Engineering and System Safety*, 124:32–55.
- Almelor, G. and Ghosal, S. (2026a). Bayesian inference for the extreme value index. Manuscript submitted for publication.
- Almelor, G. and Ghosal, S. (2026b). Bayesian inference for the weibull tail coefficient. Manuscript submitted for publication.
- Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive mcmc. *Statistics and Computing*, 18:343–373.
- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.
- Asmussen, S. and Steffensen, M. (2020). *Risk and Insurance: A Graduate Text*. Springer Nature.
- Beirlant, J., Broniatowski, M., Teugels, J. L., and Vynckier, P. (1995). The mean residual life function at great age: Applications to tail estimation. *Journal of Statistical Planning and Inference*, 45:21–48.
- Benoit, D. and Van Den Poel, D. (2017). bayesqr: A bayesian approach to quantile regression. *Journal of Statistical Software*, 76(7):1–32.
- Broniatowski, M. (1993). On the estimation of the weibull tail coefficient. *Journal of Statistical Planning and Inference*, 35:349–366.
- Chernozhukov, V. (2005). Extreme quantile regression. *Annals of Statistics*, 33(2):806–839.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, London.
- Csorgo, M. and Revesz, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, Inc., Budapest.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer Science & Business Media, New York.
- Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997). *Modeling Extremal Events*. Springer Science & Business Media, Heidelberg.
- Embrechts, P., Resnick, S., and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3(2):30–41.
- Falk, M. (1995). Some best estimators for distributions with finite endpoint. *Statistics: A Journal of Theoretical and Applied Statistics*, pages 115–125.

- Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230.
- Fisher, R. and Tippett, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, 24:180–190.
- Friel, N. and Pettitt, A. N. (2008). Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(3):589–607.
- Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*. Robert E. Krieger Publishing Co., Inc.
- Gardes, L. and Girard, S. (2006). Comparison of weibull tail-coefficient estimators. *REVSTAT – Statistical Journal*, 4:163–188.
- Gardes, L. and Girard, S. (2008). Estimation of the weibull tail-coefficient with linear combination of upper order statistics. *Journal of Statistical Planning and Inference*, 138:1416–1427.
- Gardes, L. and Girard, S. (2016). On the estimation of the functional weibull tail-coefficient. *Journal of Multivariate Analysis*, 146:29–45.
- Gelman, A., Carlin, J., and Stern, H. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton.
- Gelman, A., Roberts, G., and Gilks, W. (1996). Efficient metropolis jumping rules. *Bayesian Statistics*, 5:599–607.
- Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, Cambridge.
- Girard, S. (2004). A hill type estimator of the weibull tail coefficient. *Communications in Statistics - Theory and Methods*, 33:205–234.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une serie aleatoire. *Annals of Mathematics*, 44:423–453.
- Haario, H., Saksman, E., and Tamminen, J. (1999). Adaptive proposal distribution for random walk metropolis algorithm. *Computational Statistics*, 14:375–395.
- Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242.
- Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.
- He, F., Wang, H. J., and Tong, T. (2020). Extremal linear quantile regression with weibull-type tails. *Statistica Sinica*, 30(3):1357–1377.

- Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3:1163–1174.
- Jeffreys, H. (1935). Some tests of significance treated by the theory of probability. *Proceedings of the Cambridge Philosophy Society*, 31:203–222.
- Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.
- Koenker, R. and Bassett Jr., G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.
- Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent rv'-s, and the sample df. i. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32:111–131.
- Kotz, S., Kozubowski, T., and Podgorski, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhauser.
- Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81(11):1565–1578.
- Leadbetter, M., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York.
- Lo, A. Y. (1987). A large sample study of the Bayesian bootstrap. *The Annals of Statistics*, 15:360–375.
- Lo, A. Y. (1991). Bayesian bootstrap clones and a biometry function. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 320–333.
- McNeil, A., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management*. Princeton University Press.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1091.
- Neal, R. (2012). *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, Boca Raton.
- Newton, M. A. and Raftery, A. E. (1994). Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 56(1):3–26.
- O'Hagan, A. (1995). Fractional bayes factors for model comparison. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):99–118.
- Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

- Padoan, S. and Rizzelli, S. (2022). Consistency of bayesian inference for multivariate max-stable distributions. *The Annals of Statistics*, 50(3):1490–1518.
- Padoan, S. and Rizzelli, S. (2024). Empirical bayes inference for the block maxima method. *Bernoulli*, 30(3):2154–2184.
- Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119–131.
- Resnick, S. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag.
- Roberts, G., Gelman, A., and Gilks, W. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120.
- Roberts, G. and Rosenthal, J. (2009). Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, 18(2):349–367.
- Rubin, D. (1981). The Bayesian bootstrap. *The Annals of Statistics*, 9:130–134.
- Sriram, K., Ramamoorthi, R., and Ghosh, P. (2013). Posterior consistency of bayesian quantile regression based on the misspecified asymmetric laplace density. *Bayesian Analysis*, 8(2):479–504.
- Tokdar, S., Jiang, S., and Cunningham, E. (2022). Heavy-tailed density estimation. *Journal of the American Statistical Association*, 00:1–13.
- Van Der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.
- Yang, Y., Wang, H. J., and He, X. (2016). Posterior inference in bayesian quantile regression with asymmetric laplace likelihood. *International Statistical Review*, 84(3):327–344.
- Yu, K. and Moyeed, R. (2001). Bayesian quantile regression. *Statistics and Probability Letters*, 54:437–447.

## APPENDICES

## APPENDIX

### A

# CHAPTER 4 SUPPLEMENTARY MATERIAL - DATA MODEL 1

## A.1 Data Model 1 - Normal

Figure A.1: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Normal, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set A.

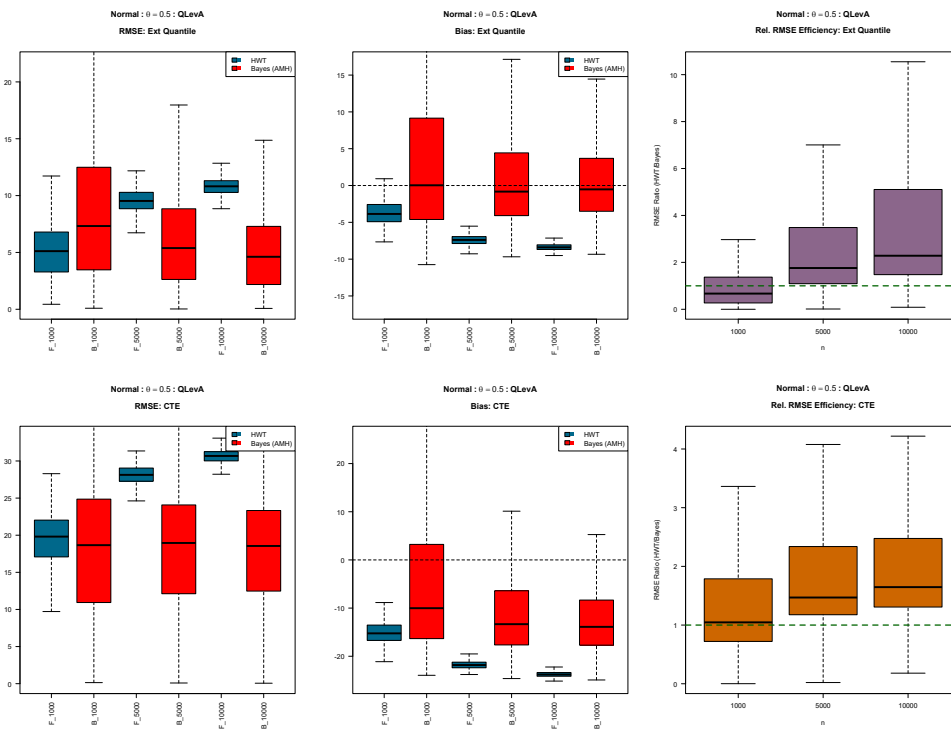


Figure A.2: RMSE and bias box plots on WTC for simulations from DM 1 - Normal, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set A.

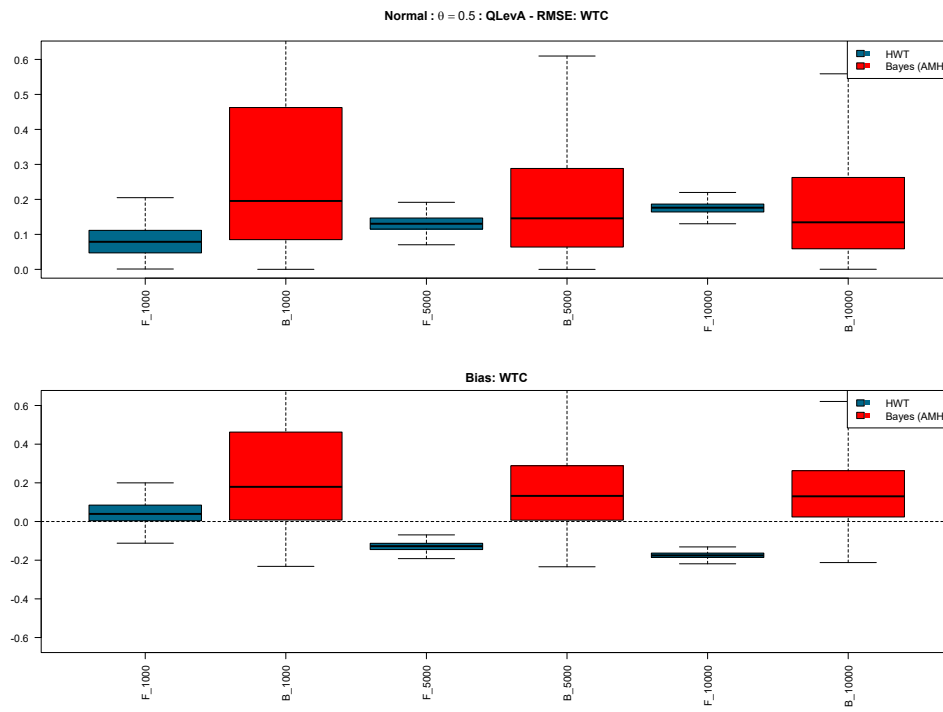


Figure A.3: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Normal, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set B.

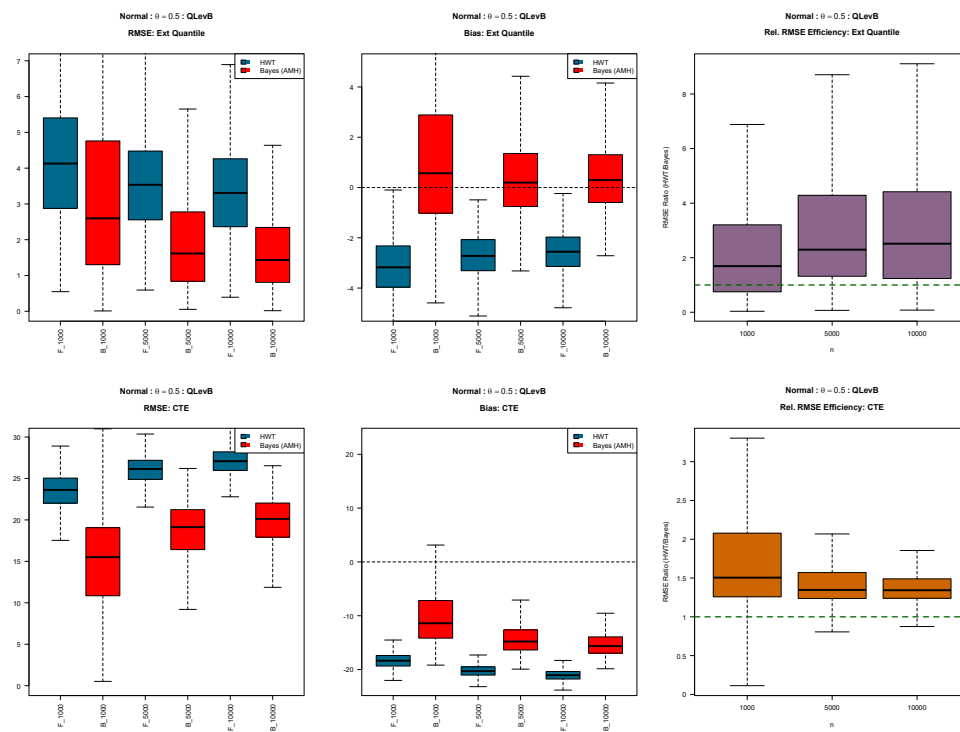
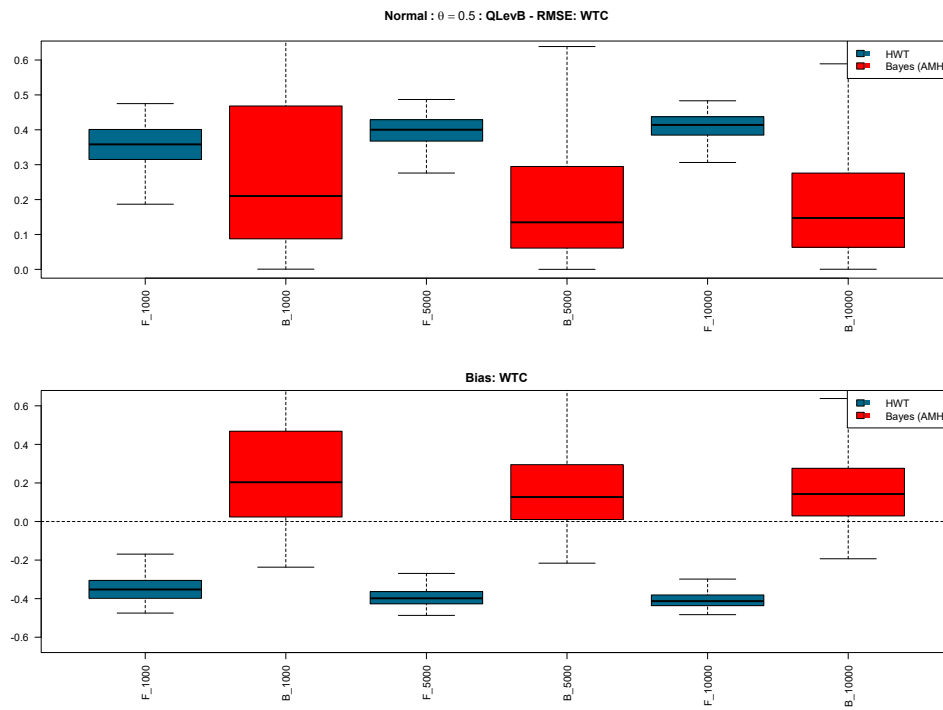


Figure A.4: RMSE and bias box plots on WTC for simulations from DM 1 - Normal, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set B.



## A.2 Data Model 1 - Logistic

Figure A.5: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Logistic, with  $\theta = 1$  using the HWT and AMH methods under quantile set A.

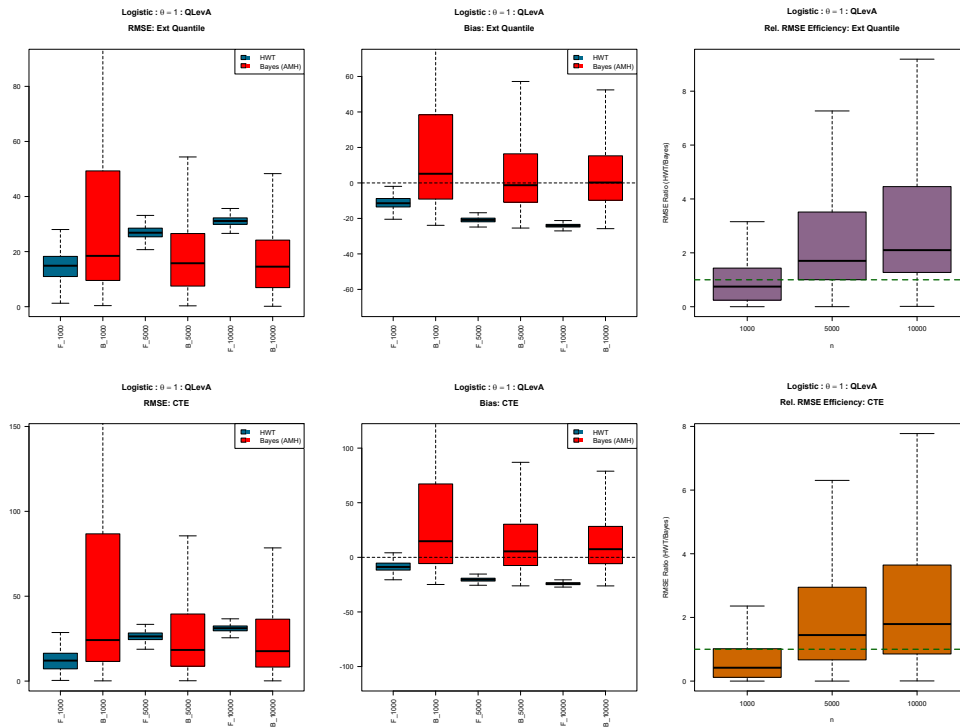


Figure A.6: RMSE and bias box plots on WTC for simulations from DM 1 - Logistic, with  $\theta = 1$  using the HWT and AMH methods under quantile set A.

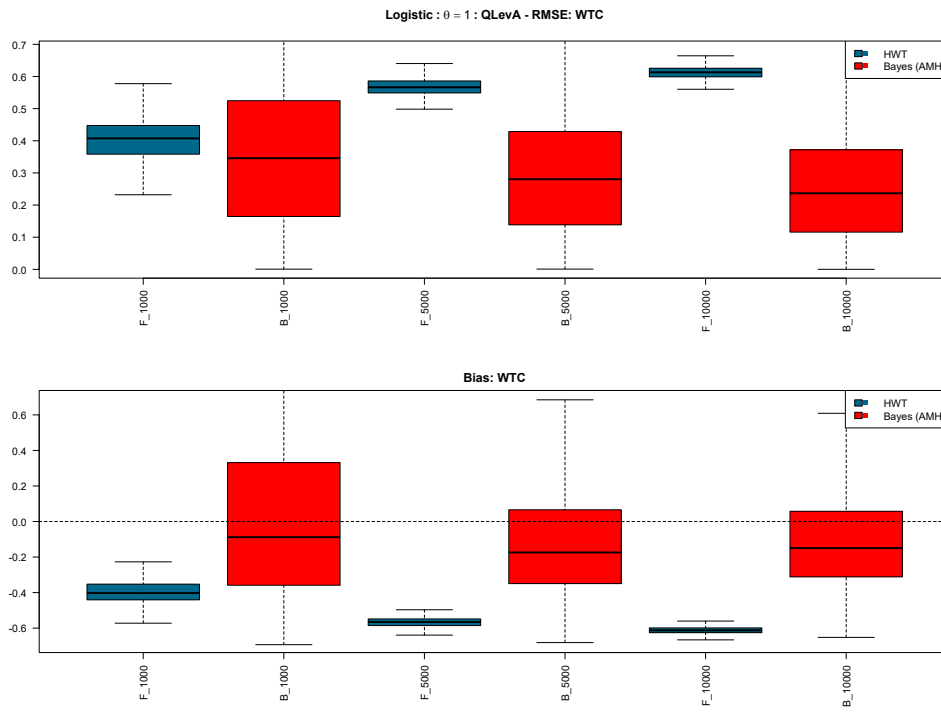


Figure A.7: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Logistic, with  $\theta = 1$  using the HWT and AMH methods under quantile set B.

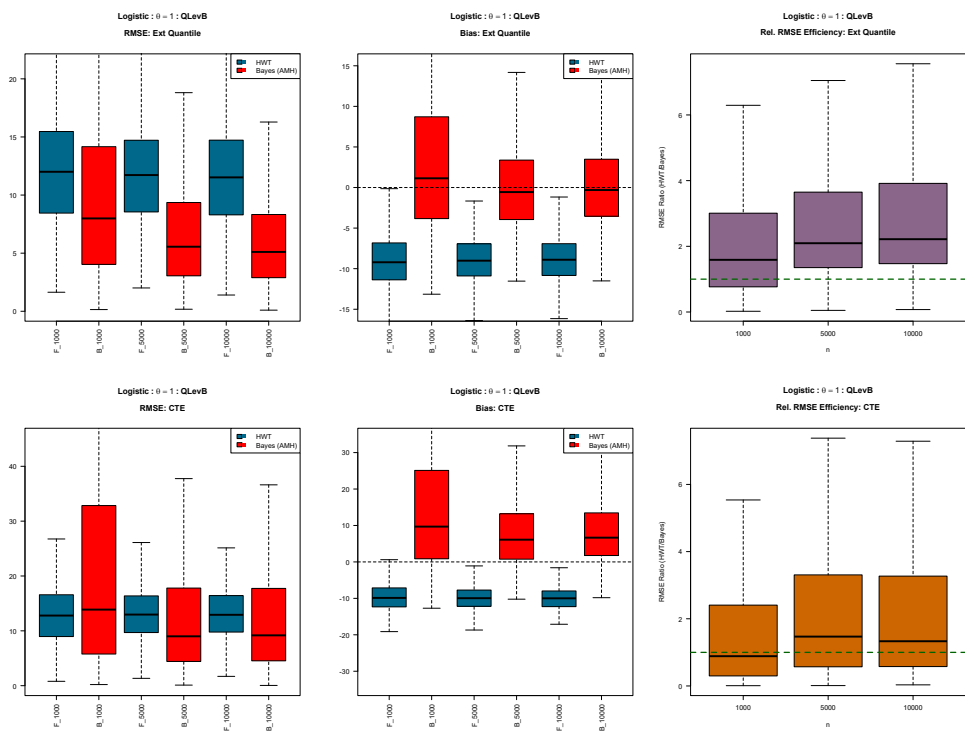
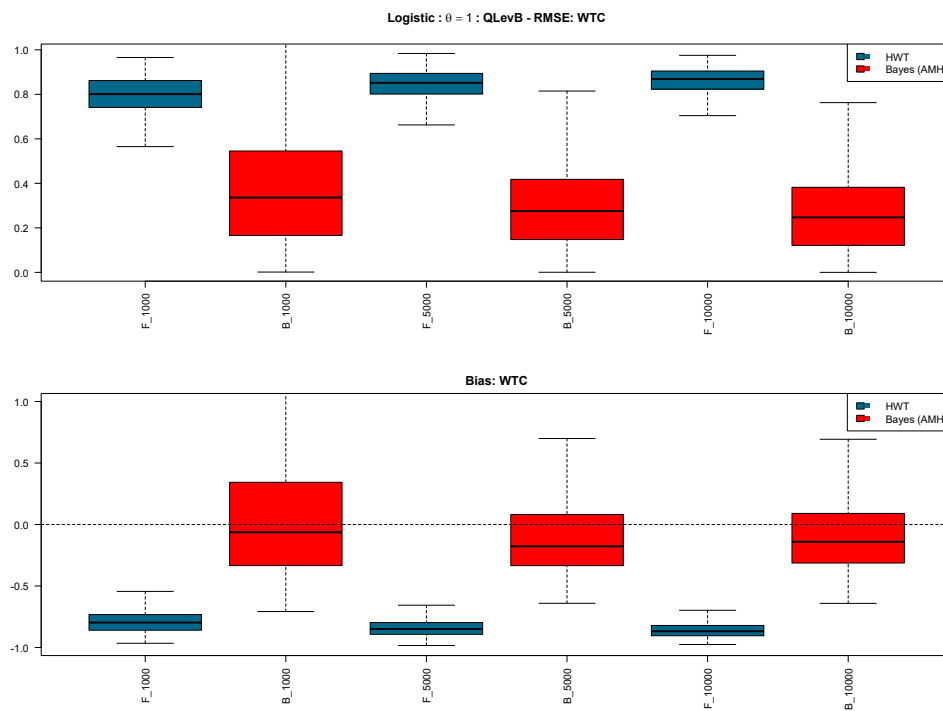


Figure A.8: RMSE and bias box plots on WTC for simulations from DM 1 - Logistic, with  $\theta = 1$  using the HWT and AMH methods under quantile set B.



## A.3 Data Model 1 - Weibull

Figure A.9: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Weibull, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set A.

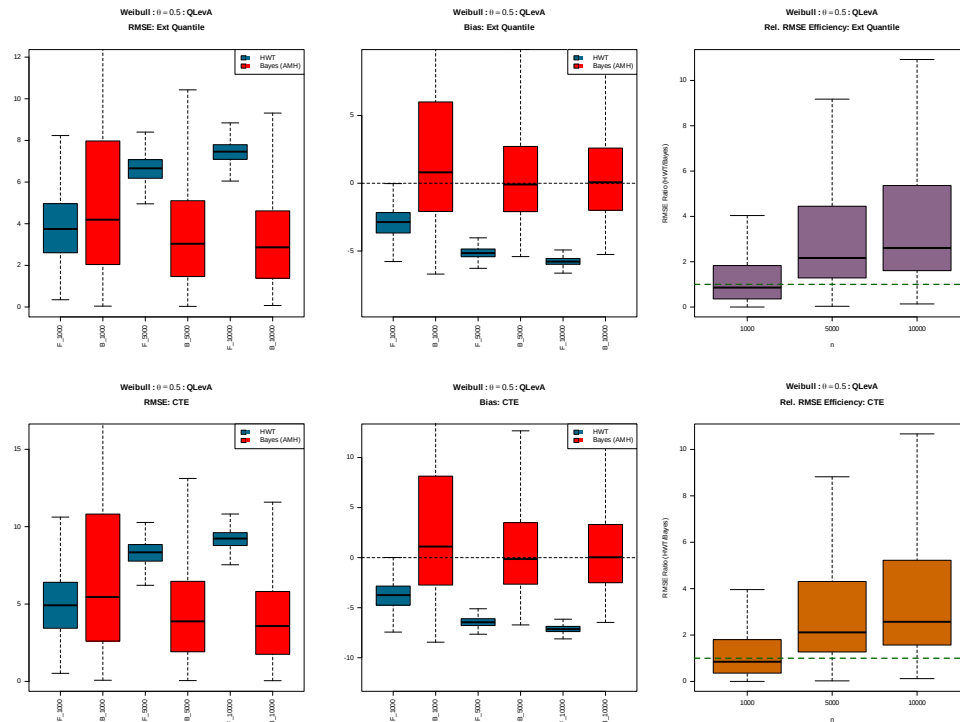


Figure A.10: RMSE and bias box plots on WTC for simulations from DM 1 - Weibull, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set A.

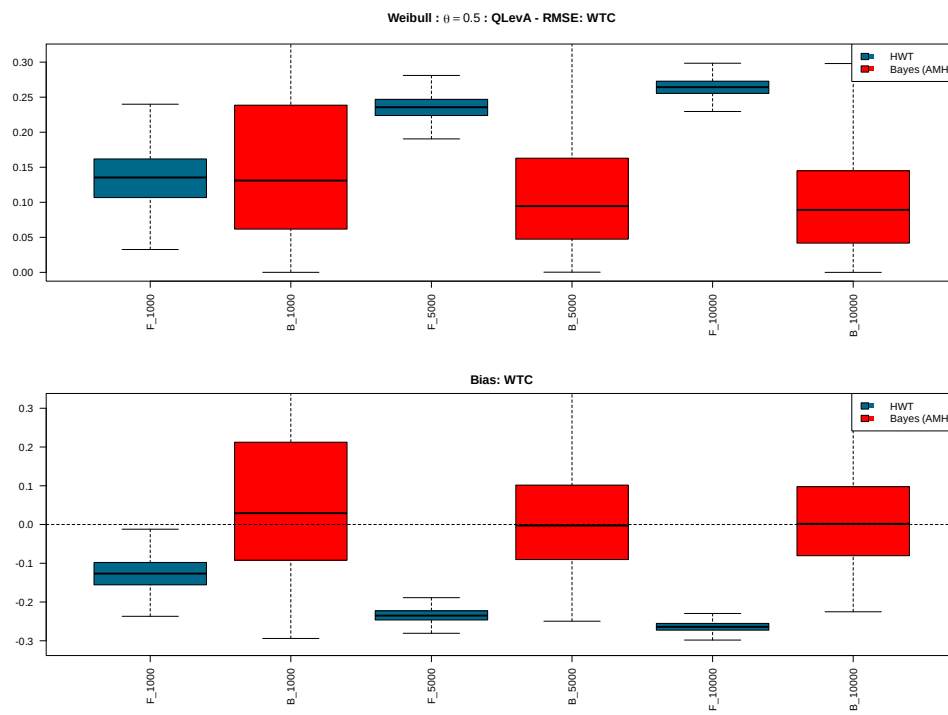


Figure A.11: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Weibull, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set B.

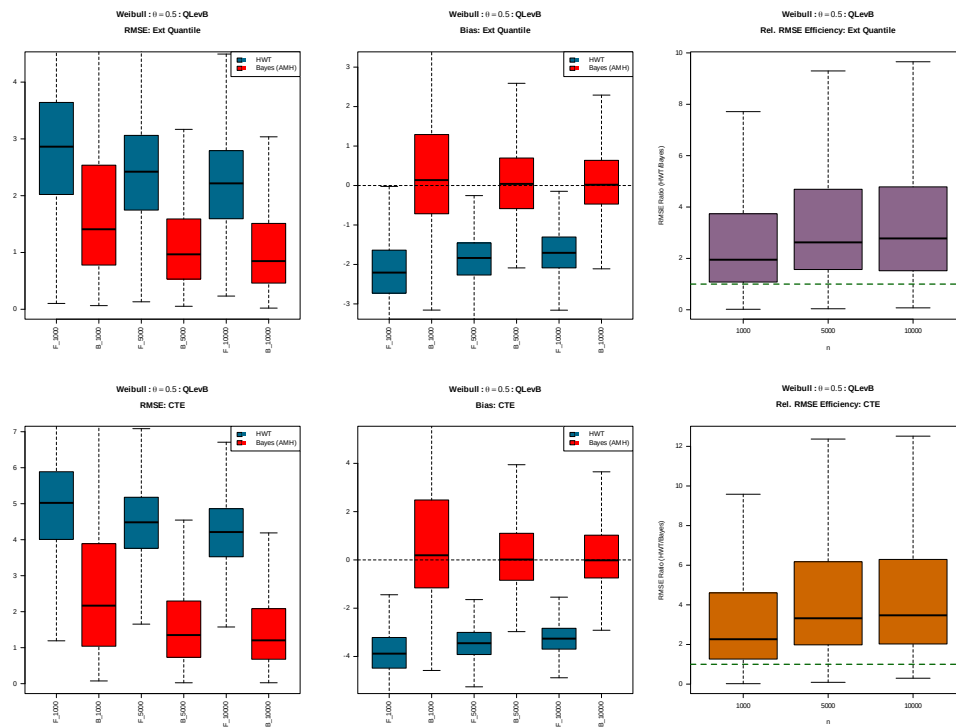


Figure A.12: RMSE and bias box plots on WTC for simulations from DM 1 - Weibull, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set B.

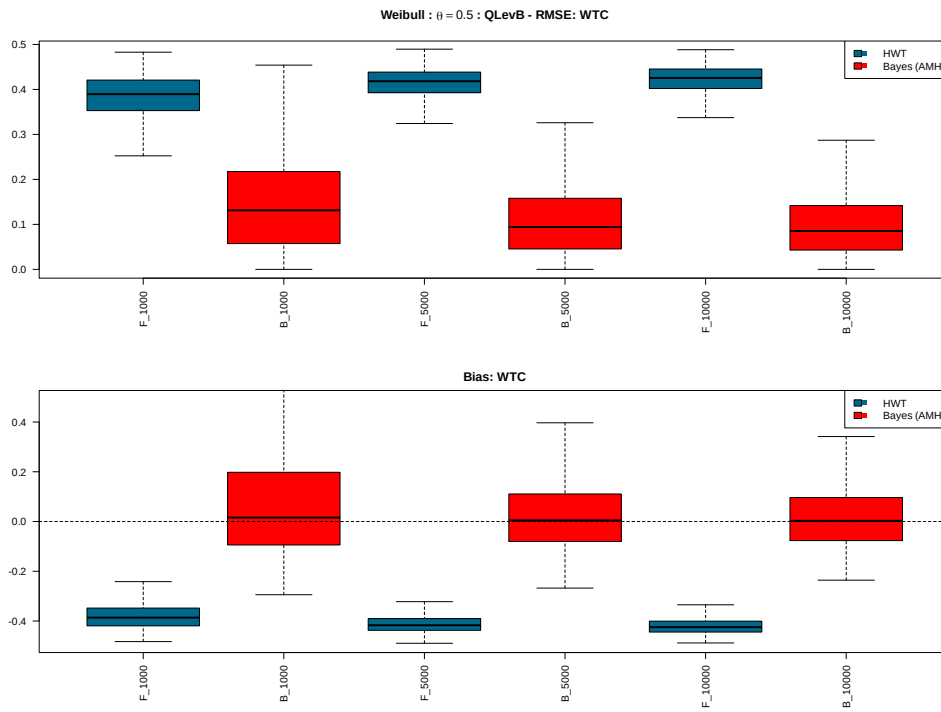


Figure A.13: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Weibull, with  $\theta = 1$  using the HWT and AMH methods under quantile set A.

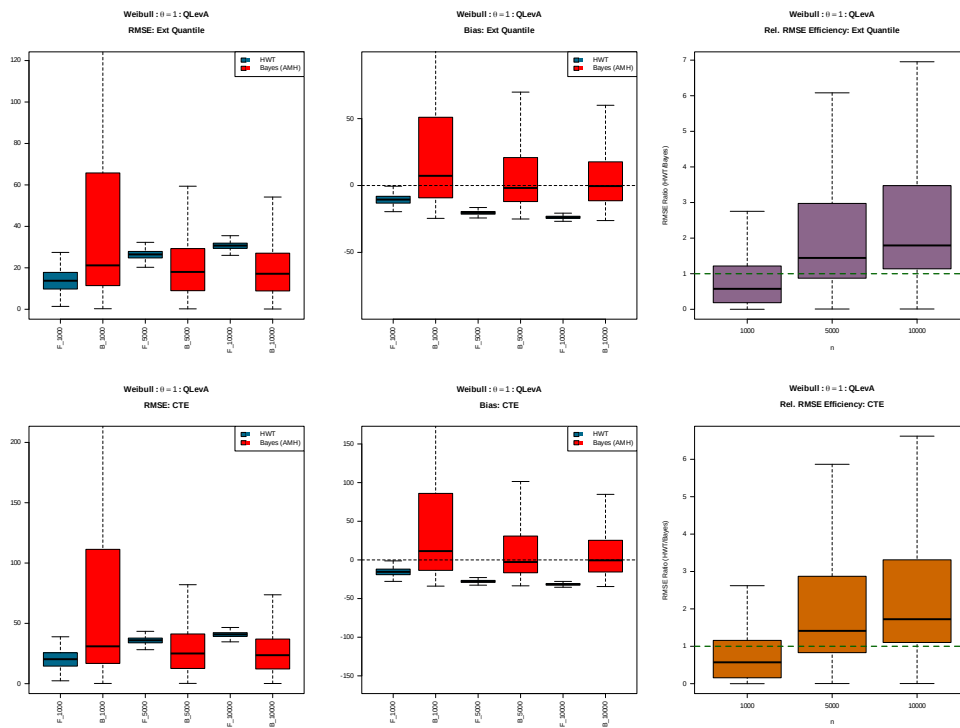


Figure A.14: RMSE and bias box plots on WTC for simulations from DM 1 - Weibull, with  $\theta = 1$  using the HWT and AMH methods under quantile set A.

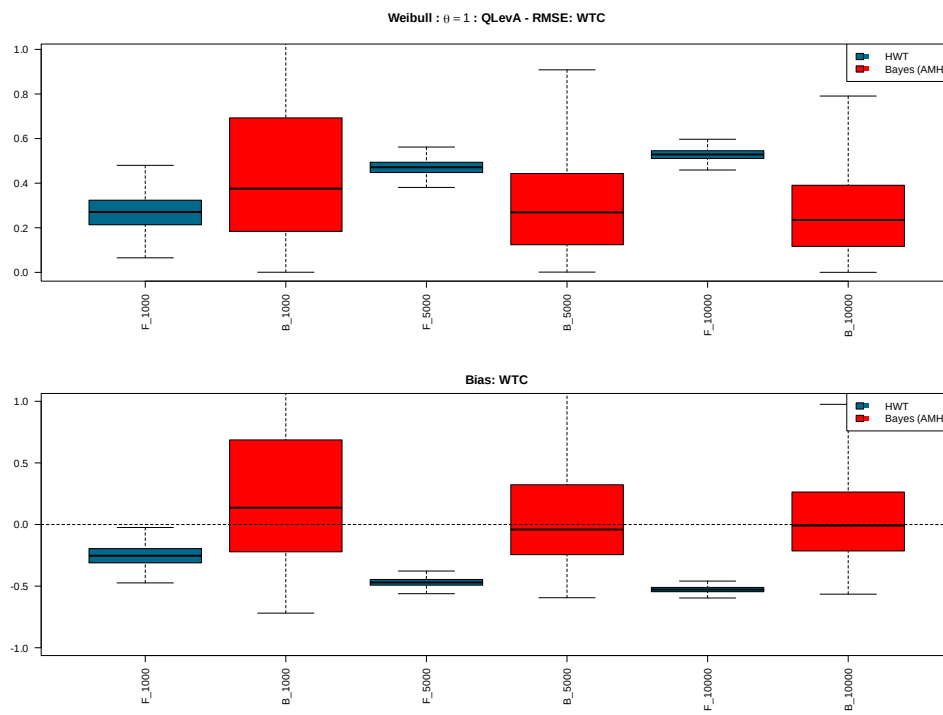


Figure A.15: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Weibull, with  $\theta = 1$  using the HWT and AMH methods under quantile set B.

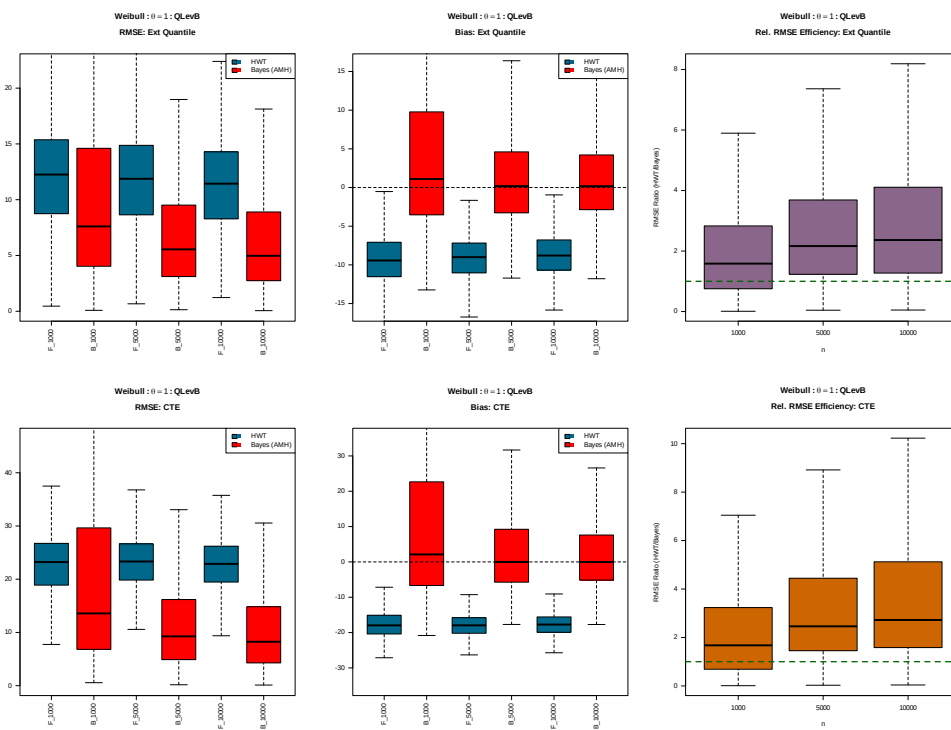


Figure A.16: RMSE and bias box plots on WTC for simulations from DM 1 - Weibull, with  $\theta = 1$  using the HWT and AMH methods under quantile set B.

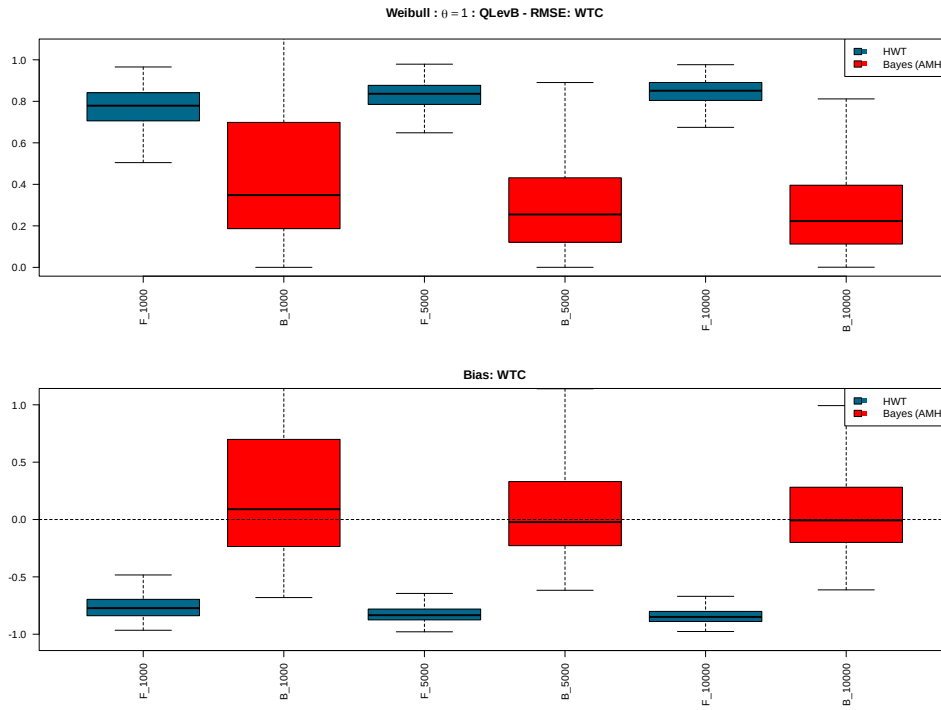


Figure A.17: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Weibull, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set A.

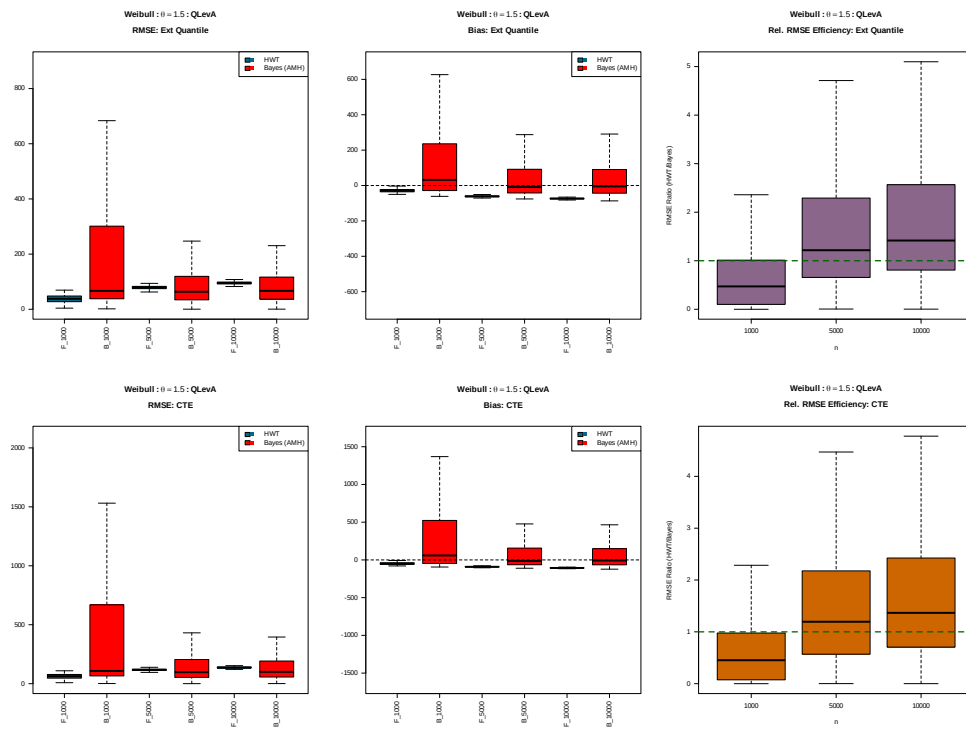


Figure A.18: RMSE and bias box plots on WTC for simulations from DM 1 - Weibull, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set A.

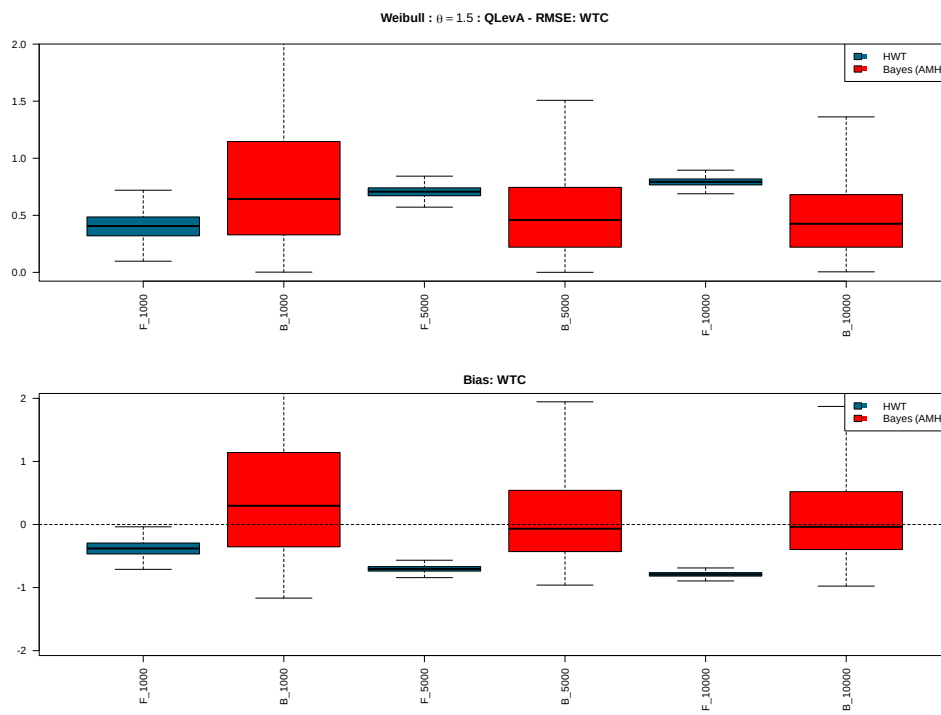


Figure A.19: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 1 - Weibull, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set B.

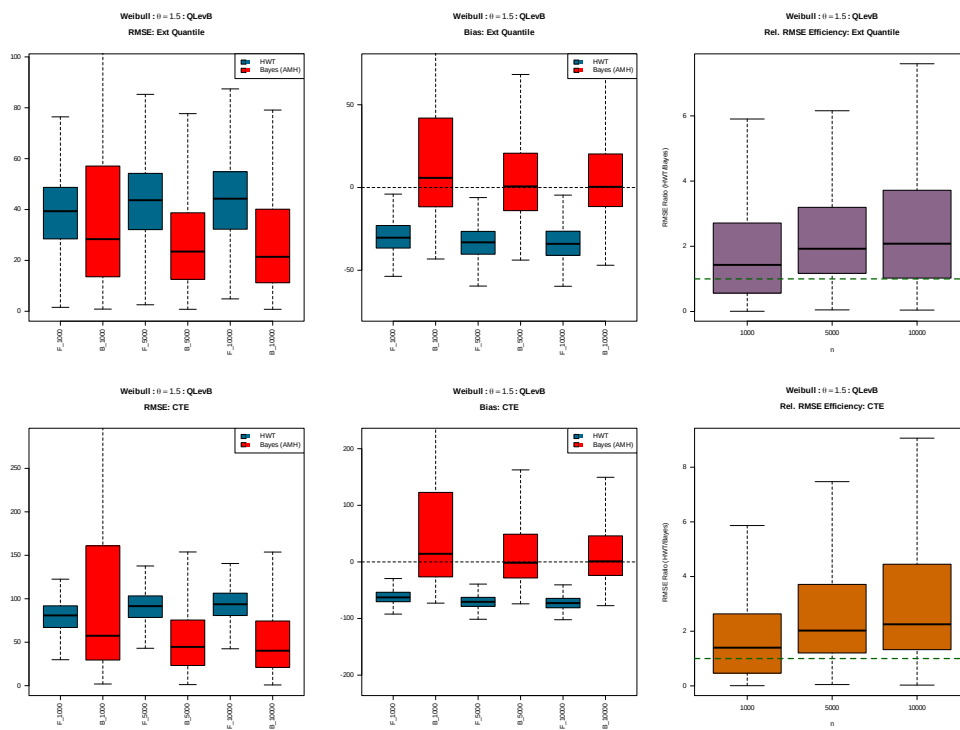
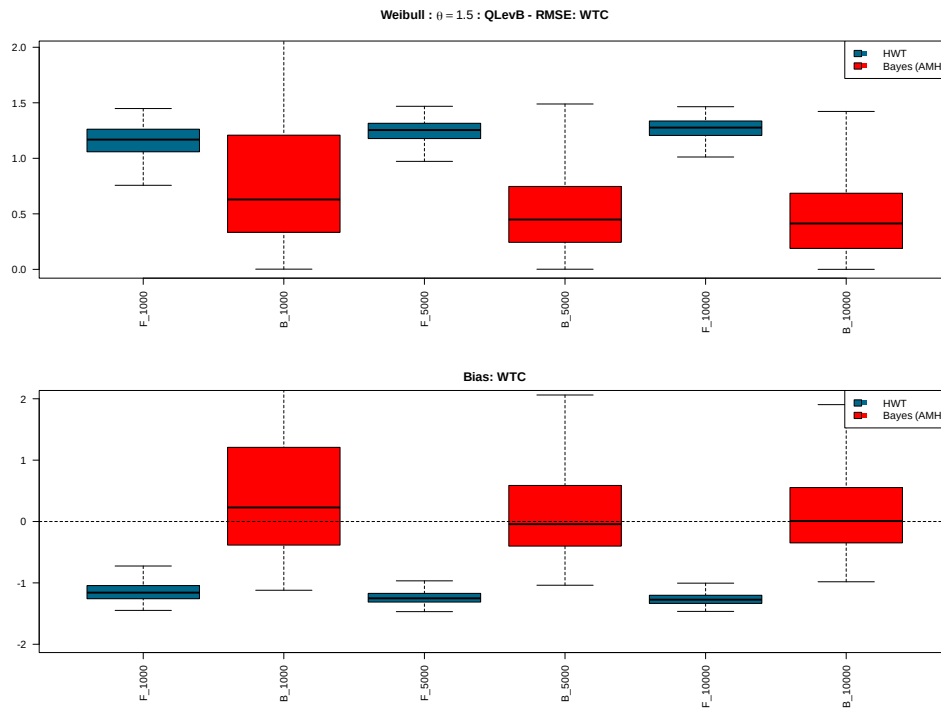


Figure A.20: RMSE and bias box plots on WTC for simulations from DM 1 - Weibull, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set B.



## APPENDIX

### B

# CHAPTER 4 SUPPLEMENTARY MATERIAL - DATA MODEL 2

## **B.1 Data Model 2 - Beta**

Figure B.1: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Beta, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set B.

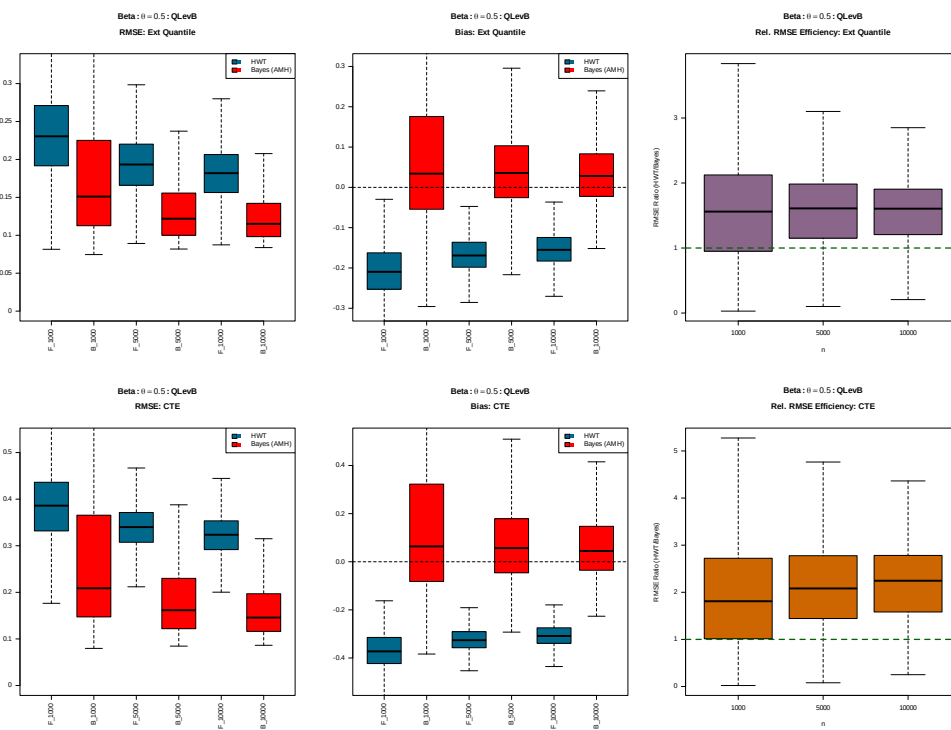


Figure B.2: RMSE and bias box plots on WTC for simulations from DM 2 - Beta, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set B.

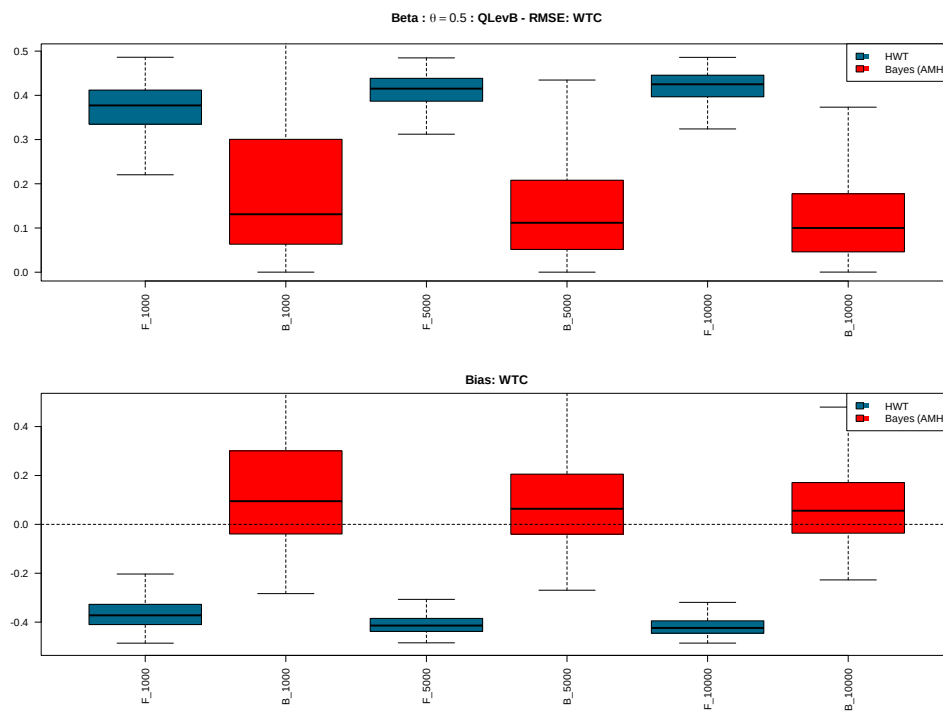


Figure B.3: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Beta, with  $\theta = 1$  using the HWT and AMH methods under quantile set A.

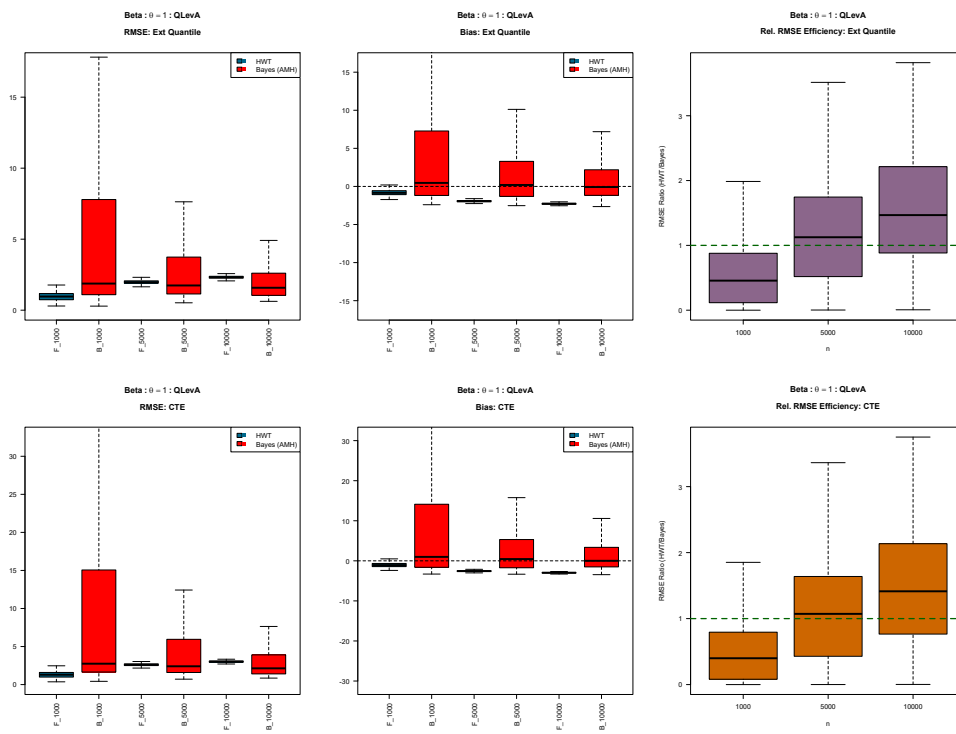


Figure B.4: RMSE and bias box plots on WTC for simulations from DM 2 - Beta, with  $\theta = 1$  using the HWT and AMH methods under quantile set A.

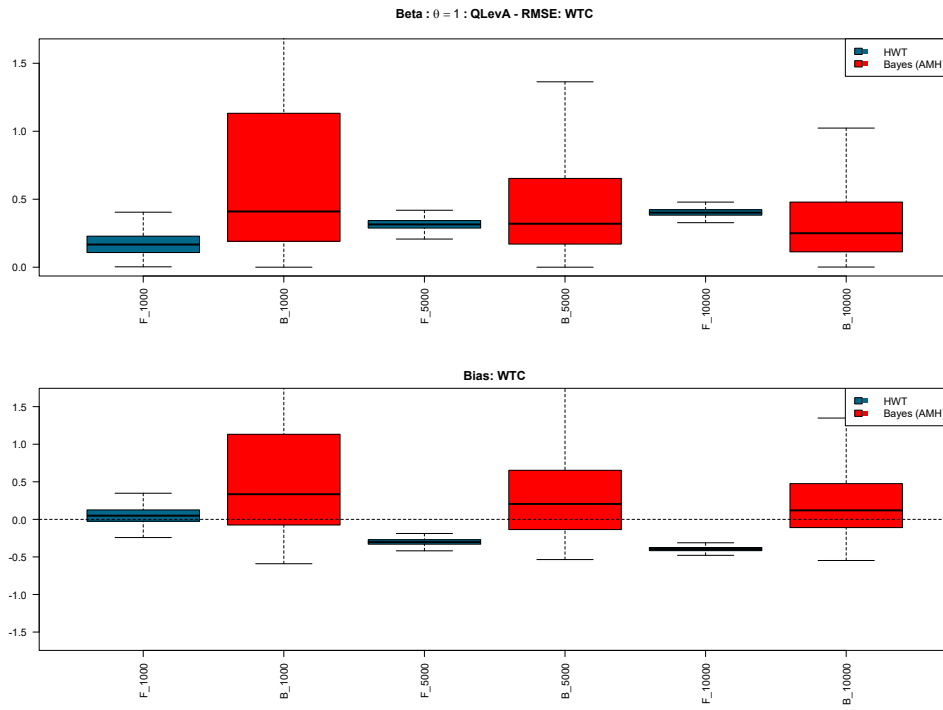


Figure B.5: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Beta, with  $\theta = 1$  using the HWT and AMH methods under quantile set B.

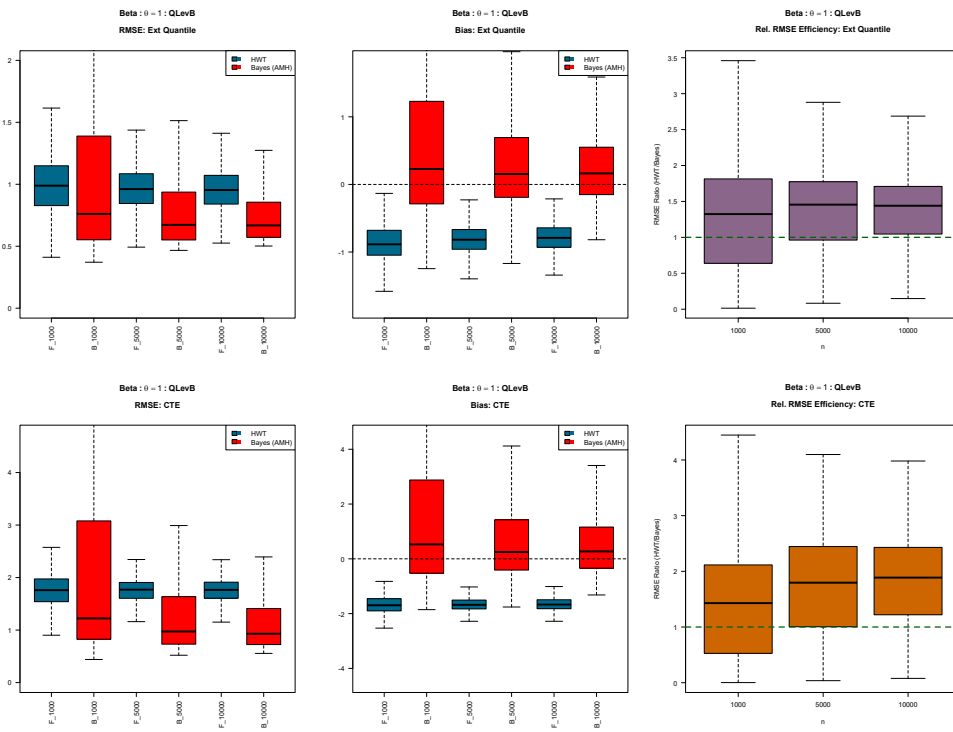


Figure B.6: RMSE and bias box plots on WTC for simulations from DM 2 - Beta, with  $\theta = 1$  using the HWT and AMH methods under quantile set B.

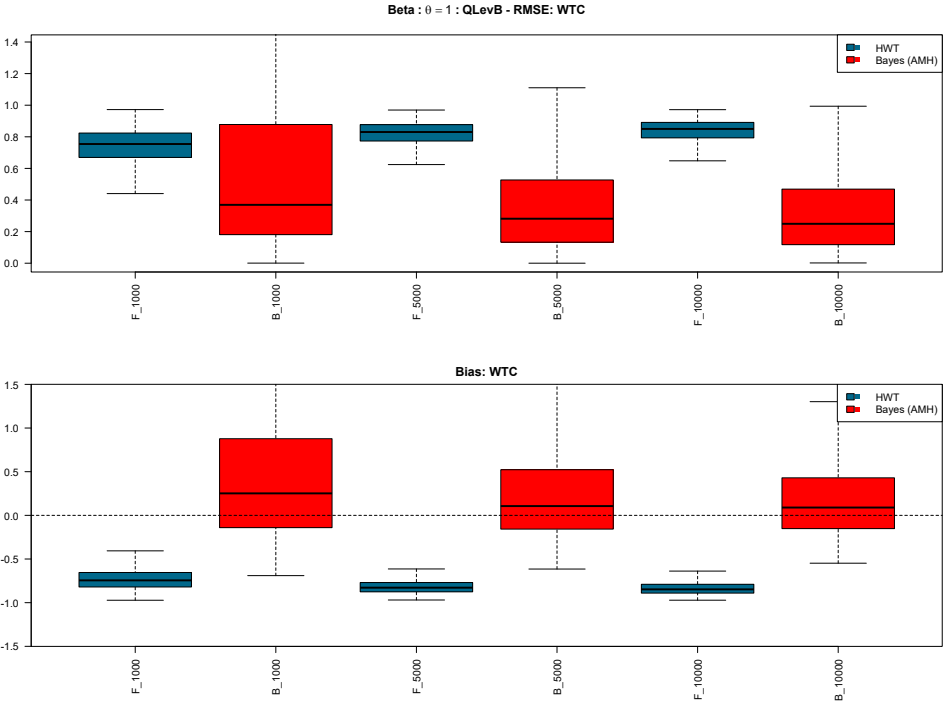


Figure B.7: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Beta, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set A.

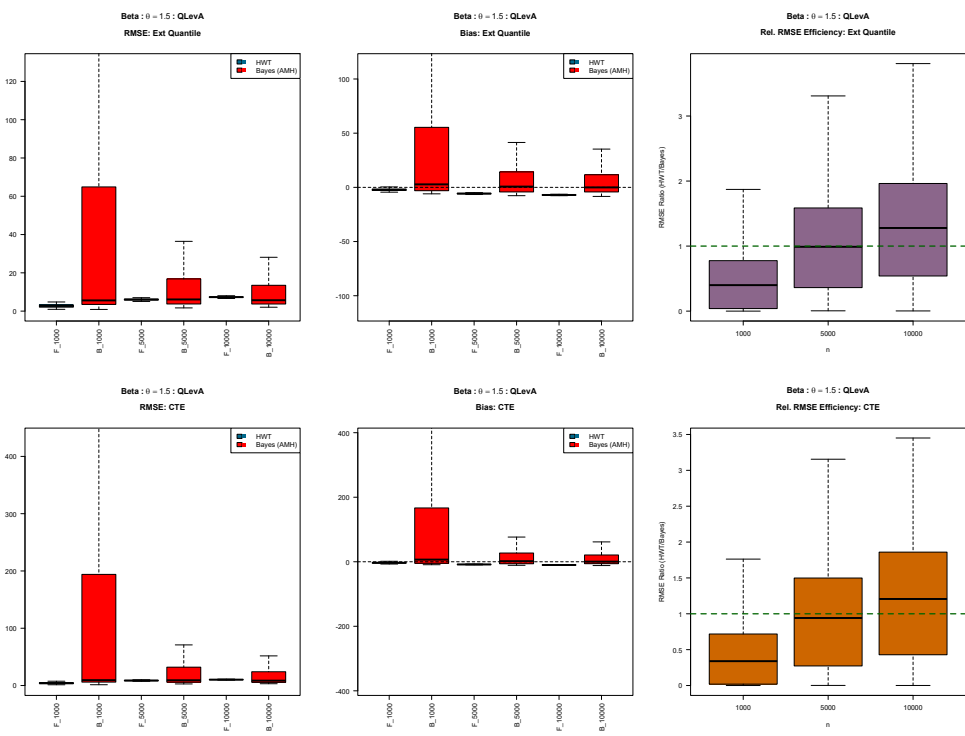


Figure B.8: RMSE and bias box plots on WTC for simulations from DM 2 - Beta, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set A.

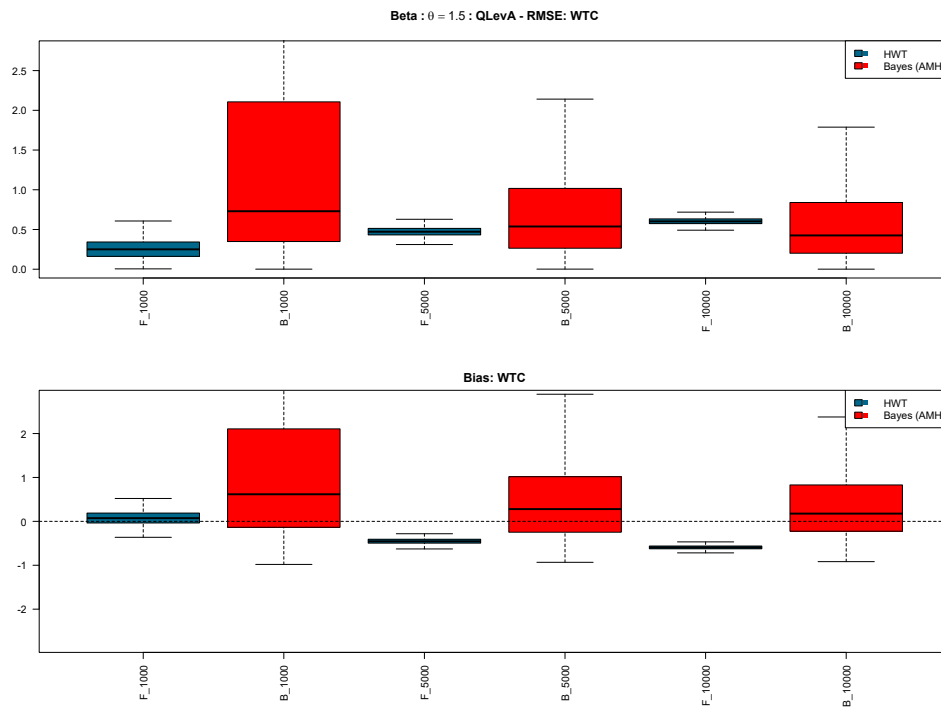


Figure B.9: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Beta, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set B.

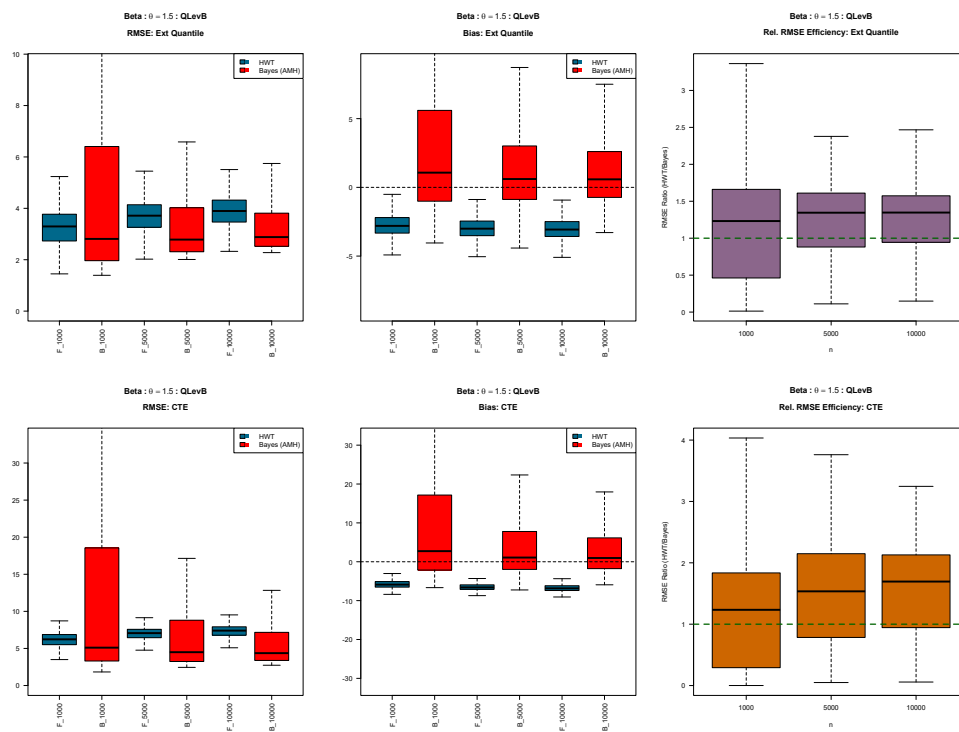
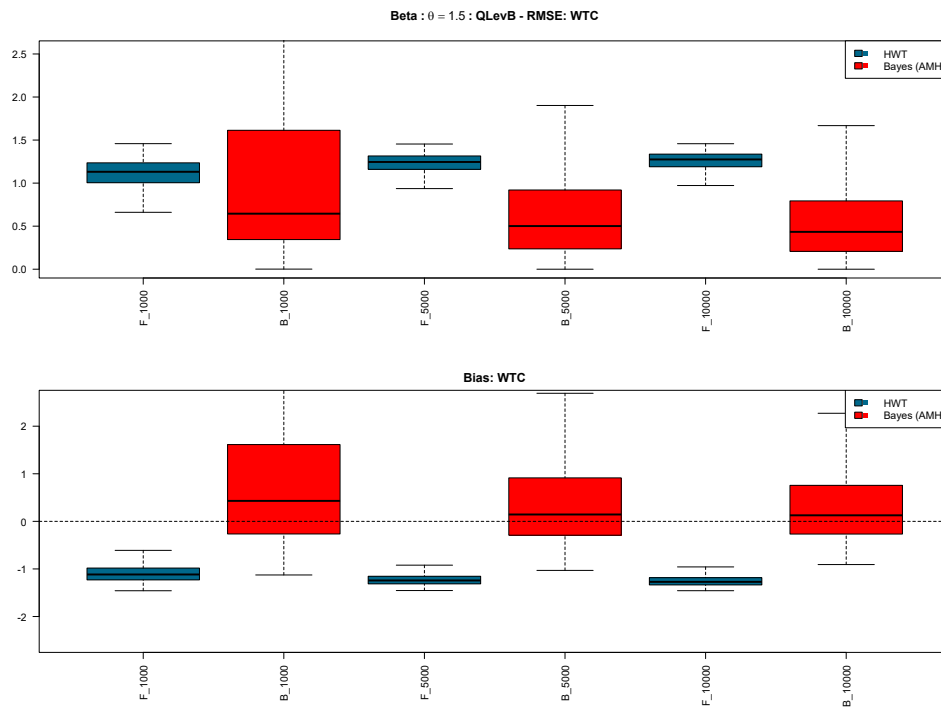


Figure B.10: RMSE and bias box plots on WTC for simulations from DM 2 - Beta, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set B.



## B.2 Data Model 2 - Trig

Figure B.11: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Trig, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set A.

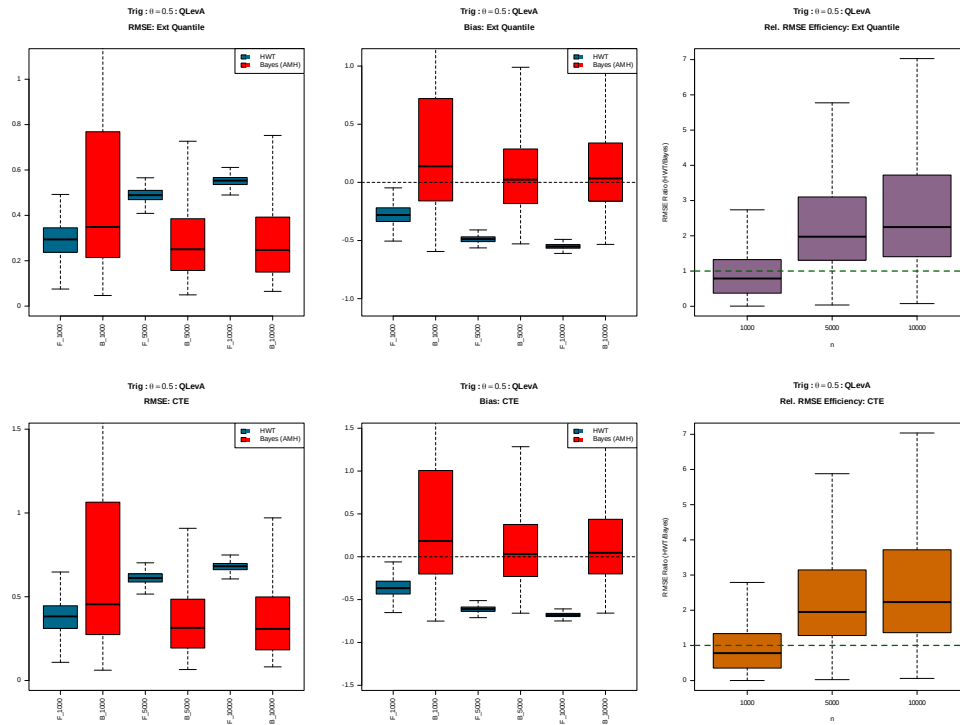


Figure B.12: RMSE and bias box plots on WTC for simulations from DM 2 - Trig, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set A.

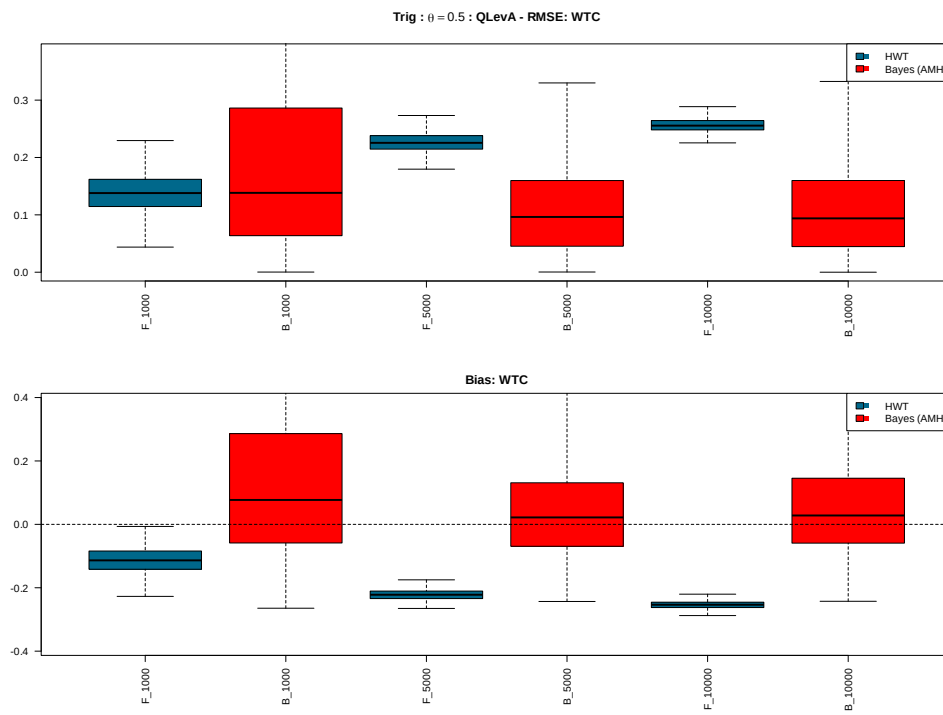


Figure B.13: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Trig, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set B.

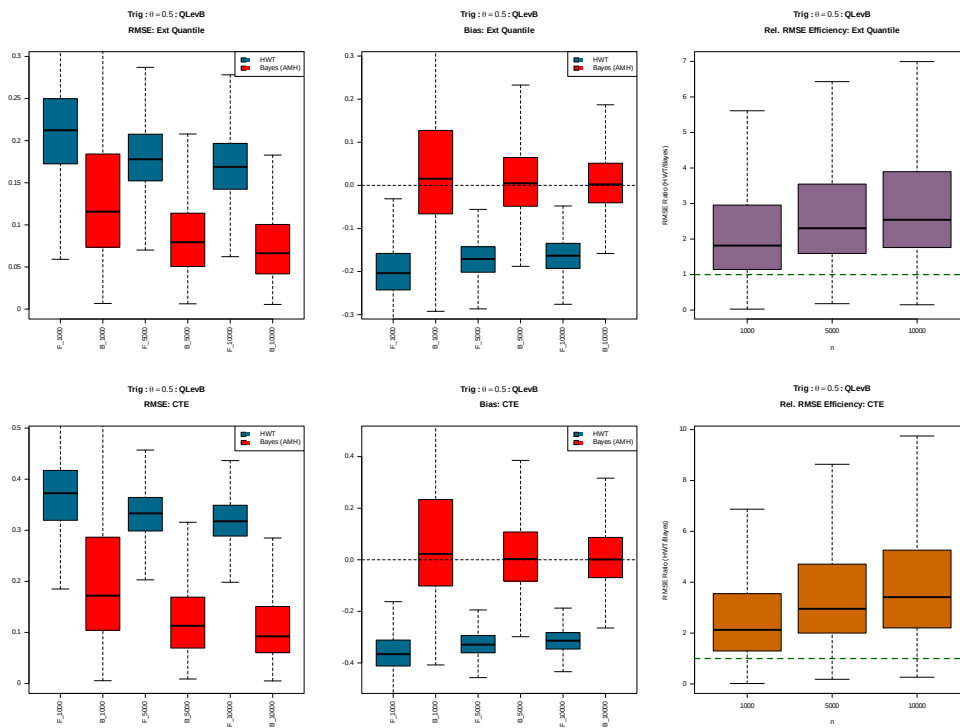


Figure B.14: RMSE and bias box plots on WTC for simulations from DM 2 - Trig, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set B.

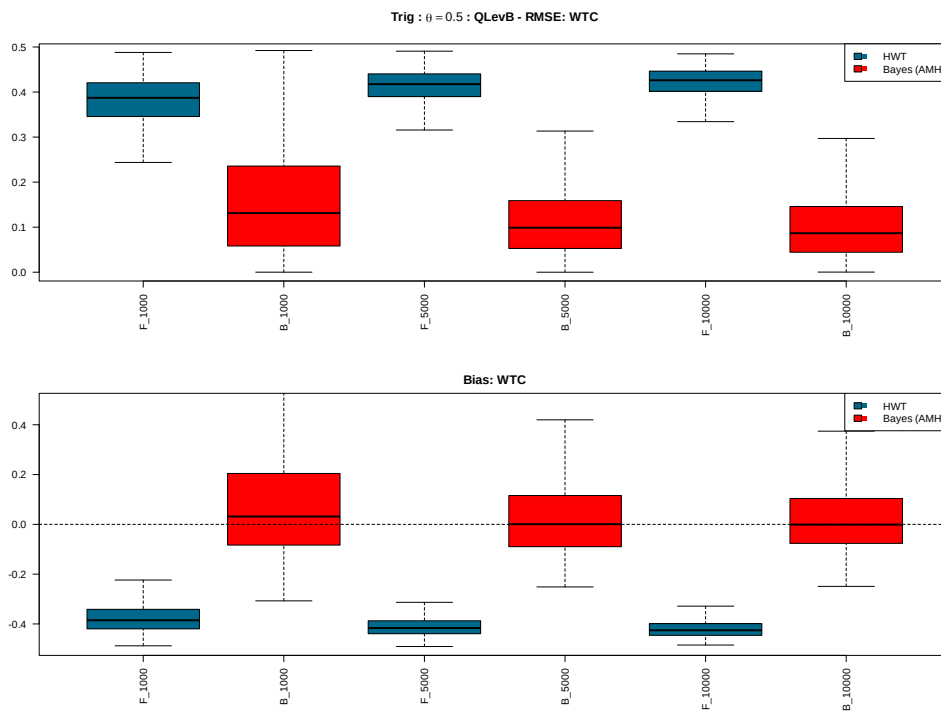


Figure B.15: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Trig, with  $\theta = 1$  using the HWT and AMH methods under quantile set A.

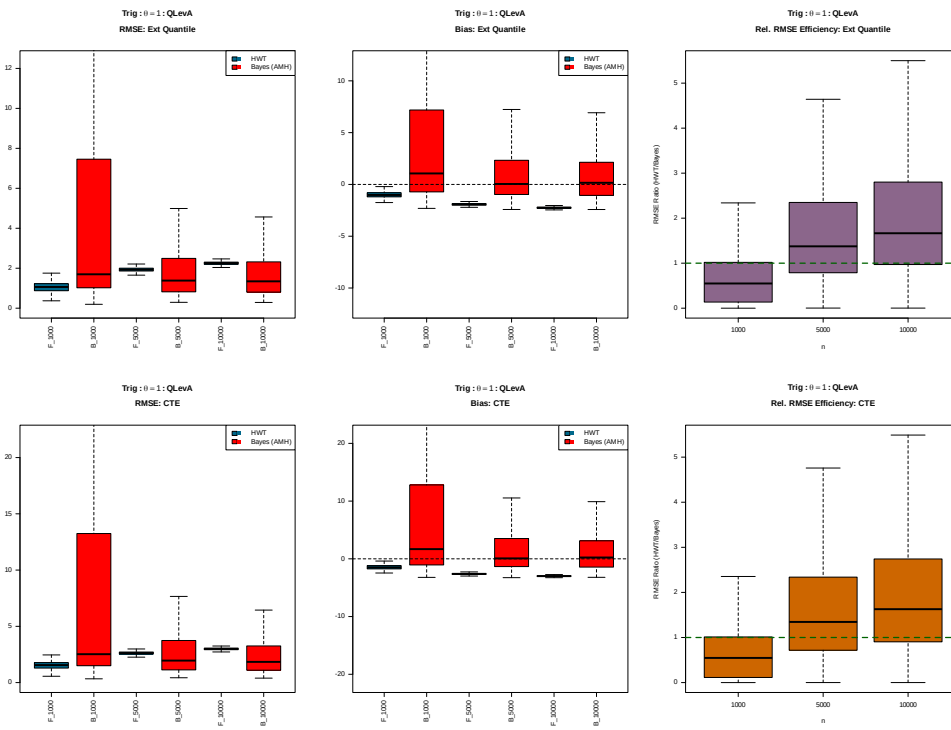


Figure B.16: RMSE and bias box plots on WTC for simulations from DM 2 - Trig, with  $\theta = 1$  using the HWT and AMH methods under quantile set A.

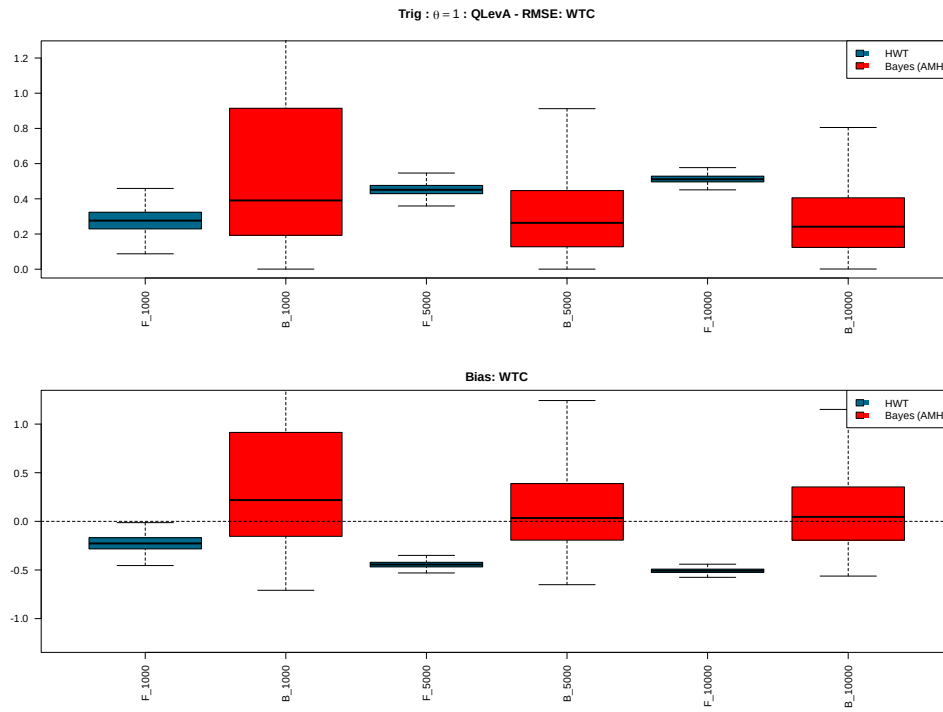


Figure B.17: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Trig, with  $\theta = 1$  using the HWT and AMH methods under quantile set B.

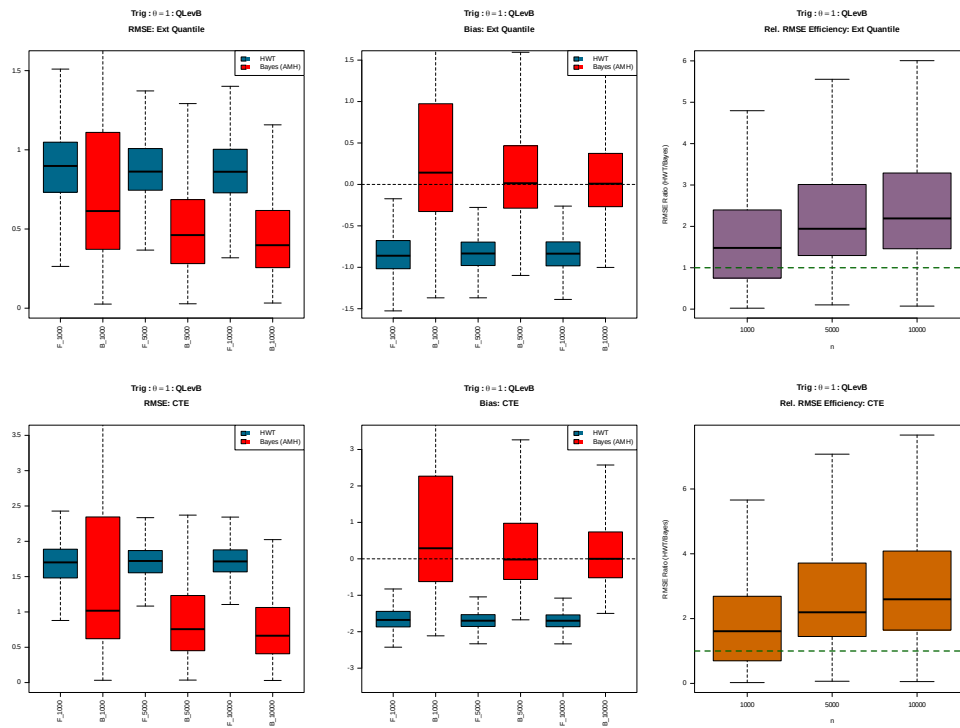


Figure B.18: RMSE and bias box plots on WTC for simulations from DM 2 - Trig, with  $\theta = 1$  using the HWT and AMH methods under quantile set B.

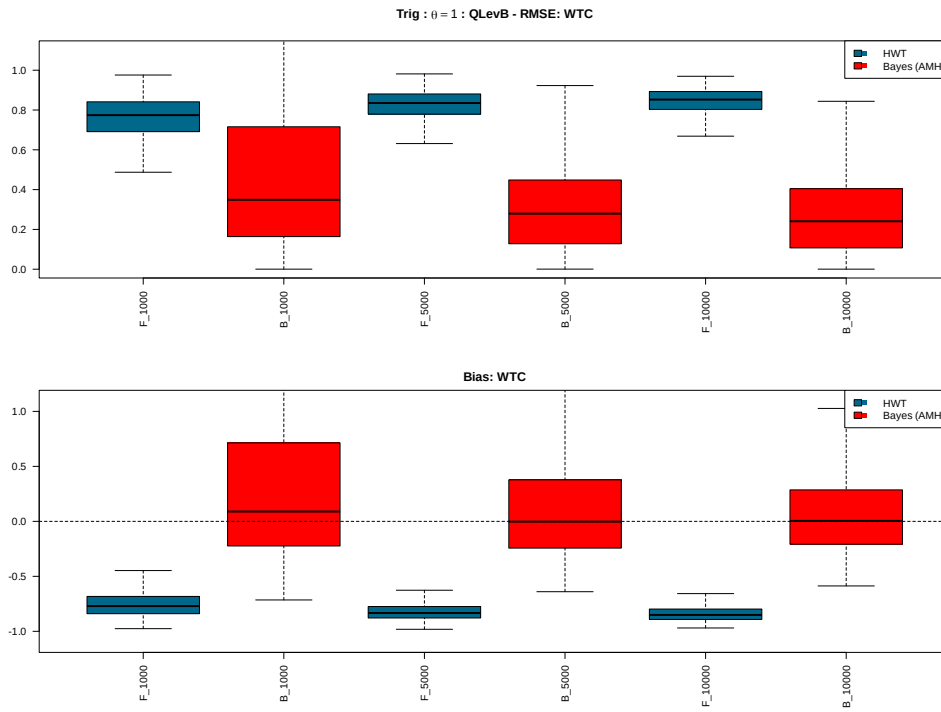


Figure B.19: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Trig, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set A.

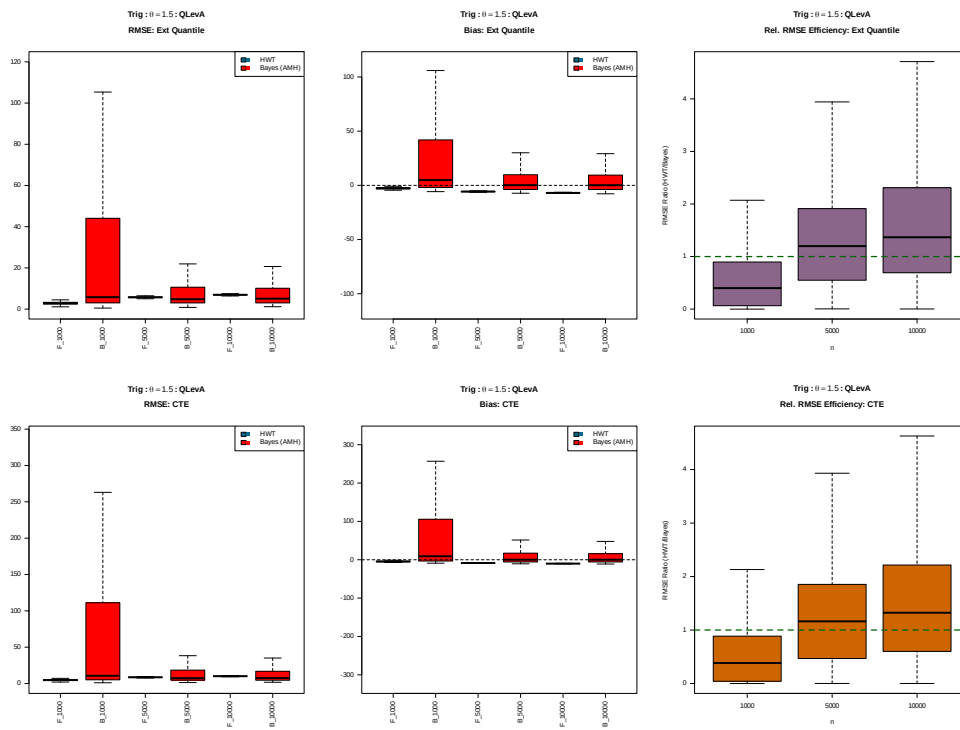


Figure B.20: RMSE and bias box plots on WTC for simulations from DM 2 - Trig, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set A.

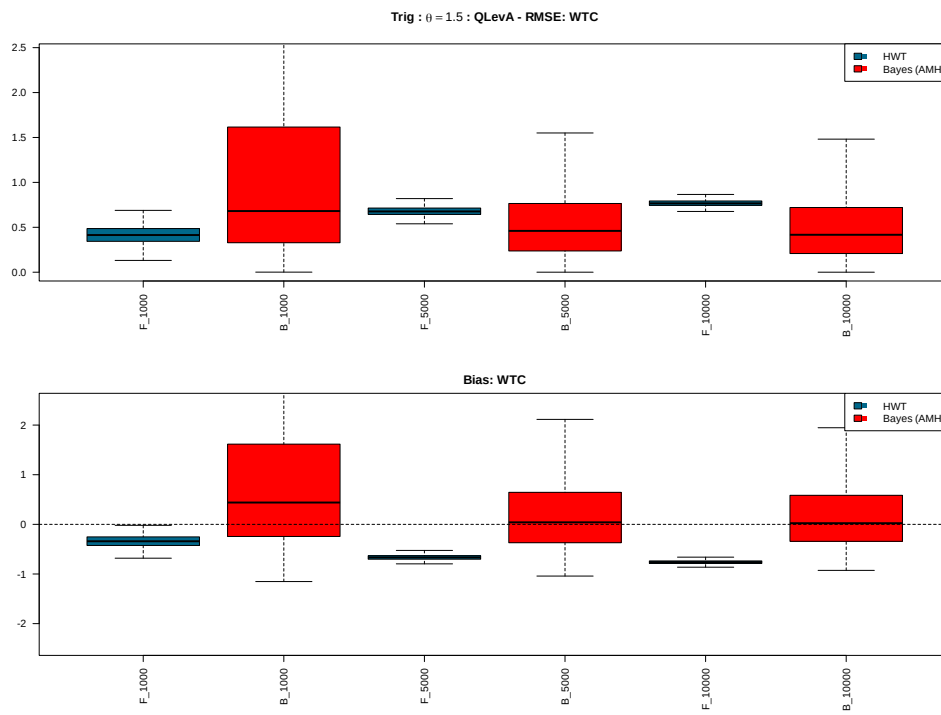


Figure B.21: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Trig, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set B.

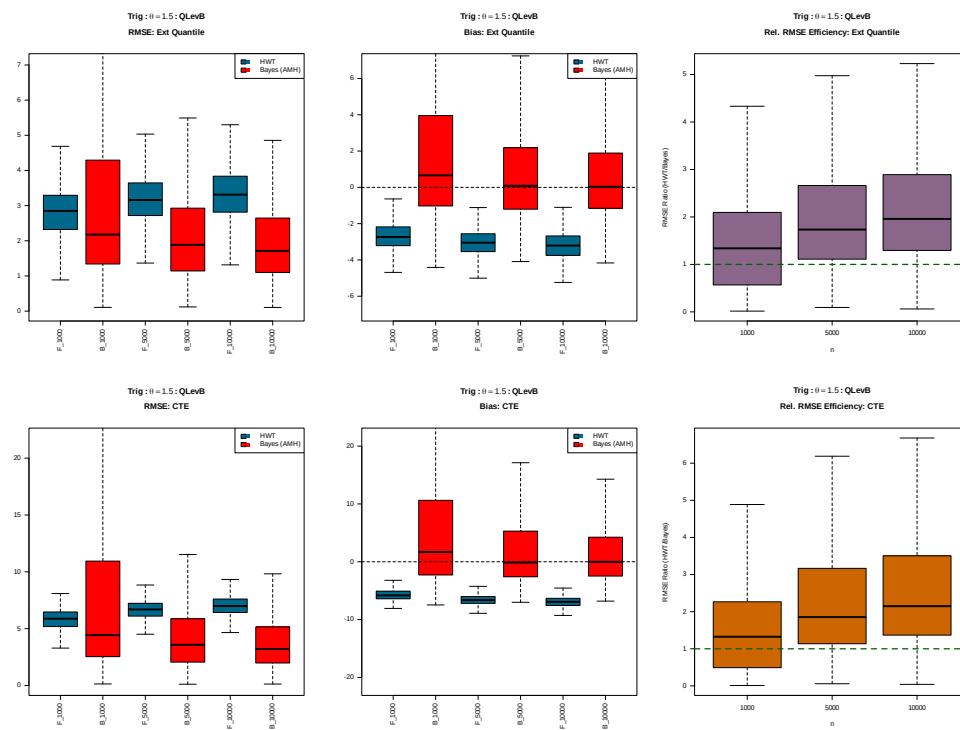
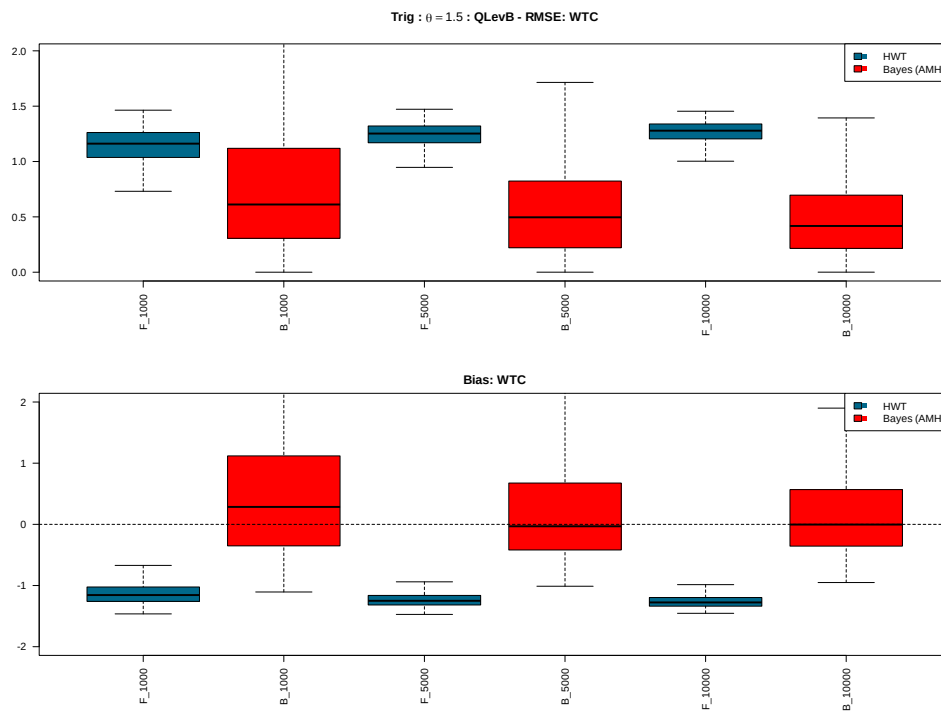


Figure B.22: RMSE and bias box plots on WTC for simulations from DM 2 - Trig, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set B.



## B.3 Data Model 2 - Poly

Figure B.23: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Poly, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set A.

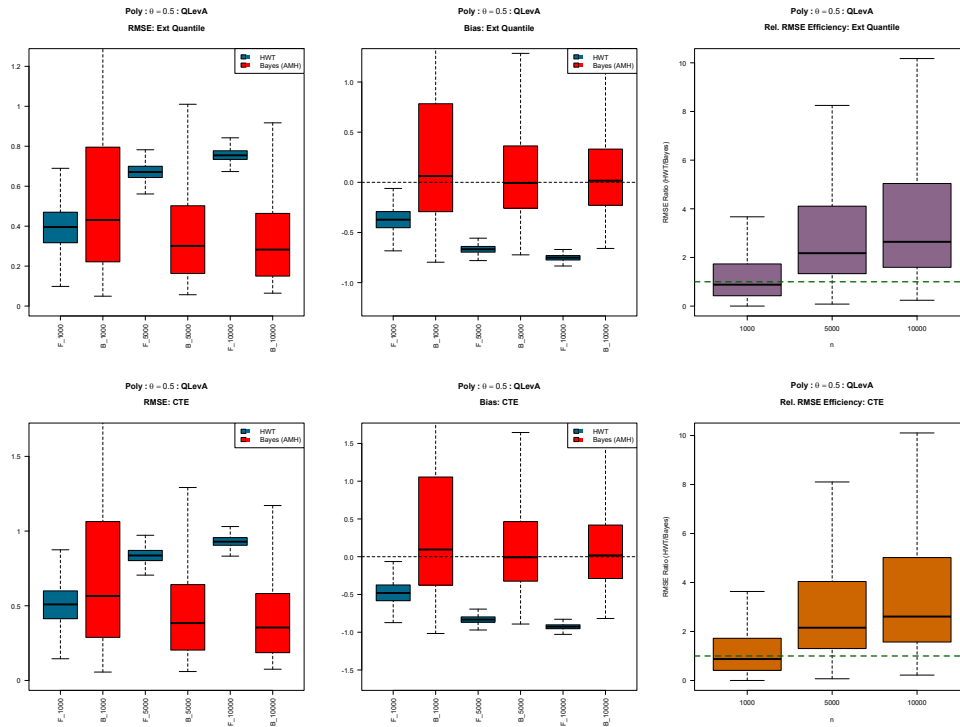


Figure B.24: RMSE and bias box plots on WTC for simulations from DM 2 - Poly, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set A.

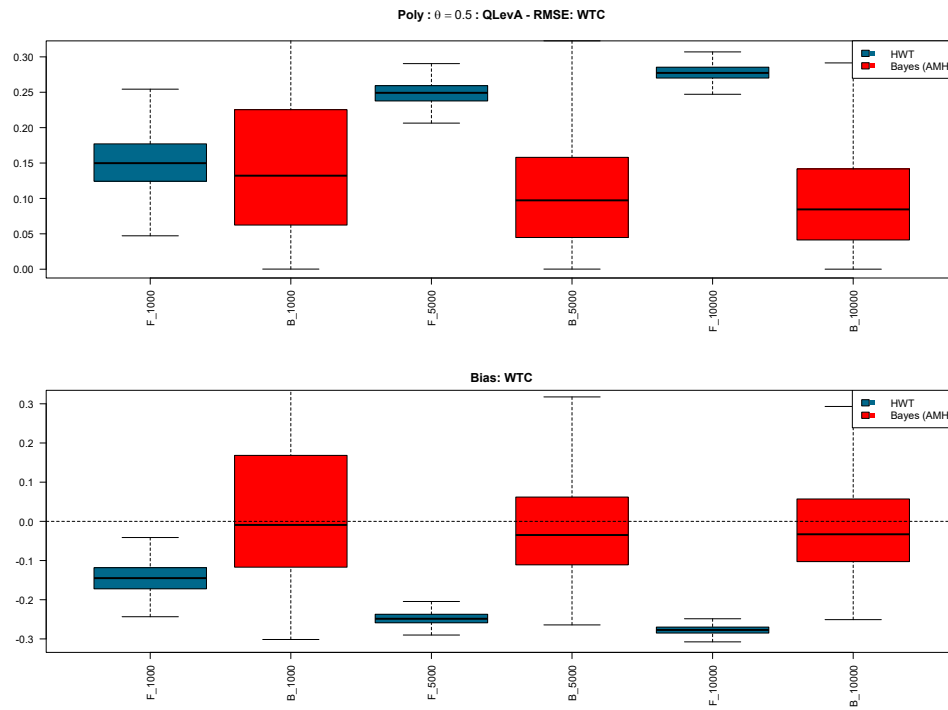


Figure B.25: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Poly, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set B.

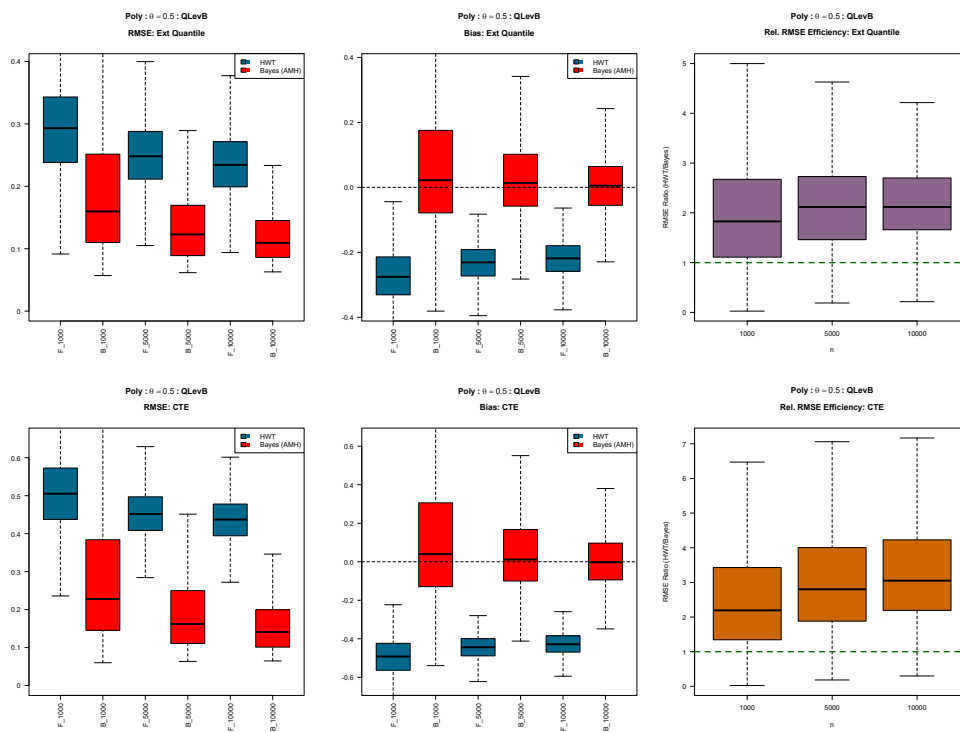


Figure B.26: RMSE and bias box plots on WTC for simulations from DM 2 - Poly, with  $\theta = 0.5$  using the HWT and AMH methods under quantile set B.

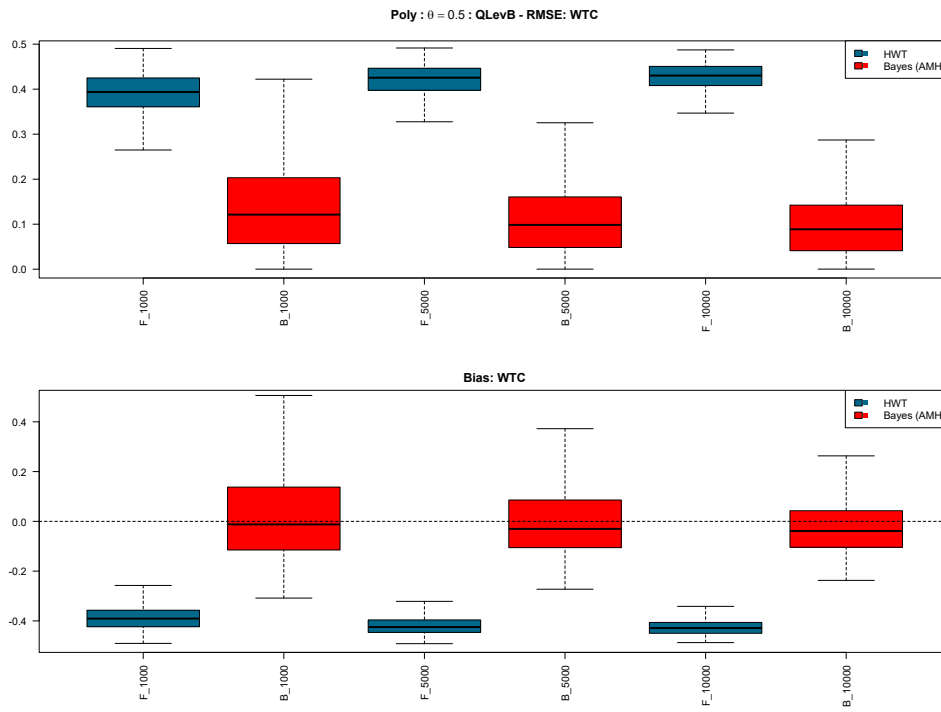


Figure B.27: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Poly, with  $\theta = 1$  using the HWT and AMH methods under quantile set A.

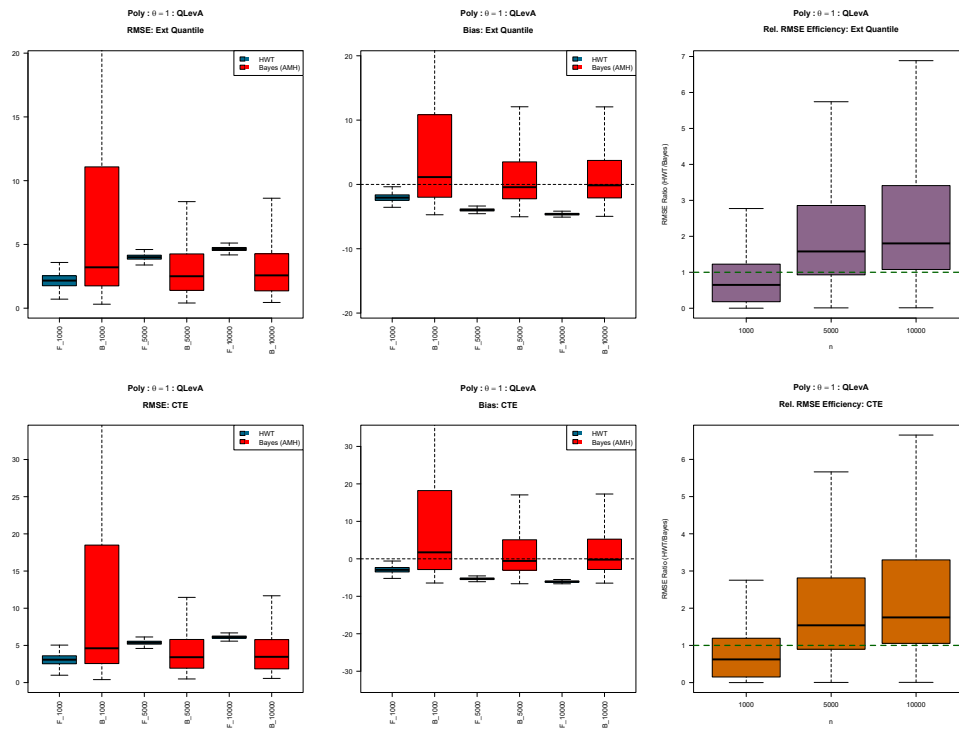


Figure B.28: RMSE and bias box plots on WTC for simulations from DM 2 - Poly, with  $\theta = 1$  using the HWT and AMH methods under quantile set A.

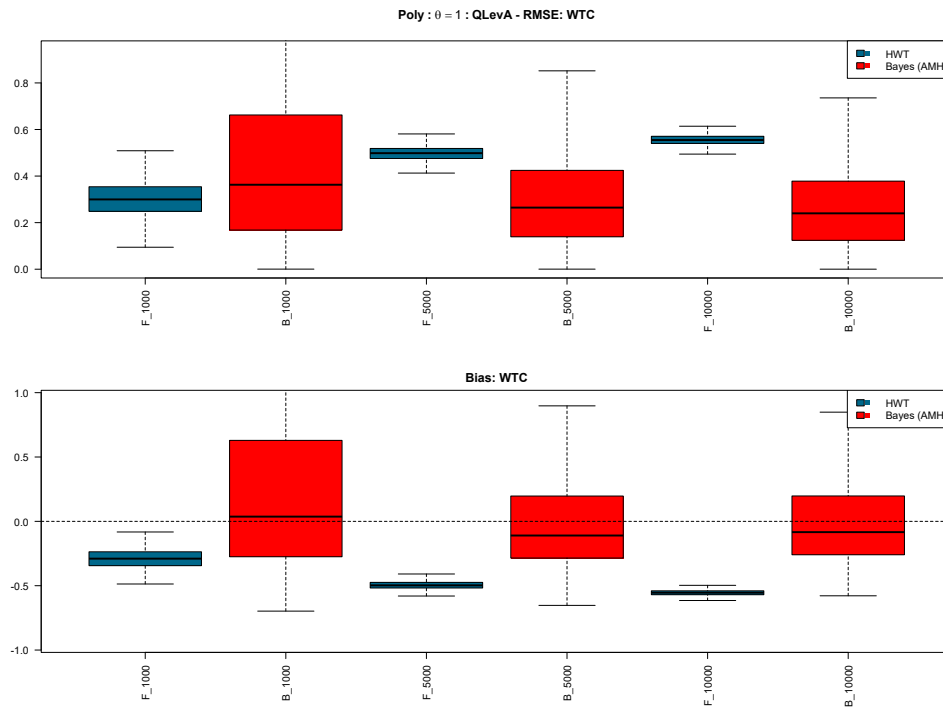


Figure B.29: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Poly, with  $\theta = 1$  using the HWT and AMH methods under quantile set B.

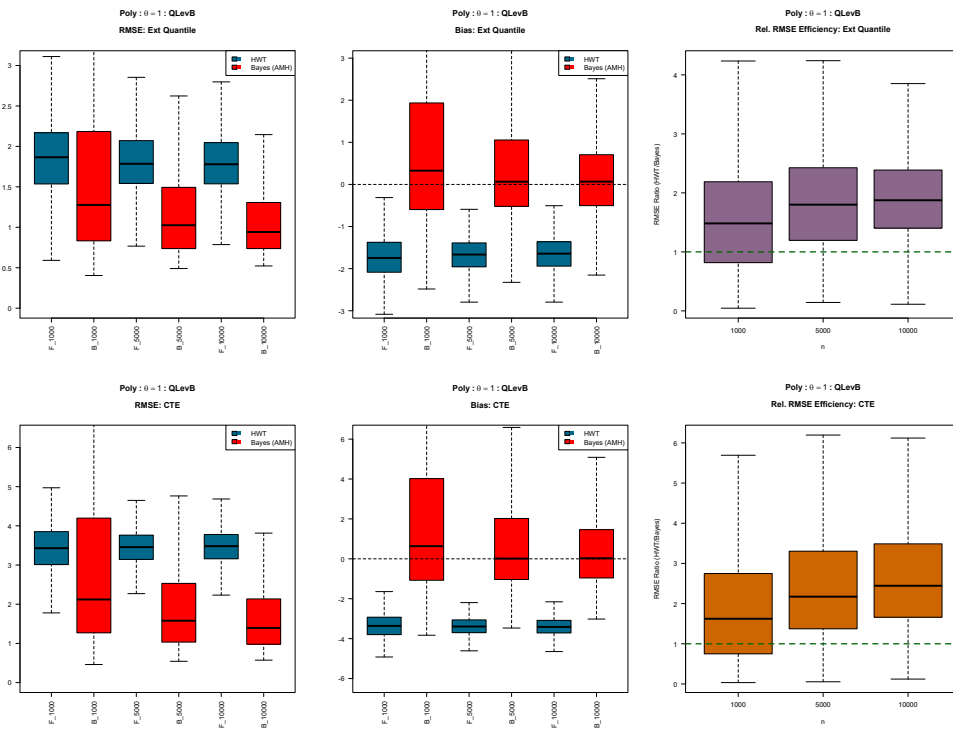


Figure B.30: RMSE and bias box plots on WTC for simulations from DM 2 - Poly, with  $\theta = 1$  using the HWT and AMH methods under quantile set B.

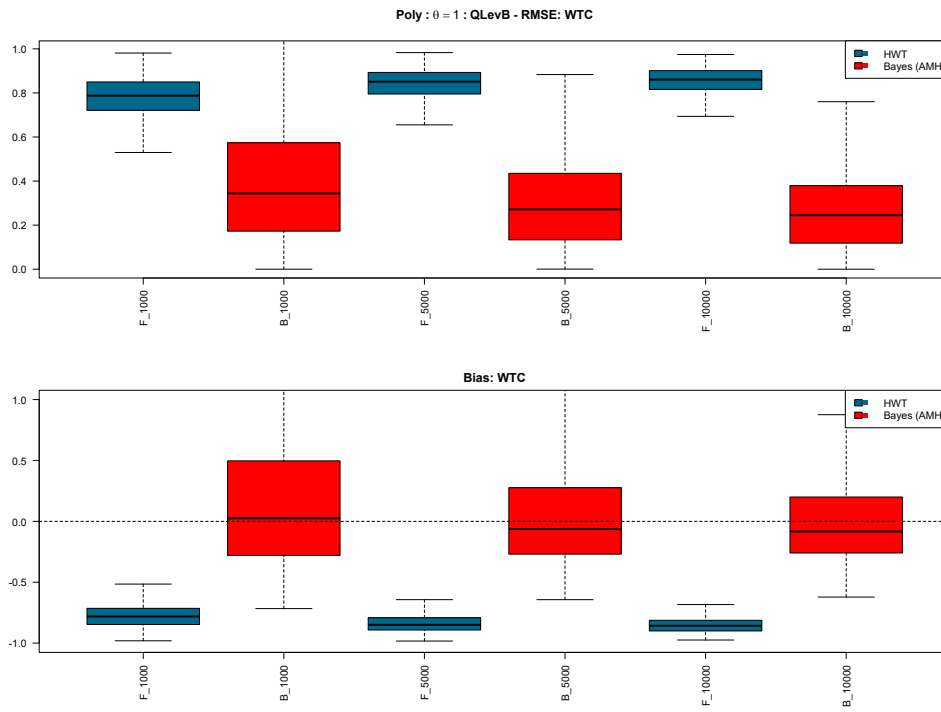


Figure B.31: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Poly, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set A.

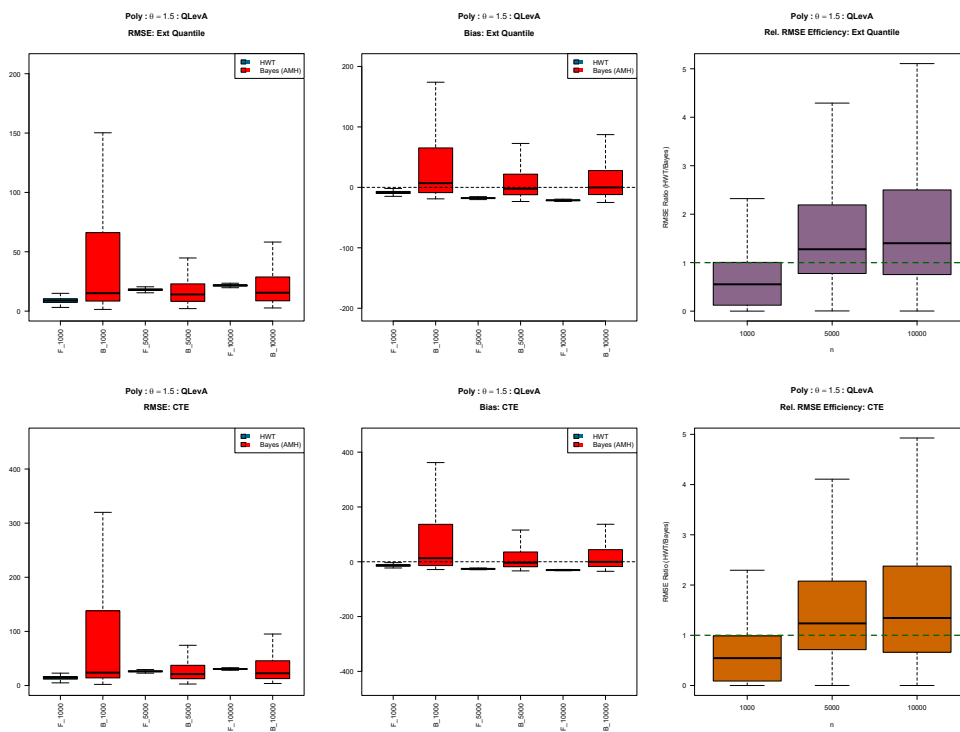


Figure B.32: RMSE and bias box plots on WTC for simulations from DM 2 - Poly, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set A.

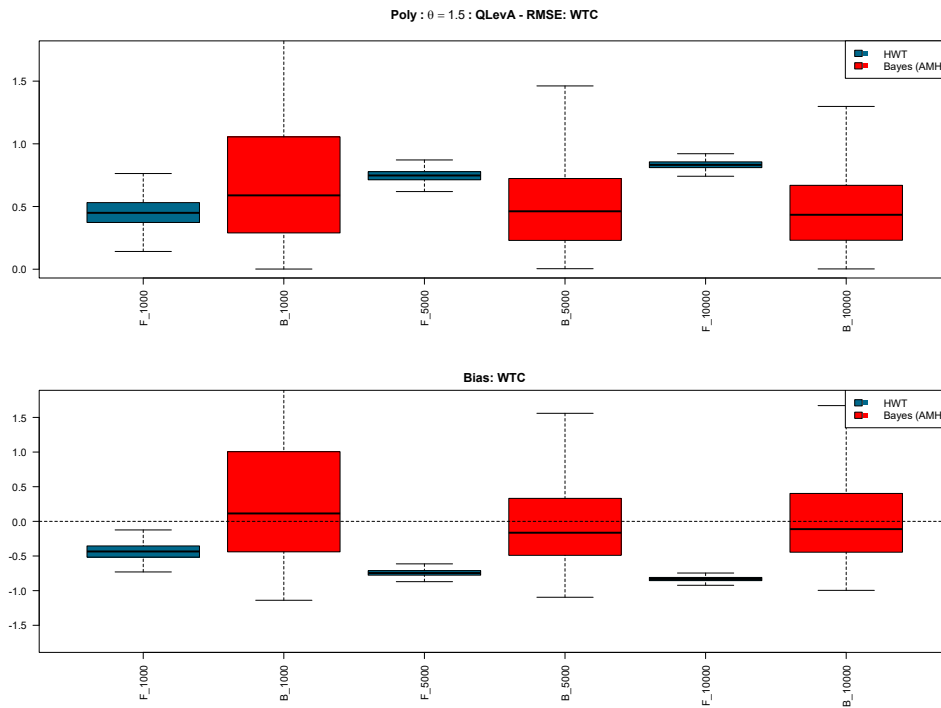


Figure B.33: RMSE, bias, and RMSE efficiency box plots on extreme quantile and CTE for simulations from DM 2 - Poly, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set B.

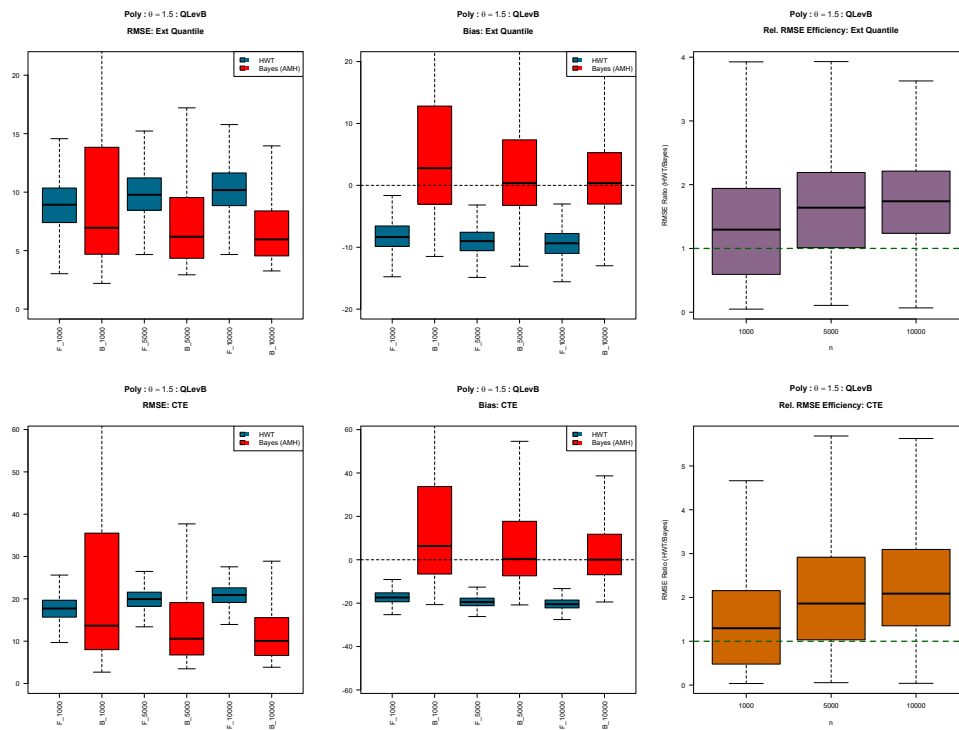


Figure B.34: RMSE and bias box plots on WTC for simulations from DM 2 - Poly, with  $\theta = 1.5$  using the HWT and AMH methods under quantile set B.

