

Statistical Analyses of Lumber Strength Properties and a Likelihood-Free Method using Empirical Likelihood

by

Yunfeng Yang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2025

© Yunfeng Yang 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Dr. Jiahua Chen
Professor, Department of Statistics
University of British Columbia

Supervisors: Dr. Samuel Wong
Associate Professor, Department of Statistics and Actuarial Science
University of Waterloo

Dr. Martin Lysy
Associate Professor, Department of Statistics and Actuarial Science
University of Waterloo

Internal Members: Dr. Pengfei Li
Professor, Department of Statistics and Actuarial Science
University of Waterloo

Dr. Audrey Béliveau
Associate Professor, Department of Statistics and Actuarial Science
University of Waterloo

Internal-External Member: Dr. Patrick Lam
Associate Professor, Department of Electrical and Computer Engineering
University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Chapter 2 has been published as: Yunfeng Yang, Martin Lysy, and Samuel W.K. Wong (2022). Multimodel Bayesian Analysis of Load Duration Effects in Lumber Reliability. In 13th International Conference on Structural Safety and Reliability (ICOSSAR 2021-2022), pages 399-407, Shanghai, China. ISBN 978-7-5478-6230-8.

Chapter 3 is based on a preprint: Yunfeng Yang, Martin Lysy, Samuel Wong (2025). A Statistical Framework for Modeling Damage from Lumber Proof Loading Experiments.

Abstract

Wood materials should meet expected strength and reliability standards for safe and stable construction. The strength of lumber and wood products may degrade over time due to sustained applied stresses, a phenomenon known as the duration-of-load (DOL) effect. The inherent variability of lumber, combined with DOL, makes structural reliability analyses particularly challenging. This thesis develops statistical methodologies to address these challenges, focusing on reliability analysis, wood strength modeling, and likelihood-free inference.

Chapter 2 evaluates the reliability of lumber, accounting for the DOL effect under different load profiles based on a multimodel Bayesian framework. Three individual DOL models previously used for reliability assessment are considered: the US model, the Canadian model, and the Gamma process model. Procedures for stochastic generation of residential, snow, and wind loads are also described. We propose Bayesian model-averaging (BMA) as a method for combining the reliability estimates of individual models under a given load profile that coherently accounts for statistical uncertainty in the choice of model and parameter values. The method is applied to the analysis of a Hemlock experimental dataset, where the BMA results are illustrated via estimated reliability indices together with 95% interval bands.

Chapter 3 explores proof-loading experiments, another industrial procedure for ensuring lumber reliability and quality, besides the DOL experiment from Chapter 2. In proof-loading, a pre-determined load is applied to remove weak specimens, but this may also weaken the surviving specimens (survivors) — a phenomenon we term the damage effect. To capture and assess this effect, we propose a statistical framework that includes a damage model and a likelihood ratio test, offering advantages over existing methods by directly quantifying the damage effect. When applied to experimental data, the proposed framework successfully detects and measures the damage effect while showing good model fit. The framework also provides correlation estimates between strength properties, potentially reducing monitoring costs in industry.

Chapter 4 investigates statistical models with intractable likelihoods, such as the Canadian model discussed in Chapter 2. To address the challenge they pose to parameter inference, various likelihood-free methods have been developed, including a recently proposed synthetic empirical likelihood (SEL) approach. We introduce a new SEL estimator based on the reparametrization trick, which greatly reduces the computational burden. The asymptotic property of our SEL estimator is derived for the situation where the number of parameters equals the number of summary statistics, leading to a method that is not only

faster, but also yields more accurate uncertainty quantification than conventional MCMC. The SEL approach is further extended by incorporating exponential tilting, which empirically improves performance when summary statistics outnumber parameters. Simulation studies validate the robustness and efficiency of our approach across various scenarios.

Acknowledgements

As I begin to write these acknowledgements, countless memories come flooding back. First and foremost, I want to express my heartfelt gratitude to my supervisors, Professor Samuel Wong and Professor Martin Lysy, for their invaluable guidance and unwavering support throughout my graduate studies. Professor Wong has been cultivating my ability to tackle statistical challenges independently. Yet, in moments when I felt completely stuck, his timely and thoughtful advice provided the guidance I needed to move forward. Thanks to his mentorship, I now have more confidence in my self-learning abilities than ever before. I am equally grateful to Professor Lysy. His sharp statistical intuition, often unraveled the core of complex problems with striking clarity. His persistent emphasis on fostering good academic habits and maintaining rigor in my work has been invaluable, significantly enhancing both the quality and efficiency of my research.

To both of my supervisors, words will always fall short of conveying the extent of my gratitude. When I faced difficulties, I often thought about how they might approach the problem and learned from their thoughtful ways of handling challenges. When I first started working with them, I was a young and inexperienced student. Now, as I complete this thesis, I feel I have grown into an independent adult, thanks to their exceptional help.

My sincere gratitude also goes to the committee members — Professor Jiahua Chen, Professor Patrick Lam, Professor Pengfei Li, and Professor Audrey Béliveau — for their time, patience in reading this thesis, thoughtful feedback, and valuable suggestions for improving this thesis.

I would also like to express my heartfelt gratitude to my colleagues and friends in the department, including Liyuan Zheng, Shiyu He, Lijia Wang, Yuxuan Zhao, Yuying Huang, and many others. I cherish the days we spent discussing statistical problems, sharing stories about life in Waterloo and Canada, and talking about our plans for the future.

Finally, I want to thank my parents for their unconditional love and care. Their unwavering support has been the foundation of everything I have achieved. Without them, this thesis and the journey in the graduate school would not have been possible.

Dedication

This thesis is dedicated to my parents, who offered me your boundless love.

Table of Contents

Examining Committee Membership	ii
Author’s Declaration	iii
Statement of Contributions	iv
Abstract	v
Acknowledgements	vii
Dedication	viii
List of Figures	xiii
List of Tables	xvii
1 Introduction	1
2 Multimodel Bayesian Analysis of Load Duration Effects in Lumber Reliability	4
2.1 Introduction	4
2.2 Methods	5
2.2.1 Models for degradation and reliability	5
2.2.2 Assessing reliability	6

2.2.3	Generating stochastic load profiles	7
2.2.4	Bayesian multimodel approach	12
2.3	Results	14
2.3.1	Experimental data	14
2.3.2	Model fitting	15
2.3.3	Reliability assessment	16
2.4	Conclusion	19
3	A Statistical Framework for Modeling Damage from Lumber Proof Loading Experiments	20
3.1	Introduction	20
3.2	Experimental data	24
3.2.1	SPF bivariate dataset	25
3.2.2	Hemlock univariate dataset	26
3.3	Damage model for lumber strength	28
3.3.1	Bivariate damage model	29
3.3.2	Univariate damage model	31
3.4	Likelihood ratio test and parameter inference	31
3.5	Simulation study	34
3.5.1	Bivariate simulation study	35
3.5.2	Univariate simulation study	41
3.6	Real data analysis	46
3.6.1	SPF bivariate dataset	46
3.6.2	Hemlock univariate dataset	49
3.7	Conclusion	54

4 Synthetic Empirical Likelihood: Asymptotic Properties and an Extension with Exponential Tilting	55
4.1 Introduction	55
4.2 Background	58
4.2.1 Empirical likelihood	58
4.2.2 Support adjustment	59
4.2.3 Bayesian EL	60
4.2.4 Bayesian synthetic likelihood	62
4.2.5 Synthetic empirical likelihood	64
4.3 Methodology	67
4.3.1 Just-identified moment condition	68
4.3.2 Over-identified moment condition	71
4.3.3 Standard error calculation	74
4.4 Simulation Study	77
4.4.1 Univariate normal distribution under just-identification	79
4.4.2 g-and-k distribution	85
4.4.3 A stochastic volatility model	87
4.4.4 Univariate normal distribution under over-identification	95
4.4.5 An M/G/1 queuing model	99
4.5 Conclusion	104
5 Conclusion	106
References	108
APPENDICES	117
A Derivation of formulas in Chapter 2	118
A.1 Derivation of equation (2.9)	118

B	Derivation of formulas in Chapter 3	119
B.1	Histograms of the Hemlock dataset	119
B.2	Derivation for the detailed expression of the likelihood (3.10) in the bivariate scenario	119
B.3	Derivation for the detailed expression of the likelihood (3.13) in the univariate scenario	121

List of Figures

2.1	Examples of stochastic live loads in a 50-year period: (a) residential loads, (b) snow load in Vancouver, (c) snow load in Quebec City, (d) wind load in Halifax. Dead load is not included in these plots.	11
2.2	Reliability analysis results under (a) residential load, (b) snow load in Vancouver, (c) snow load in Quebec City, (d) wind load in Halifax. The black lines show the estimated $\phi - \beta$ curves for the US model, Canadian model, Gamma process model, and BMA. The grey shaded regions represent 95% credible intervals obtained from BMA.	17
3.1	The density of survivor strength, generated using Equation (3.9) with the true values of α (red curve), where $\alpha = 0.6, 1, 1.4$ for (R20, T20), (R40, T40), and (R60, T60) groups, respectively. The baseline, where α is set to $\mathbf{0}$ (blue curve), is included solely for comparison to visualize the extent of damage on the survivor population.	38
3.2	The density of survivor strength, generated using Equation (3.12) with the true values of α (red curve), where $\alpha = 5, 2, 2$ for R20_1Y, R20_4Y, R5_4Y groups, respectively. The baseline, where α is set to $\mathbf{0}$ (blue curve), is included solely for comparison to visualize the extent of damage on the survivor population.	44
3.3	The Q-Q plots comparing the empirical quantiles of real data with the quantiles of simulated data from the fitted model in the bivariate real data analysis. The fitted parameters are $\hat{\mu}_X = 44.844, \hat{\mu}_Y = 5.471, \hat{\sigma}_X = 12.662, \hat{\sigma}_Y = 1.051, \hat{\rho} = 0.696$. All damage parameter values in each plot are $\hat{\alpha} = 0$	48

3.4	The Q-Q plots comparing the empirical quantiles of real data with the quantiles of simulated data from the fitted model in the univariate real data analysis. The fitted parameters are $\hat{\mu} = 50.816, \hat{\sigma} = 17.912, \hat{\eta} = 0.76$. The top-left and top-right panels evaluate the R5_4Y and R20_4Y groups, using $\hat{\alpha}_{R5_4Y} = 0$ and $\hat{\alpha}_{R20_4Y} = 0$, respectively. The bottom-left panel evaluates the R20_1Y group using the parameter estimate $\hat{\alpha}_{R20_1Y} = 4.908$, while the bottom-right panel evaluates the same group using $\alpha_{R20_1Y} = 0$	52
3.5	The Q-Q plot from the linear regression of the equal rank assumption. The empirical quantile of the strength degradation is against the estimated quantile of the fitted strength degradation. The solid line is the 45-degree line.	53
4.1	Marginal posterior densities of β_1, β_2 in binomial logistic regression model by the true posterior, (exact) EL posterior, (exact) AEL posterior, BNA of EL posterior, BNA of AEL posterior, and the true posterior.	63
4.2	The average posterior standard deviation (APSD), root mean square error (RMSE), and 95% coverage probability (CP) for the just-identified normal example, with 1000 repetitions and the number of the simulated datasets $R = 25, 100, 1000, 10000$. Error bars per R per method are approximately 6×10^{-5} for ASE, 0.003 for RMSE, and 0.007 for CP.	82
4.3	The average execution time of various methods to compute point and variance (or interval) estimates for a specific observed dataset in the just-identified normal example with $R = 1000$. The time unit is second. For MCMC methods, the speeds are obtained from a single repetition with 10000 iterations, excluding the burnin periods, while for the other methods, the speeds are averaged over 100 repetitions.	83
4.4	The left three panels display the densities of SEL, SETEL, and SL posterior with the prior $\theta \propto 1$ in the just-identified normal example, while the right three panels present the corresponding BNA densities based on SEL-Hess, SETEL-Hess, and SL-Hess. The true posterior is added as the ground truth. All densities are scaled to ensure their integration equals 1.	84
4.5	The ASE, RMSE, and 95% CP of SEL-Hess and SL-Hess for the g-and-k distribution across $R = 40, 100, 1000, 3000$, based on $N_{\text{rep}} = 1000$ repetitions for each R . The true values of parameters used to generate the observed data are $(A, B, g, k) = (3, 1, 2, 0.5)$	86

4.6	The logSL function with respect to $\boldsymbol{\xi} = (\alpha^*, \beta^*, \phi^*, \sigma^*, \rho^*)$ in the SV model, generated by 100 equally-spaced grid points, within the interval $[\xi_{\text{SL}} \pm \delta \cdot \text{SE}(\xi_{\text{SL}})]$ via SL-Hess and ND. Here $\boldsymbol{\xi}$ is a suitable transformation of $\boldsymbol{\theta}$ and $\delta = 2, 0.5, 0.1$. Each panel displays the SL function of one specific element of $\boldsymbol{\xi}$ with the red vertical line marking its SL point estimate, while all other elements are fixed at their respective SL point estimates.	90
4.7	The ASE of various proposed methods of the SV model at $R = 100, 500, 1000$ in 1000 repetitions.	92
4.8	The RMSE of various proposed methods of the SV model at $R = 100, 500, 1000$ in 1000 repetitions. Where the red lines (SETEL-Hess) are not visible in the figure, they coincide exactly with the green lines (SEL-Hess).	93
4.9	The 95% CP of various proposed methods of the SV model at $R = 100, 500, 1000$ in 1000 repetitions.	94
4.10	The ASE, RMSE and 95% CP of various methods in the over-identified normal example and 10,000 repetitions. The data comes from $N(\theta, 1)$ with the true value $\theta_0 = 0$ and the sample size $N = 25, 50$	97
4.11	The left three panels display the SEL, SETEL, and SL posterior densities in the over-identified normal example, while the right three panels present the corresponding BNA densities based on SEL-Hess, SETEL-Hess, and SL-Hess. The true posterior density is added as the ground truth. All densities are scaled to ensure their integration over whole space equals 1.	98
4.12	Densities of 5 summary statistics from the M/G/1 model with $(\theta_1, \theta_2, \lambda) = (1, 5, 0.2)$ based on 10000 simulated datasets: 25% quantile, 50% quantile, 75% quantile, minimum, maximum of $y_{1:N}$	101
4.13	The ASE, RMSE, and 95% CP of four proposed methods evaluated at $R = 20000, 40000, 80000$ across $N_{\text{rep}} = 1000$ repetitions in the M/G/1 model.	102
4.14	The logSETEL function of $\boldsymbol{\xi} = (\theta_1, \theta_2, \log(\lambda))$ in the M/G/1 model, generated by 100 equally-spaced grid points, within the interval $[\xi_{\text{SETEL}} \pm \delta \cdot \text{SE}(\xi_{\text{SETEL}})]$ via SETEL-Hess, where $\delta = 0.2, 0.02$. Each panel varies one parameter while holding all others constant at their respective SETEL point estimates. The red vertical line indicates the SETEL point estimate of the parameter being evaluated.	103

B.1	Histograms of strength distributions for truncated subsets of R100 (pieces with strength greater than 31.0275 MPa or 20.68 MPa) and for survivors from the proof-loading groups R20_1Y, R20_4Y, and R5_4Y from the Hemlock dataset.	120
-----	---	-----

List of Tables

2.1	Ground snow load parameters for various Canadian cities.	8
2.2	Parameter estimates for the US model.	16
2.3	Parameter estimates for the Canadian model.	16
2.4	Parameter estimates for the Gamma process model.	18
3.1	Summary of experiment groups from the SPF dataset. The “mean” columns denote the mean MOR or UTS values for specimens in the group which failed in that mode. The unit associated with strength is Megapascal (MPa). For example in the R20 group: 20 of the 87 specimens failed during proof loading up to MOR = 34.110 MPa, the mean MOR of these 20 specimens was 27.693 MPa, the mean UTS of the 67 specimens surviving the bending proof load was 33.430 MPa.	26
3.2	Summary of experimental groups from the Hemlock dataset. The unit associated with strength is MPa.	27
3.3	Rejection rates for the six proof loading groups given by the likelihood ratio test for the full damage case in the bivariate scenario. Rejection rates are calculated as the proportion of tests with $p < 0.05$ over 200 repetitions for each group.	36
3.4	Parameter inference results from the simulation study for the full damage case in the bivariate scenario over 200 repetitions. Conditional on the correct model (i.e., knowing all α have nonzero true values in this case), three performance measures are reported for all parameters: average estimates (AE), average standard errors (ASE), and 95% coverage probabilities (CP).	37

3.5	Rejection rates for the six proof loading groups given by the likelihood ratio test for the no damage case in the bivariate scenario. Rejection rates are calculated as the proportion of tests with $p < 0.05$ over 200 repetitions for each group.	39
3.6	Parameter inference results from the simulation study for the no damage case in the bivariate scenario over 200 repetitions. Conditional on the correct model (i.e., fixing all α as 0 in this case), three performance measures are reported for the remaining parameters: average estimates (AE), average standard errors (ASE), and 95% coverage probabilities (CP).	39
3.7	Rejection rates for the six proof loading groups given for the likelihood ratio test of the partial damage case in the bivariate scenario. Rejection rates are calculated as the proportion of tests with $p < 0.05$ over 200 repetitions for each group.	40
3.8	Parameter inference results from the simulation study for the partial damage case in the bivariate scenario over 200 repetitions. Conditional on the correct model (i.e., setting $\alpha_{R40} = 0$ and $\alpha_{T20} = 0$ in this case), three performance measures are reported for the remaining parameters: average estimates (AE), average standard errors (ASE), and 95% coverage probabilities (CP).	41
3.9	Rejection rates for the three proof loading groups given by the likelihood ratio test for the full damage case in the univariate scenario. Rejection rates are calculated as the proportion of tests with $p < 0.05$ over 200 repetitions for each group.	43
3.10	Parameter inference results from the simulation study for the full damage case in the univariate scenario over 200 repetitions. Conditional on the correct model (knowing that all α have nonzero true values in this case), three performance measures are reported for all parameters: average estimates (AE), average standard errors (ASE), and 95% coverage probabilities (CP).	43
3.11	Rejection rates for the three proof loading groups given by the likelihood ratio test for the partial damage case in the univariate scenario. Rejection rates are calculated as the proportion of tests with $p < 0.05$ over 200 repetitions for each group.	45

3.12	Parameter inference results from the simulation study for the partial damage case in the univariate scenario over 200 repetitions. Conditional on the correct model (i.e., setting $\alpha_{R20_4Y} = 0$, $\alpha_{R5_4Y} = 0$ in this case), three performance measures are reported for the remaining parameters: average estimates (AE), average standard errors (ASE), and 95% coverage probabilities (CP).	45
3.13	P-values from the proposed likelihood ratio test ($H_0 : \alpha = 0$) for the six proof-loading groups in the SPF dataset.	47
3.14	Estimate, standard error, 95% confidence intervals of the parameters from the proposed model on the SPF dataset.	47
3.15	P-values from the proposed likelihood ratio test ($H_0 : \alpha = 0$) for the three proof-loading groups in the Hemlock dataset.	49
3.16	Estimate, standard error, 95% confidence intervals for the parameters from the proposed model on the Hemlock dataset.	49

Chapter 1

Introduction

Given the widespread use of lumber in engineering and architecture, ensuring its short-term and long-term reliability is essential for safe and effective structural designs. The duration-of-load (DOL) effect, a critical consideration in this context, refers to the reduction in wood strength due to prolonged and intense applied loads. Understanding this effect is statistically complex – while various DOL models exist, they often yield different results when applied to the same dataset. Beyond the DOL effect, wood strength properties themselves, such as bending and tensile strength, are often correlated. Understanding these correlations is important for improving monitoring programs and reducing evaluation costs. Further, some DOL models incorporate random effects that make their likelihood functions intractable, which poses a significant challenge to parameter inference. Therefore, statistical methods and models are necessary to accurately capture these effects and the correlation present in experimental data of wood specimens.

To address these challenges, this thesis advances statistical methodology in three key areas. First, we introduce a Bayesian multimodel framework to effectively combine multiple DOL models, providing reliability estimates that account for model uncertainty and diverse loading profiles. Second, we propose a statistical framework to estimate the correlation between lumber strength properties and assess potential strength reduction in those specimens surviving the proof-loading testing, a useful industrial procedure for ensuring strength quality. Finally, to calibrate the parameters from models with intractable likelihoods, we develop a new synthetic empirical likelihood estimator based on the reparametrization trick and extend the method to enhance its statistical robustness.

One of the earliest investigations into the statistical relationship between the strength of wood and DOL was conducted by [Wood \(1951\)](#). Using experimental data from bending

strength tests, the author fitted a curve known as the *Madison curve* to model the percentage of short-term strength of the specimen as a function of load profile and duration of time. However, the Madison curve has limitations when predicting time-to-failure and assessing reliability under arbitrary load profiles that structures typically experience in real-world conditions.

To overcome the aforementioned challenges, *Accumulated Damage Models* (ADMs) (Gerhards, 1979; Foschi and Barrett, 1982; Foschi and Yao, 1986; Ellingwood and Rosowsky, 1991; Yang et al., 2019) have been proposed to characterize the accumulated damage of a lumber specimen as a non-decreasing function of time t , denoted by $\alpha(t)$ for $t \geq 0$. The fundamental idea behind this model is that if $\alpha(t) = 0$, the lumber specimen initially suffers from no damage, and if $\alpha(T) = 1$, the lumber specimen is destructively broken at time T . Therefore, the function is constrained to satisfy $0 \leq \alpha(t) \leq 1$ for $0 \leq t \leq T$.

The accumulated damage function $\alpha(t)$ in ADM is typically modeled as a differential equation,

$$\frac{d\alpha(t)}{dt} = f(\alpha(t), \tau(t), \boldsymbol{\theta}), \quad (1.1)$$

where $f(\cdot)$ is a specified function, $\tau(t)$ is the load profile applied to the specimen lumber at time t (detailed further below), $\boldsymbol{\theta}$ is the parameter vector in the differential equation system. The function $f(\cdot)$ and the parameter vector $\boldsymbol{\theta}$ may vary depending on the specific ADM being utilized. Two commonly-used ADMs are the U.S. model (Gerhards, 1979) and the Canadian model (Foschi and Barrett, 1982), which will be discussed in Chapter 2.

To better understand the DOL effect, specific DOL experiments are designed. The typical approach involves subjecting a lumber specimen to a pre-determined load profile, $\tau(t)$, over a specified duration t , with the time to failure recorded for each specimen (Wood, 1951; Foschi and Barrett, 1982; Gerhards, 1979). Two common types of load profiles are considered: *ramp load* and *constant load*. The ramp load applied to the specimen increases linearly until failure. Specifically, the ramp load profile is given by $\tau(t) = kt$, where k is the rate of increase and t is the time. An early study by Wood (1951) on the DOL effect, introduced a special case where the average time to failure is around 1 minute. The corresponding value of k is used to define the short-term strength of lumber specimens.

The constant load is more complex and given by

$$\tau(t) = \begin{cases} kt, & 0 \leq t \leq T_0 \\ \tau_c, & T_0 \leq t \leq T_1. \end{cases} \quad (1.2)$$

In detail, the load initially increases linearly with time, following a similar pattern as in the ramp loading test, with $\tau(t) = kt$ for $0 \leq t \leq \tau_c/k$, where k is the rate of increase, τ_c is

a predetermined load level. Once the load reaches τ_c at time $T_0 = \tau_c/k$, it is kept constant at that level until either the specimen fails or until a specified time T_1 is reached, which is typically several months or years later. In summary, there are three possible time-to-failure outcomes in constant load: during $0 \leq t \leq T_0$, specimens may fail during the initial ramp loading test; during $T_0 \leq t \leq T_1$, survivors may fail under a constant load level of τ_c ; lastly, any remaining specimens are treated as censored observations.

Note that the DOL effect only examines a single property of wood strength over long-term durations using time-to-failure data. In Chapter 3, we investigate the potential strength reduction in lumber specimens which survive proof-loading experiments. Given the data of strength measurements, our proposed statistical framework directly quantifies the reduction in strength and determines its significance, unlike DOL models that rely on time-to-failure data. If two strength properties are involved in the proof loading, our model also provides their correlation estimate.

As we shall see in Chapter 2, many DOL models have intractable likelihoods, and thus are customarily fit to data using likelihood-free methods. Yang et al. (2019) propose such a method for the Canadian model (described in Chapter 2), using an MCMC algorithm for approximate Bayesian computation (ABC) developed by Marjoram et al. (2003) to perform parameter inference. Although the approach is functional, simulation studies indicate that it is time-consuming and produces point estimates and 95% credible intervals for certain parameters that deviate from their true values. In Chapter 4, we develop a novel likelihood-free estimator based on the synthetic empirical likelihood (SEL) framework of Chaudhuri et al. (2024). We derive the asymptotic properties of our estimator for the setting where the number of parameters and summary statistics are equal, and extend the SEL framework for the setting when there are more summary statistics than parameters. Empirical evidence suggests that our estimators are not only accurate but also significantly faster than competing likelihood-free methods.

The remaining chapters of this thesis are outlined as follows. Chapter 2 evaluates lumber reliability, by accounting for the duration-of-load (DOL) effect under various load profiles using a multimodel Bayesian approach. Chapter 3 proposes a statistical framework that assesses the potential strength reduction caused by proof-loading experiments, accommodating two different scenarios. Chapter 4 explores and extends the synthetic empirical likelihood (Chaudhuri et al., 2024), using a novel likelihood approximation. Chapter 5 concludes the whole thesis.

Chapter 2

Multimodel Bayesian Analysis of Load Duration Effects in Lumber Reliability

2.1 Introduction

¹The strength of lumber and wood products may weaken over time as a result of applied stresses. This phenomenon is known as the duration-of-load (DOL) effect, and is an important factor to consider in ensuring the long-term reliability of wood-based structures. For practical reasons, experiments designed to assess DOL effects typically involve accelerated testing over a limited time period, e.g., up to a maximum of a few years. Thus, to compute DOL effects for return periods of 50 years or longer, models are needed.

Various probabilistic models have been developed for this purpose, with parameters that are calibrated from experimental data. As examples in recent reliability analyses, a study of laminated veneer lumber used the Gerhards damage accumulation model (Gilbert et al., 2019), a study of cross laminated timber used the Foschi damage model (Li and Lam, 2016), and a study of Western hemlock sawn lumber used a degradation model derived from the Gamma process (Wong and Zidek, 2019). To account for the effects of model

¹Chapter 2 has been published as: Yunfeng Yang, Martin Lysy, and Samuel WK Wong (2022). Multimodel Bayesian Analysis of Load Duration Effects in Lumber Reliability. 13th International Conference on Structural Safety & Reliability (ICOSSAR 2021-2022), ISBN 978-7-5478-6230-8. Link: <https://doi.org/10.48550/arXiv.2110.11896>.

assumptions, it can be useful to assess reliability with different models, for example as considered in [Hoffmeyer and Sørensen \(2007\)](#); [Köhler and Svensson \(2011\)](#); [Wong \(2020\)](#).

For a given model, its parameters must be estimated from data, and this uncertainty in the parameter values in turn leads to uncertainty in the computed reliability values. A Bayesian statistical approach for DOL modeling was presented in [Yang et al. \(2019\)](#), which has the advantage of coherently accounting for parameter uncertainty in reliability calculations. Nonetheless, that approach assumes that a specific model has been chosen for the analysis. The goal of this chapter is to extend the Bayesian modeling framework to combine reliability values computed from multiple DOL models. In doing so, we may compute final reliability estimates that account for both parameter and model uncertainty. The proposed multimodel framework is illustrated on construction lumber using stochastic occupancy, snow, and wind loads, with load specifications adapted from the National Building Code of Canada (NBCC) and previous studies.

2.2 Methods

2.2.1 Models for degradation and reliability

We begin by defining the damage over time for a lumber specimen via a non-decreasing function $\alpha(t)$ for time $t \geq 0$, where $\alpha(0) = 0$ signifies no damage initially and $\alpha(T) = 1$ when the specimen fails at the random time T . Also, let $\tau(t)$ denote the load applied to the specimen at time t .

Three DOL models are considered, which are briefly overviewed as follows. The U.S. model was proposed by [Gerhards \(1979\)](#), which is defined as

$$\frac{d\alpha(t)}{dt} = \exp\left(-A + B\frac{\tau(t)}{\tau_s}\right), \quad (2.1)$$

where A and B are model parameters and τ_s is the short-term strength of the specimen. The parameter τ_s is further assumed to have a lognormal distribution, i.e., $\tau_s = \tau_M \exp(wZ)$ where w is a scale parameter, Z is a standard Normal random variable, and τ_M is the median strength of the lumber population of interest.

The Canadian model is also known as Foschi-Yao model ([Foschi and Barrett, 1982](#)), which in reparametrized form specifies

$$\frac{d}{dt}\alpha(t) = [(a\tau_s)(\tau(t)/\tau_s - \sigma_0)_+]^b + [(c\tau_s)(\tau(t)/\tau_s - \sigma_0)_+]^n\alpha(t) \quad (2.2)$$

where a , b , c , n , σ_0 are random effects specific to each specimen, assumed to follow log-normal distributions, and $[x]_+ = \max(x, 0)$. The parameter σ_0 refers to a threshold that damage only happens when $\tau(t)/\tau_s - \sigma_0 > 0$.

The third DOL model is a degradation model based on the Gamma process, as proposed in [Wong and Zidek \(2019\)](#): $\alpha(t)$ is assumed to follow a Gamma process, so that the damage from time t_1 to t_2 has a gamma distribution with scale parameter ξ and shape parameter $\eta(t_2) - \eta(t_1)$ where $\eta(t)$ is a non-decreasing function that depends on $\tau(t)$. Let τ^* be a threshold below which no degradation occurs, and u a scaling parameter. Then the model for the shape parameter is

$$\eta(t) = u \sum_{i=1}^m g(\tilde{t}_i) [(\tau_i - \tau^*)_+ - (\tau_{i-1} - \tau^*)_+] \quad (2.3)$$

where $0 = \tau_0 < \tau_1 < \tau_2 < \dots < \tau_m$ is a sequence of discretized load increments that spans the range of possible loads applied, and \tilde{t}_i is the total time duration for which $\tau(t)$ exceeds τ_i . Then, an increasing function $g(\cdot)$ models the DOL effect. In [Wong \(2020\)](#), a piecewise power law was adopted, so that $g(t) \propto (t/t_i)^a a_i$ for $t_{i-1} < t \leq t_i$, where $t_0 = 0$ and t_1, t_2, \dots is a sequence of time breakpoints and a_1, a_2, \dots are the corresponding power parameters.

2.2.2 Assessing reliability

In this section, we discuss how the load $\tau(t)$ is composed of. Given a model and a set of parameter values, we may assess the long-term failure probability of a structural member under various types of loads. A stochastic load profile is simulated according to

$$\tau(t) = \phi R_o \frac{\gamma \tilde{D}_d + \tilde{D}_l(t)}{\gamma \alpha_d + \alpha_l}, \quad (2.4)$$

where ϕ is the performance factor and R_o is the characteristic strength of the lumber population considered. Further, \tilde{D}_d and $\tilde{D}_l(t)$ represent standardized dead and live loads, $\gamma = 0.25$ is the dead-to-live load ratio, $\alpha_d = 1.25$ and $\alpha_l = 1.5$ are from the NBCC 2015 edition. \tilde{D}_d is assumed to be normally distributed with mean 1.05 and standard deviation 0.1 which represents the weight of the structure and is fixed over time, while $\tilde{D}_l(t)$ can dynamically change over time and is simulated according to the specific type of load considered (see Section 2.2.3). For a simulated $\tau(t)$, the corresponding model equation, i.e., (2.1), (2.2), or (2.3), is used to predict whether the specimen fails by the end of 50 years.

2.2.3 Generating stochastic load profiles

We describe the procedures for simulating the three different load types: residency, snow, wind, for $\tilde{D}_l(t)$ in this section. Unless otherwise stated, numerical constants below are obtained from the cited references, wherein they were calibrated to data and surveys from various Canadian cities.

2.2.3.1 Residential load

Live loads for residential occupancy are modeled as the sum of two components – sustained $\tilde{D}_s(t)$ and extraordinary $\tilde{D}_e(t)$ – so that $\tilde{D}_l(t) = \tilde{D}_s(t) + \tilde{D}_e(t)$ (Foschi et al., 1989; Gilbert et al., 2019). Some examples of sustained loads can include the weight of people and furniture while the gravity of a temporary storage or a hosted party contributes to extraordinary loads.

The periods of each sustained load $\tilde{D}_s(t)$ are simulated as independent exponential random variables each with mean 10 years. During each period, the size (or intensity) of the load is simulated as independent gamma distributed random variables each with shape parameter 3.122 and scale parameter 0.0481.

For extraordinary load, $\tilde{D}_e(t)$, we similarly simulate exponential random variables to obtain periods with no extraordinary load (each with mean 1 year and size 0), alternating with short periods (each with mean 2 weeks) where the size of an extraordinary load is simulated from a gamma distribution with shape parameter 0.826 and scale parameter 0.1023.

2.2.3.2 Snow load

Snow load refers to the additional load applied to the roof of a building as a result of snowfall. Note that the amount of snow build-up per unit area on flat ground tends to differ from that of the roof of a building, due to factors that include the shape or slope of the roof, wind exposure, and melting. We refer to the former as ‘ground snow load’, and the latter as ‘roof snow load’ or simply snow load.

A typical snow load model begins by considering the annual maximum ground snow load G , which is assumed to be Gumbel distributed with a bias of \tilde{G} and a coefficient of variation (CoV) denoted $\text{CoV}(G)$ (Foschi et al., 1989). A random value G^* from a Gumbel distribution G can be simulated using the formula

$$G^* = B + \frac{-\log(-\log(p))}{A} \quad (2.5)$$

where $A = 1.282/(\text{CoV}(G) \times \bar{G})$, $B = \bar{G} - (0.577/A)$ and p is a random value drawn from the standard uniform distribution $U(0,1)$. The load associated with 50-year return period, denoted by G_{50} , is calculated by letting $p = 49/50$ in (2.5); in other words, G_{50} is the 0.98 quantile of the probability distribution of G . Calculated values of A and B are provided in Foschi et al. (1989) for various Canadian cities based on their snow histories, and reproduced in Table 2.1.

Table 2.1: Ground snow load parameters for various Canadian cities.

City	A	B
Vancouver	0.0977	5.0123
Halifax	0.1028	19.4276
Arvida	0.1255	29.4438
Ottawa	0.1082	20.8780
Saskatoon	0.1695	15.4561
Quebec City	0.3222	17.0689

The duration of winter is assumed to be five months of the year (from November 1 to April 1), and there is assumed to be no snow in the other seven months of the year (Foschi et al., 1989). For simulation purposes, each winter is divided into N_S segments of equal duration. Then within each segment, there is a certain probability of snow; if snow occurs in a segment, the ground snow load is simulated from a Gumbel distribution, where all the segment loads are assumed to be independent and identically distributed. These steps are detailed as follows.

First, the probability that there is no snow for an entire winter denoted by p_0 is calculated as

$$p_0 = e^{-e^{AB}}, \quad (2.6)$$

obtained by setting $G^* = 0$ in (2.5). Equivalently, this means there is no snow in all N_S segments that year, so the probability of snow in one segment denoted by p_e must satisfy

$$(1.0 - p_e)^{N_S} = p_0. \quad (2.7)$$

Then based on (2.6) and (2.7), p_e is calculated as

$$p_e = 1.0 - \exp \left[-\frac{1}{N_S} \exp(AB) \right], \quad (2.8)$$

If a segment has snow, then the ground snow load for the segment, which is denoted as G_s and also named the segment ground snow load, can be simulated according to

$$G_s = B + \frac{1}{A} [-\log(-N_S \cdot \tilde{p})], \quad (2.9)$$

where $\tilde{p} = \log(1 - p_e + p_e p)$ and p is a random uniform number $U(0, 1)$. The derivation of (2.9) can be found in A.1.

Second, define $g_s = G_s/G_{50}$ as the standardized ground snow load, that is, the ratio of a segment ground snow load G_s to the 50-year return period load G_{50} . A random value for g_s is simulated by

$$g_s = B' + \frac{1}{A'} [-\log(-N_S \cdot \tilde{p})], \quad (2.10)$$

where $B' = AB/(AB + 3.9019)$ and $A' = AB + 3.9019$ (i.e., $3.9019 = \log(-\log(49/50))$).

Third, following Bartlett et al. (2003) the standardized snow load q_s on the roof of a building is modelled as

$$q_s = r \cdot g_s, \quad (2.11)$$

where r is the ground-to-roof snow load transformation factor, assumed to have a log-normal distribution with a bias of 0.6 and a CoV of 0.42.

For practical implementation, we set the number of segments per winter to be $N_S = 10$, so that the length of each segment is a half-month. To summarize, for each segment we simulate a random number r_n from the standard uniform distribution $U(0, 1)$. If $r_n < p_e$, then a random snow load q_s is simulated and in (2.4) we set the live load $\tilde{D}_l(t) = q_s$ for that segment. Otherwise, the snow load is zero for that segment and we set $\tilde{D}_l(t) = 0$. For the non-winter portion of the year, $\tilde{D}_l(t) = 0$.

2.2.3.3 Wind load

Wind loads refer to the pressure of wind against the surface of a building. A model for the annual maximum wind load W has been previously conceptualized as the product of four random variables (Bartlett et al., 2003),

$$W = \zeta C_e C_p C_g, \quad (2.12)$$

where ζ is the reference velocity pressure, C_e is the exposure factor, C_p is the external pressure coefficient and C_g is the gust factor.

Following [Bartlett et al. \(2003\)](#), we define

$$\eta = C_e C_p C_g, \quad (2.13)$$

as the combination of exposure, pressure coefficient and gust, which is assumed to have a log-normal distribution with a bias of 0.68 and a CoV of 0.22.

Then ζ is determined by

$$\zeta = \frac{1}{2} \rho V^2, \quad (2.14)$$

where ρ is the density of air and treated as a constant (1.2929kg/m³ for dry air at 0° Celsius), while V is the wind velocity and modeled with a Gumbel distribution.

[Bartlett et al. \(2003\)](#) provides calibrated values for the Canadian cities Regina, Rivière-du-Loup and Halifax: the Gumbel-distributed annual maximum wind velocity V has a bias of $(1 + 3.050 \cdot \text{CoV}_a)/(1 + 2.592 \cdot \text{CoV}_a)$, where the corresponding CoV_a for the three cities are 0.108, 0.170, and 0.150, respectively. The standardized wind load w , defined as the ratio of W to the wind load for a 50-year return period W_{50} , is then given by

$$w = \frac{W}{W_{50}} = \frac{V^2 \eta}{(V^2 \eta)_{50}}, \quad (2.15)$$

where V and η are simulated from their respective Gumbel and log-normal distributions, and $(V^2 \eta)_{50} = 1.5913$ is the 0.98 quantile of the probability distribution for $V^2 \eta$ obtained via Monte Carlo simulation.

Wind loads occur over relatively short periods and only the strong winds are typically considered ([Gilbert et al., 2019](#)). Thus, we simulate an independent sequence of values for w according to (2.15), which represent the live load $\tilde{D}_l(t)$ in (2.4) during the periods of strong winds that correspond to the annual maximum of each year. These wind loads have duration 3 hours ([Bartlett et al., 2003](#)), and we assume that they occur at a random time once per year. Between these occurrences, we set $\tilde{D}_l(t) = 0$.

Figure 2.1 plots examples of simulated stochastic live loads over a 50-year period: (a) residential load, (b) snow load in Vancouver, (c) snow load in Quebec City, (d) wind load in Halifax, as discussed in this section. The dead load, which is fixed for the lifetime of the structure, is not included in these plots. In the first residential load plot, the horizontal lines represent sustained loads while the vertical lines (spike) refer to extraordinary loads. Similarly, the vertical lines in the snow load and wind load plots indicate snow loads and wind loads, respectively. Further, we notice the phenomenon that Quebec City have larger and more frequent snow load than Vancouver, which is consistent with the local wether condition.

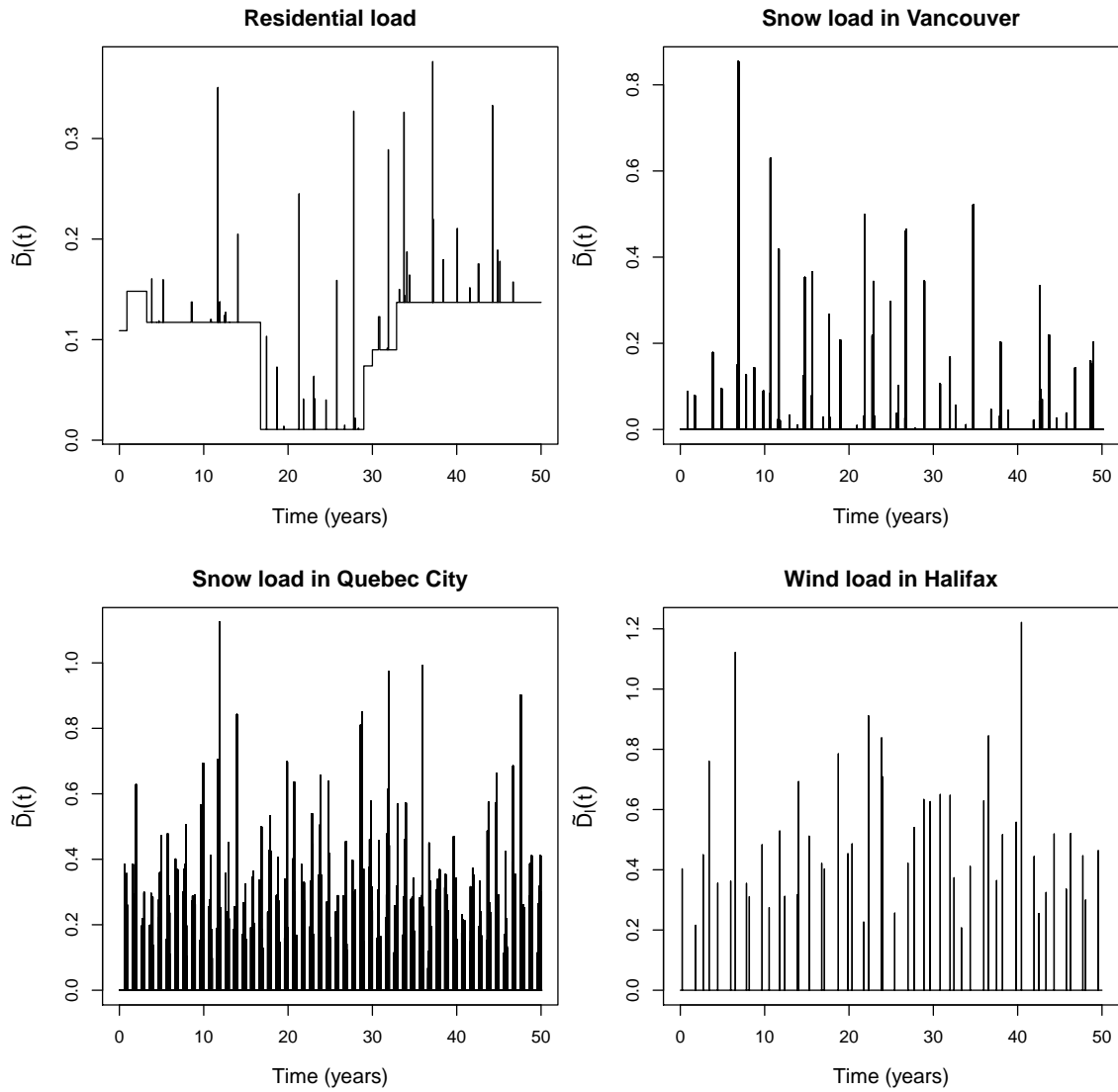


Figure 2.1: Examples of stochastic live loads in a 50-year period: (a) residential loads, (b) snow load in Vancouver, (c) snow load in Quebec City, (d) wind load in Halifax. Dead load is not included in these plots.

2.2.4 Bayesian multimodel approach

We first review the Bayesian approach to assess reliability for an individual (single) model. Let $\Delta = 1$ if a lumber specimen fails within a given timeframe (e.g., 50 years under a chosen load profile) and $\Delta = 0$ otherwise. Then for a given reliability model with parameters $\boldsymbol{\theta}$, the failure probability is

$$p_F = g(\boldsymbol{\theta}) = \Pr(\Delta = 1 \mid \boldsymbol{\theta}). \quad (2.16)$$

By writing $p_F = g(\boldsymbol{\theta})$, we emphasize the fact that the failure probability is a function of $\boldsymbol{\theta}$. Since the true value of $\boldsymbol{\theta}$ is unknown, it must be estimated from a sample of observed data $y = (y_1, \dots, y_n)$ (e.g., observed failure times in an accelerated testing experiment). In a Bayesian context, this is achieved by specifying a prior distribution $p(\boldsymbol{\theta})$, from which we obtain the posterior distribution

$$p(\boldsymbol{\theta} \mid y) = \frac{p(y \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(y)}. \quad (2.17)$$

The Bayesian estimator of p_F is then the posterior failure probability given the data,

$$\begin{aligned} \hat{p}_F &= \Pr(\Delta = 1 \mid y) \\ &= \int \Pr(\Delta = 1 \mid \boldsymbol{\theta})p(\boldsymbol{\theta} \mid y)d\boldsymbol{\theta}. \end{aligned} \quad (2.18)$$

Typically the integral in (2.18) cannot be evaluated in closed form; rather it is stochastically approximated in the following steps:

1. Obtain draws $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(N)}$ from $p(\boldsymbol{\theta} \mid y)$. This is usually done via Markov chain Monte Carlo (MCMC) sampling techniques.
2. For each draw $\boldsymbol{\theta}^{(i)}$ generate $N_{\text{prof}} = 10^5$ stochastic load profiles using the methods of Section 2.2.3, and let $\Delta_{ij} = 1$ if load profile j resulted in a failure and $\Delta_{ij} = 0$ otherwise. The failure probability $p_F^{(i)} = \Pr(\Delta = 1 \mid \boldsymbol{\theta}^{(i)})$ is then approximated as

$$p_F^{(i)} \approx \frac{1}{N_{\text{prof}}} \sum_{j=1}^{N_{\text{prof}}} \Delta_{ij}. \quad (2.19)$$

3. Finally, the Bayesian estimator of p_F is approximated by

$$\hat{p}_F \approx \frac{1}{N} \sum_{i=1}^N p_F^{(i)}. \quad (2.20)$$

Note that this Bayesian failure probability estimator can in fact be written as

$$\hat{p}_F = \Pr(\Delta = 1 \mid y) = E[p_F \mid y], \quad (2.21)$$

i.e., \hat{p}_F is the expected value of p_F under the posterior failure probability distribution $p(p_F \mid y) = p(g(\boldsymbol{\theta}) \mid y)$. In this sense, we may quantify the statistical uncertainty about p_F by calculating the 95% credible interval; namely, the 2.5% and 97.5% quantiles of $p(p_F \mid y)$. These are readily computed by taking the 2.5% and 97.5% sample quantiles of $p_F^{(1)}, \dots, p_F^{(N)}$ obtained in Step 2 above.

The Bayesian estimator and credible interval described above apply to a single model for failure probability. The purpose of multimodel Bayesian inference is to combine information from several candidate models into the estimation of p_F . Consider a set of K candidate models M_1, \dots, M_K , with corresponding parameter vectors $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K$, and prior distributions $p(\boldsymbol{\theta}_k \mid M_k)$. Let $p(M_k)$ denote the prior probability that the true model is M_k , such that $\sum_{k=1}^K p(M_k) = 1$. Then the Bayesian model-averaging (BMA) estimate of failure probability is

$$\begin{aligned} \hat{p}_F &= \Pr(\Delta = 1 \mid y) \\ &= \sum_{k=1}^K \Pr(\Delta = 1 \mid M_k, y) p(M_k \mid y) \end{aligned} \quad (2.22)$$

where $\Pr(\Delta = 1 \mid M_k, y)$ is the posterior failure probability given the data for each model M_k as given by (2.18), and

$$p(M_k \mid y) = \frac{p(y \mid M_k) p(M_k)}{\sum_{l=1}^K p(y \mid M_l) p(M_l)} \quad (2.23)$$

is the posterior probability that the true model is M_k . As was the case for the single model estimator, the BMA estimator $\hat{p}_F = E[p_F \mid y]$ is the mean of the posterior failure probability distribution

$$p(p_F \mid y) = \sum_{k=1}^K p(p_F \mid M_k, y) p(M_k \mid y), \quad (2.24)$$

which leads to the following stochastic approximation for \hat{p}_F and its credible interval under BMA:

1. For each model M_k , follow the single-model setup to obtain MCMC draws and corresponding failure probabilities via (2.19), which we denote by $\boldsymbol{\theta}_k^{(1)}, \dots, \boldsymbol{\theta}_k^{(N)}$ and $p_F^{(ik)}$ respectively.

2. Calculate the posterior model probability $p(M_k | y)$. Since this calculation is analytically intractable except in a few special cases (e.g., [Raftery et al., 1997](#)), instead we use the approximation

$$p(M_k | y) \approx \frac{\exp(-\text{BIC}_k/2)p(M_k)}{\sum_{l=1}^K \exp(-\text{BIC}_l/2)p(M_l)}, \quad (2.25)$$

where BIC_k is the Bayesian information criterion (BIC) (e.g., [Kass and Raftery, 1995](#)):

$$\text{BIC}_k = -2 \log p(y | \hat{\boldsymbol{\theta}}_k, M_k) + \dim(\boldsymbol{\theta}_k) \log(n), \quad (2.26)$$

where $\hat{\boldsymbol{\theta}}_k$ is the maximum likelihood estimate of $\boldsymbol{\theta}_k$, $\dim(\boldsymbol{\theta}_k)$ is the number of parameters in model M_k , and n is the sample size of the observed data y .

3. Draw Z_1, \dots, Z_N from a categorical distribution on K integers such that $\Pr(Z_i = k) = p(M_k | y)$ as calculated in (2.25), and let $p_F^{(i)}$ for BMA be defined as

$$p_F^{(i)} = p_F^{(iZ_i)}, \quad i = 1, \dots, N. \quad (2.27)$$

4. Finally, analogous to the single-model setup, we now approximate the BMA estimator of p_F in (2.24) by

$$\hat{p}_F \approx \frac{1}{N} \sum_{i=1}^N p_F^{(i)}. \quad (2.28)$$

This is justified by the fact that $p_F^{(1)}, \dots, p_F^{(N)}$ are draws from the BMA posterior distribution (2.24). We may thus construct the BMA 95% credible interval for p_F by computing the 2.5% and 97.5% sample quantiles of $p_F^{(1)}, \dots, p_F^{(N)}$ obtained in Step 3, in the same way as for the single-model setup.

2.3 Results

2.3.1 Experimental data

The data used for computing reliability values in this chapter are the lumber sample specimens from the western Hemlock experiment first described in [Foschi and Barrett \(1982\)](#). To summarize briefly, the specimens were divided into ramp load and constant

load groups to maintain a similar distribution of modulus of elasticity across groups. In a ramp load test group, the load was increased linearly over time t at a given rate τ_k until the specimen failed, that is $\tau(t) = \tau_k t$. In a constant load test group, the load first increased at rate τ_k until reaching the constant load level τ_c , that is $\tau(t) = \tau_k t$ for $0 \leq t \leq \tau_c/\tau_k$; then the load was maintained at τ_c until the specimen failed or the end of the testing time period was reached (this ranged from 3 months to 4 years, depending on the group). The constant load specimens that survived to the end of the testing period were then broken using a ramp load test, see [Wong \(2020\)](#) for details. The characteristic strength R_o for this population is taken to be 20.68 MPa, which is its empirical 5th percentile.

2.3.2 Model fitting

The parameters of the three models described in Section 2.2.1 – the US, Canadian, and Gamma process models – were calibrated to the experimental data using the techniques described in [Wong \(2020\)](#). Given the failure time y_j for each data specimen j , the load function $\tau_j(t)$ applied to that specimen, and the model parameters θ_k , $k = 1, 2, 3$ for model k , the likelihood function for model k is given by

$$L_k(\theta_k | y) = \prod_{j=1}^n p_k(y_j | \tau_j(t), \theta_k), \quad (2.29)$$

where $y = (y_1, \dots, y_n)$ and the specific form of each model is derived in [Wong \(2020\)](#).

For each model, $N = 500$ sets of parameter values $\theta_k^{(1)}, \dots, \theta_k^{(N)}$ were sampled from the posterior distribution $p_k(\theta_k | y)$, for the purpose of estimating p_F as described in Section 2.2.4. Sampling from the posterior for the US and Gamma process models was performed using MCMC techniques, with a Laplace approximation applied to the US model posterior to facilitate computations. For the intractable likelihood function of the Canadian model, an approximate Bayesian computation (ABC) technique was used ([Yang et al., 2019](#)). Posterior means and 95% credible intervals for all model parameters are presented in Tables 2.2-2.4 as obtained by [Wong \(2020\)](#).

Table 2.2: Parameter estimates for the US model.

Parameter	Post. Mean	95% Cred. Interval
A	68.5	(65.0, 71.9)
B	79.7	(75.9, 83.4)
w	0.426	(0.421, 0.431)

Table 2.3: Parameter estimates for the Canadian model.

Parameter	Post. Mean	95% Cred. Interval
μ_a	-12.6	(-13.2, -12.2)
σ_a	0.41	(0.16, 0.43)
μ_b	3.66	(2.99, 4.11)
σ_b	0.09	(0.06, 0.30)
μ_c	-46.4	(-58.9, -13.0)
σ_c	0.21	(0.06, 0.87)
μ_n	-1.89	(-2.38, 0.09)
σ_n	0.33	(0.06, 0.55)
μ_{σ_0}	0.39	(-0.93, 0.90)
σ_{σ_0}	0.15	(0.07, 0.50)

2.3.3 Reliability assessment

Suppose the probability of failure p_F is provided for a given performance factor ϕ and stochastic load profile in (2.4). The first order reliability method (see e.g., Madsen et al., 2006) converts p_F into a reliability index $\beta = -\Phi^{-1}(p_F)$, where Φ is the standard Normal cumulative distribution function. By computing β for a range of values of ϕ , we obtain a curve that describes the relationship between β and ϕ .

Reliability estimates in the form of $\phi - \beta$ curves are displayed in Figure 2.2, for the different load profile scenarios described in Section 2.2.3. The broken black lines display the Bayesian posterior mean estimate $\hat{\beta}^{(k)}$ for each model $k = 1, 2, 3$, which is calculated as

$$\hat{\beta}^{(k)} = \frac{1}{N} \sum_{i=1}^N \beta^{(ik)} = \frac{1}{N} \sum_{i=1}^N -\Phi^{-1}(p_F^{(ik)}), \quad (2.30)$$

where $p_F^{(ik)}$ is the failure probability for parameter set $\theta^{(ik)}$ as computed in Section 2.2.4 for each individual model. Here, the specific load profile scenario and value of ϕ will govern the simulated loads in (2.4), and in turn the failure probability obtained from (2.19).

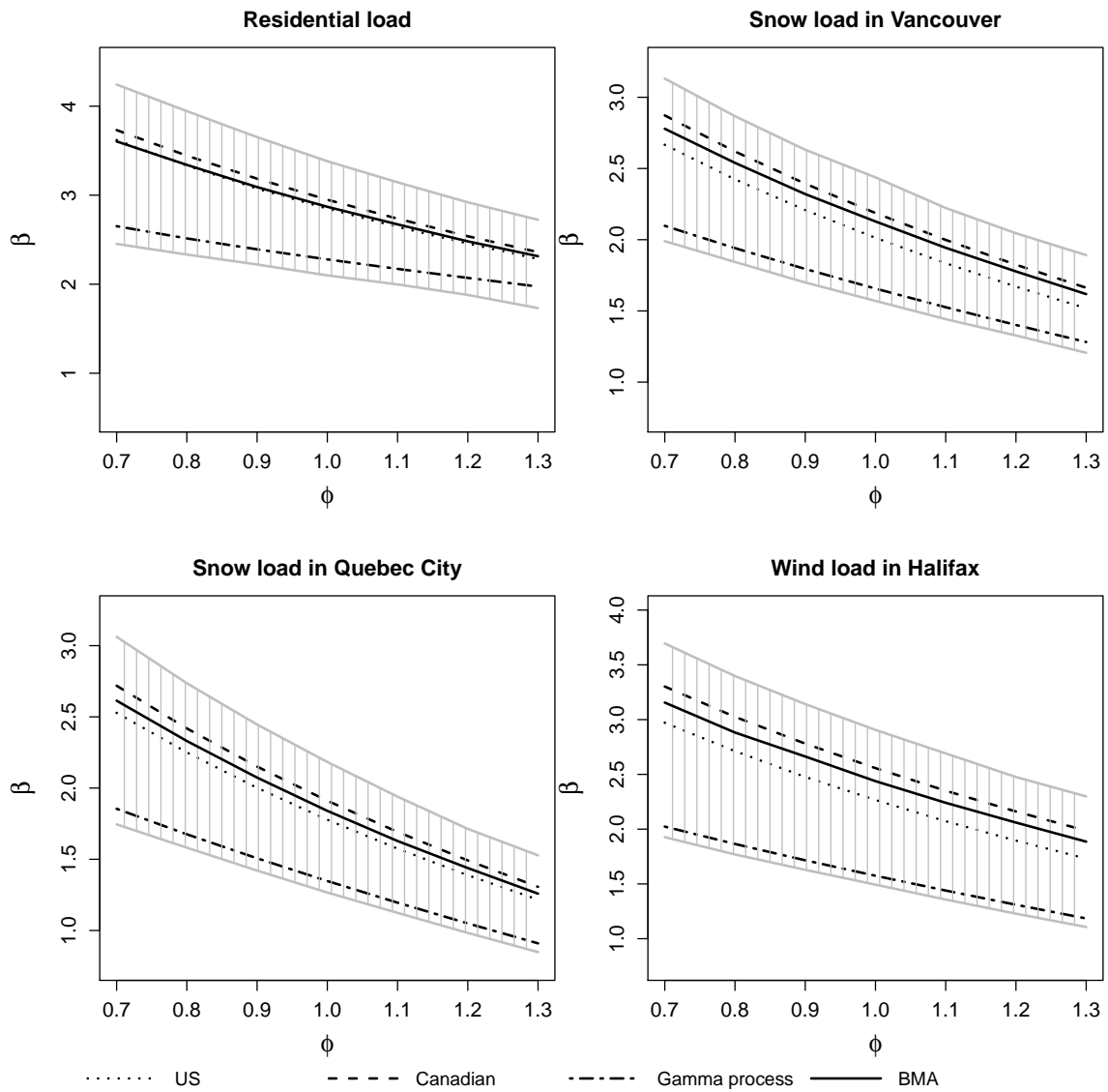


Figure 2.2: Reliability analysis results under (a) residential load, (b) snow load in Vancouver, (c) snow load in Quebec City, (d) wind load in Halifax. The black lines show the estimated $\phi - \beta$ curves for the US model, Canadian model, Gamma process model, and BMA. The grey shaded regions represent 95% credible intervals obtained from BMA.

Table 2.4: Parameter estimates for the Gamma process model.

Parameter	Post. Mean	95% Cred. Interval
u	0.084	(0.077, 0.104)
a_1	3.7×10^{-9}	$(4.6 \times 10^{-14}, 2.1 \times 10^{-3})$
a_2	0.027	(0.018, 0.028)
a_3	0.094	(0.054, 0.103)
t_1	0.00144	(0.00015, 0.00493)
t_2	2327	(289, 2890)
τ^*	4.35	(0, 4.45)
ξ	0.27	(0.20, 0.30)

The solid black lines in Figure 2.2 correspond to the multimodel BMA estimate of β . Also displayed in grey are the BMA 95% credible intervals. To obtain these, we take $p_F^{(i)}$ in (2.27) and compute $\beta^{(i)} = -\Phi^{-1}(p_F^{(i)})$ so that $\beta^{(1)}, \dots, \beta^{(N)}$ are draws from the BMA posterior distribution. Then the BMA estimate and 95% credible intervals for β are calculated in the same way as for p_F described in Section 2.2.4, i.e., by taking the mean and 2.5%/97.5% quantiles of $\beta^{(1)}, \dots, \beta^{(N)}$. The BICs calculated for the US, Canadian and Gamma process models are -5898, -6188 and -6184, respectively (Wong, 2020). Under the equal probability prior $p(M_k) = 1/3$, $k = 1, 2, 3$, the posterior probabilities (2.25) of the US, Canadian, and Gamma process models are 0.00, 0.88, and 0.12, respectively. The negligible posterior probability of the US model is due to its BIC being significantly higher than that of the other two models, indicating that the US model provides a comparatively poor fit to the data.

In all four load profile scenarios in Figure 2.2, the Canadian model is the most optimistic among the three individual models (i.e., estimating the highest β), while the Gamma process model estimates a noticeably lower reliability index than the others. The BMA estimates are closer to those of the Canadian model than those of the Gamma process model, since the Canadian model accounts for most of the posterior model probability mass (88%). Overall, the BMA 95% credible intervals contain all the estimates of the individual models. Interestingly, the BMA and US model estimates in the residential load scenario are very similar, even though the US model has zero posterior probability and therefore does not contribute to the BMA estimate.

The results across the different scenarios allow us to make several observations:

1. The reliability indices computed for snow loads in Vancouver are consistently higher than for Quebec City. This is a sensible result since Quebec City typically has a

colder and snowier winter than Vancouver.

2. The reliability index β under residential loads is higher than the other three load profile scenarios for the same values of ϕ . Referring to Figure 2.1, we see that the sustained component of residential loads is relatively low and its extraordinary component tends to be less extreme than the peak live loads due to snow and wind. This coincides with our understanding that most of the damage to specimens, and hence failures, occur during the relatively short periods when they experience the highest peak loads (Murphy et al., 1987).
3. Evidence of the DOL effect can be seen by comparing the snow load scenario in Quebec City and the wind load scenario in Halifax. While the peak loads for these two scenarios are similar (see Figure 2.1, bottom panels), snow loads are sustained for a relatively longer duration (e.g., half a month or more) compared to wind loads which are nearly instantaneous (with duration 3 hours in the simulation). Thus, it is sensible that β in the Quebec City snow load scenario is lower than that of the wind load scenario in Halifax, as more damage occurs from the longer duration of the snow loads.

2.4 Conclusion

This chapter presented a multimodel Bayesian approach for the reliability analysis of lumber products that are susceptible to load duration effects. The main advantage of the proposed BMA method is its ability to coherently account for both model and parameter uncertainty in the reliability estimates. Rather than having to choose a specific model, practitioners may run the analyses with multiple models and produce a combined estimate and 95% interval via BMA. This is of practical importance since DOL models tend to use accelerated test data to assess long-term reliability, and results may be sensitive to the assumptions of individual models. BMA provides a solution by producing a combined estimate and range of outcomes according to the likelihood of each model. We demonstrated the utility of BMA by taking models fitted to a Hemlock dataset and assessing the reliability of that lumber population under residential, snow and wind loads.

Chapter 3

A Statistical Framework for Modeling Damage from Lumber Proof Loading Experiments

3.1 Introduction

¹ The mechanical properties of lumber include its load-bearing capacity under various modes of stress encountered in engineering applications, such as tension, bending, shear, and compression (Green et al., 1999). Standard test methods are prescribed for measuring each of these so-called “strength properties” on a piece of lumber (ASTM Standard D4761, 2005). Two of these strength properties are considered in this chapter: the ultimate tensile strength (UTS) and the modulus of rupture (MOR). UTS is measured by applying tension to the ends of a specimen until it breaks, while MOR is measured by bending the specimen until failure. *Proof loading* is an industrial procedure used to ensure the quality and reliability of wooden specimens and products. While different wood grades are expected to maintain their strength properties over time, regular monitoring, such as through proof loading, is necessary to ensure that these properties remain consistent with industry standards. By subjecting specimens to a stress load with both pre-determined level and duration, then removing weaker pieces that fail under the load, proof loading ensures that the overall quality of a lumber population meets the specified design strength

¹Chapter 3 is based on a preprint: Yunfeng Yang, Martin Lysy, Samuel Wong (2025). A Statistical Framework for Modeling Damage from Lumber Proof Loading Experiments.

requirements (e.g., [Strickler et al., 1970](#); [Madsen, 1976](#); [Marin and Woeste, 1981](#); [Woeste et al., 1986](#)).

Proof loading can take various forms, in regards to the length of time (short or long) and strength properties (same or different) considered. In this chapter, we primarily investigate two scenarios: (1) *bivariate scenario*, where specimens are proof loaded under one mode (MOR or UTS) to a pre-determined stress level for a selected short-term duration (e.g., 1, 5, or 10 minutes), with survivors subsequently tested to failure under the other mode; (2) *univariate scenario*, where specimens are proof loaded under one mode to a pre-determined stress level for a selected long-term duration (e.g., 1, 2, or 4 years), with survivors tested to failure under the same mode. During proof-loading, the applied stress may weaken/reduce the strength of surviving specimens (or survivors, for short), mimicking the stress experienced by load-bearing members in a building. Therefore, evaluating the strength reduction of these survivors, which we term the *damage effect*, is critical for ensuring the long-term reliability and safety of wood-based structures.

Previous research has explored damage effect in lumber under applied stress from various perspectives, two of which are introduced here. The first one, referred to as the threshold-based damage, assumes that no damage accumulates until the applied stress exceeds a specific fraction (e.g., 50%) of the material’s short-term strength, supported by [Smith et al. \(2003\)](#), [Tseng and Yu \(1997\)](#), [Yang et al. \(2019\)](#), and [Wong and Zidek \(2019\)](#). The second one, supported by experimental evidence from datasets in [Steiner and Wesolowsky \(1995\)](#), [Cai \(2015\)](#), and [Cai et al. \(2021\)](#), suggests that weaker specimens are more vulnerable and experience greater damage effect under applied stresses. Building on these two perspectives, the first key objective of this chapter is to develop a statistical framework for evaluating potential damage effect to proof-loading survivors.

In addition to ensuring the quality of a lumber population, another well-studied application of proof loading in the bivariate scenario is to estimate the correlation ρ between two strength properties (e.g., [Suddarth et al., 1978](#)). The correlation is valuable in practice for reducing the cost of lumber monitoring programs, as it enables the measurement of one property to predict the other based on their relationship. However, estimating this correlation presents a unique challenge, as each property can only be measured through a destructive test that breaks the specimen: we cannot physically “break the same board twice” ([Evans et al., 1984](#)). As two strength properties cannot be observed simultaneously, proof-loading designs have been extensively discussed in the literature as an approach to collect the necessary data for estimating their correlation (e.g., [Johnson and Galligan, 1983](#); [Evans et al., 1984](#); [Green et al., 1984](#); [Steiner and Wesolowsky, 1995](#); [Johnson and Lu, 2007](#)).

In the single (or basic) proof loading design (SPLD) (Johnson and Galligan, 1983; Evans et al., 1984; Green et al., 1984), specimens are first tested in a designated loading mode, say mode A , up to a pre-determined level l and for a selected duration. Those specimens that fail provide their strength measurement in mode A , while survivors are then destructively tested in the other mode, mode B , to measure the strength. Using data from SPLD, maximum likelihood estimation (MLE) can be applied to estimate ρ as well as other parameters that govern the probability distributions of the two strength properties.

While the SPLD and MLE are based on principled statistical methods, simulation studies have shown that the resulting estimates of ρ can be quite unstable in practice (Green et al., 1984; Evans et al., 1984; Cai, 2015). This can in turn lead to wide confidence intervals with much uncertainty about the true value of the correlation when applied to experimental data. Various extensions to the SPLD have thus been suggested to alleviate this difficulty. De Amorim and Johnson (1986) proposed using a symmetric design: half the specimens are proof loaded under mode A and then destructively tested under mode B , while the other half are proof loaded under mode B and then destructively tested under mode A ; this type of design is also discussed in Johnson and Lu (2007). Cai (2015) described the “SPLD with a shoulder” approach: in addition to the SPLD test samples, so-called “shoulder samples” are also included. These shoulder samples are composed of specimens that are destructively tested to failure under one mode alone, to provide the marginal distribution of the strength property. De Amorim and Johnson (1986) also discussed such an approach and called it a hybrid design. Finally, experiments have also been carried out using multiple proof-loading levels (Johnson and Galligan, 1983; Cai, 2015). The second key objective of this chapter is to provide the most precise estimate of ρ possible, based on the experimental data obtained through proof-loading experiments, as described in Cai (2015).

Methods for estimating ρ have generally assumed that these proof-loading survivors are not damaged, simplifying the estimation procedure. However, whether this assumption actually holds remains an open question. Several experiments have investigated the potential damage effect to survivors caused by the proof loading with short-term duration, and their findings are summarized as follows. Strickler et al. (1970) and Woeste et al. (1986) concluded that proof loading does not significantly damage lumber strength, even at high stress levels. Madsen (1976) and Marin and Woeste (1981) found that proof loading improves the minimum strength levels of the lumber (e.g., the weakest specimens are removed) while generally not causing damage to the material, regardless of testing conditions. Katzengruber et al. (2006) further supported these findings, demonstrating that nearly all specimens withstand higher stresses in a double proof load procedure, reinforcing the overall reliability of proof loading as a quality assurance technique. Steiner and

Wesolowsky (1995) pointed out that some damage to weaker pieces, i.e., specimens with breaking strengths that only slightly exceed the proof-loading level, might be possible.

To the best of the authors' knowledge, Cai (2015) was the first reported work to explicitly include a damage parameter in the estimation of a correlation model for lumber strength properties. Specifically, if proof loading under mode A causes no damage, the observed strength of a survivor under mode B is the original (undamaged) strength y . However, if damage occurs, the observed strength under mode B is a reduced value y^* , known as *residual strength*. This necessitates models that relate the observed residual strength y^* to the original strength y , accounting for factors such as proof-loading level and duration.

Cai (2015) postulated a simple model for the residual strength, namely $y^* = y - \alpha/y$ where α is a parameter for the damage effect, which captures the vulnerability of weaker specimens (i.e., weaker survivors with inherently lower values of y , might sustain more damage than stronger survivors). However, Cai (2015) estimated α separately from ρ and nonetheless assumed no damage up to a 20th percentile proof loading level. In a subsequent study, Cai et al. (2021) applied copula models to analyze the same dataset. Their assessment of proof-loading damage relied on quantile-quantile (Q-Q) plots, comparing experimental data to simulated data from their fitted model under the no-damage assumption. While visual inspection of Q-Q plots qualitatively suggested potential damage, particularly for weaker specimens, this graphical approach could not quantify the magnitude of damage or assess its statistical significance.

In the univariate scenario, pieces are first proof loaded at a specific level l in a chosen mode for a longer-term duration. Survivors are then destructively tested in the same mode to measure their strength. Due to potential damage from proof loading, the observed residual strength of survivors, may be reduced to y^* , where $y^* \leq y$, similar to the bivariate scenario. Duration-of-load (DOL) models describe the weakening of lumber over time due to applied loads, including the Madison curve (Wood et al., 1960), the Gerhards (or US) model (Gerhards and Link, 2007), the Foschi (or Canadian) model (Foschi and Yao, 1986), and the Gamma process model (Wong and Zidek, 2019). The details of DOL models have been reviewed in Chapters 1 and 2. In addition, Karacabeyli and Lum (1994) proposed a model to quantify damage effect by modeling the strength degradation in survivors using the equal rank assumption. Specifically, the strength degradation is defined as $1 - \text{Survivors MOR}/(\text{Shoulders MOR at the same rank})$, where higher degradation reflects greater damage. A linear regression was then applied, with strength degradation as the response variable and proof-loading level and duration as covariates, to explain and predict strength degradation.

Despite their utility, both DOL models and the method of equal rank assumption have their limitations. DOL models primarily focus on fitting failure time without providing interpretable parameters for damage quantification. While the method of equal rank assumption is capable of quantifying damage, it lacks rigorous tools to assess the significance of strength degradation (i.e., damage effect) and can produce practically meaningless negative degradation values. Furthermore, the regression used in this equal-rank method only predicts (and provides confidence intervals for) the mean strength degradation for each experimental setting, leading to potentially large residuals for individual specimens.

The considerations above motivate the contribution of this chapter. Building on the perspectives of threshold-based damage and the vulnerability of weaker specimens, we propose a statistical framework, comprising a damage model and a likelihood ratio test, to directly characterize the damage effect caused by proof loading. The proposed framework is flexible, accommodating both bivariate and univariate scenarios. The damage model leverages the joint likelihood of all available experimental data, allowing damage parameters and strength property parameters to be estimated together, including the correlation coefficient ρ in the bivariate scenario. The likelihood ratio test is designed to assess the presence of significant damage in proof-load survivors.

The remainder of the chapter is organized as follows. Section 3.2 describes the experimental data collected at the FPIInnovations testing laboratory for both the bivariate and univariate scenarios. Section 3.3 introduces the proposed damage model, while Section 3.4 outlines the parameter inference method and the likelihood ratio test. Simulation studies in various cases are presented in Section 3.5, followed by the analysis of real data in Section 3.6. Finally, the chapter concludes in Section 3.7.

3.2 Experimental data

The two experimental datasets (or real data) analyzed in this chapter were produced at the FPIInnovations testing laboratory in Vancouver, British Columbia, and its predecessor, Forintek. The experiments focused on assessing the strength properties of lumber pieces, specifically their load-bearing capacity under various stress modes and levels.

The first dataset, called the SPF dataset, consists of spruce-pine-fir (SPF) specimens tested in 2011 for studying the relationships between lumber strength properties, as initially reported in Cai (2015). In the SPF dataset, specimens in proof-loading groups were proof-loaded under either bending or tension, and survivors subsequently tested to failure under the other mode, thereby corresponding to the bivariate scenario. The second dataset, called

the Hemlock dataset, consists of the Western Hemlock specimens which originate from the proof loading experiments of [Foschi and Barrett \(1982\)](#). In the Hemlock dataset, both proof-loading and the measurement of survivor strength were conducted in the bending mode, thereby corresponding to the univariate scenario.

Each dataset contains proof-loading groups and shoulder groups. The shoulder groups, also referred to as ‘baseline groups’, ‘control groups’, or ‘reference group’ in other literature, consist of specimens tested directly to failure. They provide baseline measurements of marginal strength properties and estimated quantiles of strength distributions used to determine the proof-loading levels. In both datasets, to account for piece-to-piece variations, groups are divided to ensure a similar distribution of modulus of elasticity (MOE) across them, since the measurement of MOE is non-destructive unlike MOR and UTS measurements.

3.2.1 SPF bivariate dataset

In the SPF dataset, three bundles of nominal 12 foot 2-by-4 specimens were used for the experiment: two bundles of SPF (spruce-pine-fir) 1650f-1.5E, and one bundle of SPF #2. This combined sample of specimens might thus be considered to represent, to some extent, the SPF lumber population commonly used for construction applications in Canada.

The bending machine utilized a four-point set up with a test span of 73.5 inches. The tension machine was set to a test span of 96 inches with tensile forces applied parallel to grain. These machines were used to apply proof loads and to measure MOR and UTS respectively, in accordance with the test specifications in [ASTM Standard D4761 \(2005\)](#).

The specimens were divided into eight groups: six proof-loading groups following a symmetric design, and two additional groups serving as shoulder groups. For three of the proof-loading groups, specimens were proof loaded in the tension mode up to the estimated 20th, 40th, or 60th percentile of the UTS distribution respectively, so that approximately 20%, 40%, or 60% of the specimens failed during tension proof loading and yielded UTS measurements; surviving specimens (survivors) were then tested in the bending mode to obtain MOR measurements. We refer to these three groups as T20, T40, and T60. Analogously, for the other three proof-loading groups, specimens were proof loaded in the bending mode up to the estimated 20th, 40th, or 60th percentile of the MOR distribution respectively, and UTS was measured on the survivors. These groups are denoted as R20, R40, and R60.

In addition, the tension and bending shoulder groups directly measured UTS and MOR to failure without prior proof loading, providing the marginal distributions for UTS and

MOR, respectively. These groups are denoted as T100 and R100. Each shoulder group contains 174 specimens. Based on these shoulder groups, the 20th, 40th and 60th percentiles of UTS and MOR were estimated for this population of lumber, and subsequently used to set the pre-determined proof-loading levels for the six remaining groups. Table 3.1 summarizes some statistics for the eight groups, including the proof-loading levels in each mode, the number and percentage of specimens broken by proof loading. All units associated with strength are Megapascal (MPa).

Group	sample size	proof load level	mean MOR	mean UTS	# (percent) broken by proof load
R100	174	—	45.619	—	—
T100	174	—	—	37.846	—
R20	87	MOR = 34.110	27.693	33.430	20 (23.0%)
R40	87	MOR = 42.132	33.064	34.923	40 (46.0%)
R60	87	MOR = 48.843	36.729	40.660	54 (62.1%)
T20	87	UTS = 20.409	46.494	17.815	14 (16.1%)
T40	87	UTS = 27.482	49.929	20.314	38 (43.7%)
T60	87	UTS = 33.879	55.962	23.603	52 (59.8%)

Table 3.1: Summary of experiment groups from the SPF dataset. The “mean” columns denote the mean MOR or UTS values for specimens in the group which failed in that mode. The unit associated with strength is Megapascal (MPa). For example in the R20 group: 20 of the 87 specimens failed during proof loading up to MOR = 34.110 MPa, the mean MOR of these 20 specimens was 27.693 MPa, the mean UTS of the 67 specimens surviving the bending proof load was 33.430 MPa.

3.2.2 Hemlock univariate dataset

In the Hemlock dataset, both proof-loading and the measurement of survivor strength were conducted in the MOR mode. Specimen samples of nominal 2-by-6 Western hemlock lumber, classified as No. 2 grade or better, were tested using a three-point bending setup with a 3.51m (138-inch) span. Two proof-loading levels were designed: 4500 psi (31.028 MPa) and 3000 psi (20.685 MPa), which represent the 5th and 20th percentiles of MOR. The specimens were divided into four groups: R100, R20_1Y, R20_4Y, R5_4Y, described as follows:

1. A set of 140 pieces was directly tested to failure via MOR in the R100 shoulder group.

2. A set of 300 pieces was tested in R20_1Y group, under a proof loading level of 31.028 MPa for a duration of 1 year. In total, 57 pieces failed before the proof loading level was reached, 97 pieces failed during the 1-year proof loading duration, and 146 pieces survived after the 1-year proof loading duration.
3. A set of 101 pieces was tested in R20_4Y group, under a proof loading level of 31.028 MPa for a duration of 4 years. In total, 23 pieces failed before the proof loading level was reached, 41 pieces failed during the 4-year proof loading duration, and 37 pieces survived after the 4-year proof loading duration.
4. A set of 198 pieces were tested in R5_4Y group, under a proof loading level of 20.685 MPa for a duration of 4 years. In total, 5 pieces failed before the proof loading level was reached, 42 pieces failed during the 4-year proof loading duration, and 151 pieces survived after the 4-year proof loading duration.

The statistics of the experimental groups are summarized in Table 3.2. The unit associated with strength is MPa. Figure B.1 in Appendix presents histograms of strength distributions for truncated subsets of the R100 (pieces with strength greater than 31.0275 MPa or 20.68 MPa) and for survivors from the proof-loading groups R20_1Y, R20_4Y, and R5_Y. Compared to the truncated R100, the three proof-loading groups exhibit lower density in the left tail of their strength distributions, suggesting that weaker pieces may be more likely to be weakened or removed by proof loading. Moreover, although both share the same proof-loading level, the R20-1Y group displays a heavier right tail than R20-4Y (e.g., with a greater proportion of survivors exceeding 80 MPa in strength), resulting in a higher mean of survivor strength in Table 3.2.

Group	sample size	duration	proof load level	mean of survivor strength	# (percent) broken by proof load
R100	140	—	—	—	—
R20_1Y	300	1 Year	31.028	61.436	154 (51.3%)
R20_4Y	101	4 Years	31.028	55.614	64 (63.3%)
R5_4Y	198	4 Years	20.685	51.796	47 (23.7%)

Table 3.2: Summary of experimental groups from the Hemlock dataset. The unit associated with strength is MPa.

3.3 Damage model for lumber strength

For the notation setup, let Y denote the random variable for the original (undamaged) strength of a specimen in the mode of interest, Y^* the residual strength in the mode after proof loading, and X the strength in the proof-loading mode. In the bivariate scenario, X and Y represent distinct strength properties measured in different modes, whereas in the univariate scenario, X and Y are under the same mode, giving $X = Y$. Their respective observations are denoted as y , y^* , and x . Following Cai (2015), we assume the pair (X, Y) (after suitable transformation) follows a bivariate normal distribution.

As introduced in Section 3.1, we propose the damage model, building on the assumptions of threshold-based damage and the vulnerability of weaker specimens. The model, which integrates the threshold parameter η , is given by

$$Y^* = h(Y, \alpha) \cdot I\{l > \eta \cdot X\} + Y \cdot I\{l \leq \eta \cdot X\}, \quad (3.1)$$

where $I\{\cdot\}$ is the indicator function that equals 1 when the condition inside the brackets is satisfied and 0 otherwise, l is the proof-loading level, η is the threshold parameter $0 \leq \eta \leq 1$, α is the damage parameter, and $h(Y, \alpha)$ is the damage function that satisfies $h(Y, \alpha) \leq Y$ conditional on $l > \eta \cdot X$. The proposed damage model indicates that when the proof loading stress level l exceeds $\eta \cdot X$, the survivor strength may be reduced to $Y^* = h(Y, \alpha)$, where α may depend on proof-loading level and duration. The reduction in strength, or damage effect, is represented by $Y - Y^*$.

To capture the vulnerability of weaker specimens, the damage function is considered as

$$h(Y, \alpha) = Y - \alpha \cdot \frac{\mu_Y}{\mu_X} \left(\frac{l}{\eta} - X \right), \quad (3.2)$$

where $\alpha \geq 0$, μ_Y and μ_X are the mean parameters of Y and X from the bivariate normal. The proposed damage function exhibits several key properties. First, weaker specimens (lower X) experience greater damage effect (higher $Y - Y^*$). Second, the ratio μ_Y/μ_X ensures $h(Y, \alpha)$ and Y remain on the same scale, since $(l/\eta - X)$ is in the unit of X . Third, the damage effect increases with α . Note that when $\alpha = 0$, the damage model (3.1) reduces to $Y^* = Y$, indicating that survivors experience no damage. In this case, η becomes indeterminate, as the model does not depend on η . Last, we comment that the damage function can be further extended as $Y - \alpha \cdot \frac{\mu_Y}{\mu_X} \left(\frac{l}{\eta} - X \right)^b$, where $b > 0$ is the power parameter. However, in this chapter, setting $b = 1$ is sufficient to give reasonable results while maintaining simplicity.

The remainder of this section discusses the detailed model and the likelihood expression for each scenario.

3.3.1 Bivariate damage model

This subsection begins by reviewing the bivariate normal model for two strength properties based on the SPLD experiment. This model, developed by Cai (2015), does not account for damage effect. Subsequently, we present our extension of this model to incorporate the damage effect.

3.3.1.1 Bivariate normal model for SPLD with a shoulder

The two random variables X and Y , representing the lumber strength measurements in modes A and B , respectively, are assumed to follow a bivariate normal distribution with parameters $\boldsymbol{\beta} = (\mu_X, \sigma_X, \mu_Y, \sigma_Y, \rho)$, where μ_X and μ_Y are the means of X and Y , σ_X and σ_Y are their standard deviations, and ρ is the correlation between X and Y . The probability density function (PDF) of (X, Y) is denoted as $f_{X,Y}(x, y; \boldsymbol{\beta})$. The marginal distributions of X and Y are normal, with PDFs $f_X(x; \mu_X, \sigma_X)$ and $f_Y(y; \mu_Y, \sigma_Y)$.

Consider a proof-loading group that proof loads specimens under mode A to load level l , followed by testing survivors to failure under mode B . Assuming no damage effect, the observed (strength) value w for one specimen from the group is

$$w = \begin{cases} x & \text{for } x < l, \\ y & \text{for } x \geq l. \end{cases} \quad (3.3)$$

Note that x and y do not need to have the same scale or unit.

Cai (2015) showed that the likelihood of w is expressed as

$$\tilde{L}_p(\boldsymbol{\beta}; w) = \begin{cases} f_X(x; \mu_X, \sigma_X) & \text{for } x < l, \\ \int_l^\infty f_{X,Y}(x, y; \boldsymbol{\beta}) dx & \text{for } x \geq l. \end{cases} \quad (3.4)$$

where the integration expressed as

$$\int_l^\infty f_{X,Y}(x, y; \boldsymbol{\beta}) dx = \frac{\exp\left[-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right]}{\sqrt{2\pi}\sigma_Y} \int_\tau^\infty \frac{1}{\sqrt{2\pi}} e^{-v^2/2} dv, \quad (3.5)$$

and

$$\tau = \frac{l - \mu_X - \rho\sigma_X(y - \mu_Y)/\sigma_Y}{\sigma_X\sqrt{1 - \rho^2}}. \quad (3.6)$$

The addition of a shoulder group for mode B provides data from the marginal distribution of Y , and the likelihood of the observed value y for one specimen from the shoulder group is simply the PDF of the normal distribution, which is $L_s(\mu_Y, \sigma_Y; y) = f_Y(y; \mu_Y, \sigma_Y) = 1/(\sqrt{2\pi}\sigma_Y) \exp[-(y - \mu_Y)^2/\sigma_Y^2]$.

So a proof-loading group with N_p specimens $\mathbf{w} = \{w_1, \dots, w_{N_p}\}$ and a shoulder group with N_s specimens $\mathbf{y} = (y_1, \dots, y_{N_s})$, would have the likelihood

$$\tilde{L}(\boldsymbol{\beta}; \mathbf{w}, \mathbf{y}) = \left\{ \prod_{i=1}^{N_p} \tilde{L}_p(\boldsymbol{\beta}; w_i) \right\} \left\{ \prod_{i=1}^{N_s} L_s(\mu_Y, \sigma_Y; y_i) \right\}. \quad (3.7)$$

Analogous expressions of w can be derived by interchanging x and y , for the case where the proof load is under mode B and survivors are tested under mode A .

3.3.1.2 Damage model for proof loading effect

If damage effect exist in the proof-loading group, the observed data for a specimen becomes

$$w = \begin{cases} x & \text{if } x < l, \\ y^* & \text{if } x \geq l. \end{cases} \quad (3.8)$$

The damage model in the bivariate scenario gives

$$Y^* = \left[Y - \alpha \cdot \frac{\mu_Y}{\mu_X} \left(\frac{l}{\eta} - X \right) \right] \cdot I\{l > \eta \cdot X\} + Y \cdot I\{l \leq \eta \cdot X\}. \quad (3.9)$$

Then the likelihood of w , incorporating the damage parameter α , is

$$L(\boldsymbol{\beta}, \alpha, \eta; w) = \begin{cases} f_X(x; \mu_X, \sigma_X) & \text{for } x < l, \\ \int_l^\infty f_{X,Y^*}(x, y^*; \boldsymbol{\beta}, \alpha, \eta) dx & \text{for } x \geq l, \end{cases} \quad (3.10)$$

where $f_{X,Y^*}(x, y^*; \boldsymbol{\beta}, \alpha, \eta)$ is the joint PDF of (x, y^*) . For the case of $x \geq l$, Appendix B.2 shows that

$$\int_l^\infty f_{X,Y^*}(x, y^*; \boldsymbol{\beta}, \alpha, \eta) dx = \int_l^{\frac{l}{\eta}} f_{X,Y}(x, h^{-1}(y^*, \alpha); \boldsymbol{\beta}) dx + \int_{\frac{l}{\eta}}^\infty f_{X,Y}(x, y^*; \boldsymbol{\beta}) dx, \quad (3.11)$$

where $h^{-1}(y^*, \alpha) = y^* + \alpha \cdot \frac{\mu_Y}{\mu_X} \left(\frac{l}{\eta} - x \right)$.

3.3.2 Univariate damage model

Let Y represent the random variable for the original strength of a specimen in a chosen mode. It is assumed to follow a normal distribution, $Y \sim N(\mu, \sigma^2)$, with observed value y . After a long-term proof-loading, the strength of survivors may degrade, resulting in a residual strength Y^* (with observed value y^*) such that $y^* \leq y$ for a specimen. Moreover, the proof-loading experiments ensure that the residual strength of a survivor, exceeds the proof-loading level l , thus satisfying $l < y^* \leq y$. We exclusively focus on survivors, as the primary objective in this chapter is to assess the potential reduction in their strength.

Setting $Y = X$ in the damage function (3.2) yields the univariate damage model, given by

$$Y^* = \left[Y - \alpha \cdot \left(\frac{l}{\eta} - Y \right) \right] \cdot I\{l > \eta \cdot Y\} + Y \cdot I\{l \leq \eta \cdot Y\}. \quad (3.12)$$

Let $\boldsymbol{\beta} = (\mu, \sigma)$ denote the parameters that govern the distribution of original strength property in the univariate scenario. Let $f_{\text{TN}}(y^*; \mu, \sigma, a, b)$ denote the PDF of the truncated normal distribution evaluated at y^* , where the underlying normal distribution has mean μ , variance σ^2 , and is truncated at the interval (a, b) . Then the likelihood of the observed data y^* , is given by

$$L_p(\boldsymbol{\beta}, \alpha, \eta; y^*) = \begin{cases} (\alpha + 1)^{-1} f_{\text{FN}}(h^{-1}(y^*, \alpha); \mu, \sigma, h^{-1}(l, \alpha), \infty) & \text{for } l < y^* < l/\eta \\ f_{\text{FN}}(y^*; \mu, \sigma, h^{-1}(l, \alpha), \infty) & \text{for } y^* \geq l/\eta, \end{cases} \quad (3.13)$$

where $h^{-1}(y^*, \alpha) = (\alpha + 1)^{-1}(y^* + \alpha \cdot l/\eta)$. The detailed derivation of the likelihood (3.13), along with an introduction to the truncated normal distribution, is provided in Appendix B.3.

Note that the likelihood (3.13) of η is discontinuous at $\eta = l/y^*$ unless $\alpha = 0$. In contrast, in the bivariate setting, the likelihood (3.10) of all parameters is continuous over the entire parameter space, including η , as the function (3.11) does not exhibit any discontinuity.

3.4 Likelihood ratio test and parameter inference

Suppose we have the set of proof-loading groups G_p and the set of shoulder groups G_s . Let $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \eta)$ be the full parameter vector, where $\boldsymbol{\beta}$ is the parameter vector of the distribution governing the original strength properties, $\boldsymbol{\alpha} = (\alpha_g : g \in G_p)$ is the vector of

damage parameters, where α_g represents the parameter for proof-loading group g in G_p , and η is the threshold parameter. Recall that, in the bivariate scenario, $\boldsymbol{\beta} = (\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$, while in the univariate scenario, $\boldsymbol{\beta} = (\mu, \sigma)$.

In general, we use w to represent the observed data of a specimen from a proof-loading group and s to represent that from a shoulder group. In the bivariate proof-loading scenario, if the proof load is in mode A and survivors are tested in mode B, the observation w is given by (3.8). An analogous expression applies for w when modes A and B are interchanged. In the univariate proof-loading scenario, $w = y^*$. For shoulder groups, the observation is either $s = x$ or $s = y$, depending on whether the testing mode is A or B. Specifically, let $w_{i,g}$ denote the i -th observation in proof-loading group $g \in G_p$ and $s_{j,k}$ denote the j -th observation in shoulder group $k \in G_s$.

An important question of interest is whether a proof-loading group suffers from a significant damage effect or not. Formally, this can be expressed as a hypothesis test: given a proof-loading group $g \in G_p$, the null hypothesis H_0 and alternative hypothesis H_A are

$$H_0 : \alpha_g = 0, \quad H_A : \alpha_g > 0. \quad (3.14)$$

The hypothesis can be tested using the likelihood ratio test (LRT). In order to construct the LRT, we define the function $L_{\text{LRT}}(\boldsymbol{\beta}, \alpha_g, \eta)$, which combines the likelihood contributions from the tested proof-loading group g and all shoulder groups. The function is given by

$$L_{\text{LRT}}(\boldsymbol{\beta}, \alpha_g, \eta) = \underbrace{\left\{ \prod_{i=1}^{N_g} L_p(\boldsymbol{\beta}, \alpha_g, \eta; w_{i,g}) \right\}}_{\text{from the tested proof-loading group } g} \underbrace{\left\{ \prod_{k \in G_s} \prod_{j=1}^{N_k} L_s(\mu_k, \sigma_k; s_{j,k}) \right\}}_{\text{from all shoulder groups}}. \quad (3.15)$$

where N_g and N_k are the sample sizes for groups $g \in G_p$ and $k \in G_s$, respectively. Here, $\prod_{i=1}^{N_g} L_p(\boldsymbol{\beta}, \alpha_g, \eta; w_{i,g})$ is the likelihood for the proof-loading group g , based on Equation (3.10) in the bivariate scenario or Equation (3.13) in the univariate scenario, and $\prod_{j=1}^{N_k} L_s(\mu_k, \sigma_k; s_{j,k})$ is the likelihood for the shoulder group k , based on the normal PDF.

The test statistic is

$$\lambda_{\text{LR}} = -2 \left[\log L_{\text{LRT}}(\tilde{\boldsymbol{\beta}}^0, \alpha_g^0, \eta^0) - \log L_{\text{LRT}}(\tilde{\boldsymbol{\beta}}, \tilde{\alpha}_g, \tilde{\eta}) \right], \quad (3.16)$$

where $\tilde{\boldsymbol{\beta}}^0$ maximizes $\log L_{\text{LRT}}$ with respect to $\boldsymbol{\beta}$ given $\alpha_g^0 = 0$ and $\eta^0 = 1$, $(\tilde{\boldsymbol{\beta}}, \tilde{\alpha}_g, \tilde{\eta})$ maximizes $\log L_{\text{LRT}}$ with respect to $(\boldsymbol{\beta}, \alpha_g, \eta)$. The commonly used χ^2 approximation for the likelihood ratio test statistic is not considered here for two main reasons. First, unlike

under H_A , η is indeterminate and unidentifiable under H_0 . Second, the LRT based on χ^2 is inherently a two-sided test, while in this context, the interest lies in a one-sided test, specifically $H_A : \alpha_g > 0$. Therefore, we use simulation to obtain the distribution of the test statistic λ_{LR} under H_0 numerically, which facilitates the calculation of P-value.

For the given observed data, the steps of performing the likelihood ratio testing via simulation are summarized as follows:

1. Given the observed dataset of the proof-loading group g being tested and shoulder groups, calculate the test statistic λ_{LR} and $\tilde{\beta}^0$.
2. Generate the simulated dataset of the tested proof-loading group and shoulder groups using $\tilde{\beta}^0$, $\alpha_g = 0$ and $\eta = 1$, with the same sample size as observed dataset.
3. Calculate the test statistic λ_{LR}^* for the simulated data from Step 2.
4. Repeat Step 2-3 N_λ times (e.g., $N_\lambda = 1000$). The empirical distribution of the N_λ simulated test statistics λ_{LR}^* is the numerical distribution of λ_{LR} under $H_0 : \alpha_g = 0$.
5. Calculate the P-value as the proportion of simulated λ_{LR}^* values that are greater than the test statistic λ_{LR} from Step 1.

This LRT may be applied to each proof-loading group in turn. The final (reduced) parameter vector θ is obtained by setting $\alpha_g = 0$ in $\hat{\theta}$ for groups where $H_0 : \alpha_g = 0$ is not rejected. The resulting joint likelihood $L(\theta)$ is then given as,

$$L(\theta) = \underbrace{\left\{ \prod_{g \in G_p} \prod_{i=1}^{N_g} L_p(\beta, \alpha_g, \eta; w_{i,g}) \right\}}_{\text{from all proof-loading groups}} \underbrace{\left\{ \prod_{k \in G_s} \prod_{j=1}^{N_k} L_s(\mu_k, \sigma_k; s_{j,k}) \right\}}_{\text{from all shoulder groups}}. \quad (3.17)$$

Compared to the LRT function (3.15), the joint likelihood $L(\theta)$ (3.17) includes all proof-loading groups and shoulder groups, providing an overall statistical inference using the entire dataset.

Parameter estimation can be conducted by maximum likelihood estimation (MLE). Thus, the point estimate of the parameters, obtained via MLE, is

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} L(\theta). \quad (3.18)$$

The estimated covariance matrix $\hat{\mathbf{V}}$ of $\hat{\boldsymbol{\theta}}$ is calculated as

$$\hat{\mathbf{V}} = \left[- \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log L(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right]^{-1}, \quad (3.19)$$

which is the inverse of the observed Fisher information matrix. The $1 - \gamma$ level confidence interval for the i -th element $\boldsymbol{\theta}_i$ of $\boldsymbol{\theta}$ is calculated as $\left[\hat{\boldsymbol{\theta}}_i \pm |z_{\frac{\gamma}{2}}| \cdot \sqrt{\hat{\mathbf{V}}_{ii}} \right]$, where $\gamma \in (0, 1)$, z_{γ} is the γ quantile value of the standard normal distribution, $\hat{\boldsymbol{\theta}}_i$ is the i -th element of $\hat{\boldsymbol{\theta}}$, and $\hat{\mathbf{V}}_{ii}$ is the i -th diagonal element of $\hat{\mathbf{V}}$.

When the likelihood function is discontinuous in certain parameters, the standard asymptotic theory for MLE does not apply. In such cases, these parameters are held fixed at their point estimates, and variance estimation and confidence interval construction are carried out only for the remaining parameters.

3.5 Simulation study

In this section, we present our simulation study for both the bivariate and univariate scenarios. The simulation study is designed to closely reflect the real data, with sample sizes for each simulated lumber group matching those observed in the real dataset. The true values of $\boldsymbol{\beta}$ in the simulation are based on the fitted values in the real data analysis (Section 3.6). For each scenario, we consider various damaged cases, characterized by different true values of $\boldsymbol{\alpha}$ and η . One case replicates the real data analysis, using the fitted $\boldsymbol{\alpha}$ and η values as the true values. Other cases modify specific $\boldsymbol{\alpha}$ and η values to assess the robustness of the proposed method. In each case, we evaluate the performance of the likelihood ratio test in selecting significant α and conduct parameter inference to assess the precision and uncertainty of estimates, both based on $N_{\text{rep}} = 200$ repetitions.

Specifically, for the likelihood ratio test, the numerical distribution under $H_0 : \alpha = 0$ is obtained by $N_{\lambda} = 1000$ simulated datasets. The rejection rate of H_0 is calculated as the proportion of $p < 0.05$ across the repetitions. For parameter inference, three metrics are assessed: averaged estimates (AE) of the parameters, average standard errors (ASE) of the estimates, and coverage probabilities (CP), which is the proportion of confidence intervals that enclose the true parameter value.

As noted in [Fan and Li \(2002\)](#), assessing variable selection (i.e., testing $H_0 : \alpha = 0$ in our study) and coverage simultaneously can be challenging. When the variable selection and coverage are evaluated together, any damage parameter not rejected by the likelihood

ratio test is set to 0 in a given repetition. However, if the true $\alpha > 0$, setting it to 0 results in under-coverage, as the confidence interval does not include the true value for those repetitions. Hence we assess coverage probabilities (and other two metrics) conditional on the correct model, i.e., parameters α with nonzero true values are estimated across all repetitions, while those with true values of 0 are treated as known. In other words, LRT is not used when calculating the three metrics for parameter inference.

3.5.1 Bivariate simulation study

To mimic the experimental dataset, each simulated dataset consists of the shoulder groups $G_s = \{\text{R100}, \text{T100}\}$ and the proof-loading groups $G_p = \{\text{R20}, \text{R40}, \text{R60}, \text{T20}, \text{T40}, \text{T60}\}$. The damage parameters for the six simulated proof-loading groups are denoted by $\boldsymbol{\alpha} = (\alpha_{\text{R20}}, \alpha_{\text{R40}}, \alpha_{\text{R60}}, \alpha_{\text{T20}}, \alpha_{\text{T40}}, \alpha_{\text{T60}})$. The sample sizes match those from the experimental dataset described in section 3.2, with $N = 87$ specimens in each proof loading group and $2N = 174$ specimens in each shoulder group, resulting in a total sample size of $87 \times 6 + 174 \times 2 = 870$.

The true value of $\boldsymbol{\beta}$ is set as $\boldsymbol{\beta} = (\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) = (45, 5.5, 13, 1, 0.7)$, similar to the fitted values from real data analysis in section 3.14. The true value of the threshold parameter is set to $\eta = 0.7$ if applicable. As to true values of the damage parameter $\boldsymbol{\alpha}$, three cases are considered: (i) All six proof loading groups are damaged. (ii) All six proof-loading groups are not damaged. (iii) Parts of six proof loading groups are damaged while the rest of them are not damaged. Case (ii) is designed to mimic the real data analysis result, while cases (i) and (iii) are added to assess the robustness of our proposed method.

3.5.1.1 Simulation case (i): full damage effect

In this case, the true values of damage parameters $\boldsymbol{\alpha} = (\alpha_{\text{R20}}, \alpha_{\text{R40}}, \alpha_{\text{R60}}, \alpha_{\text{T20}}, \alpha_{\text{T40}}, \alpha_{\text{T60}})$ are set to $(0.6, 1, 1.4, 0.6, 1, 1.4)$. In other words, all six proof loading groups are damaged. Further, the true parameters indicate that the groups under higher proof loading levels have higher damage effect.

To provide visual insight, Figure 3.1 presents the density of survivor strength, generated using Equation (3.9) with the true values of $\boldsymbol{\alpha}$ (red curve). The baseline, where $\boldsymbol{\alpha}$ is set to $\mathbf{0}$ (blue curve), is included solely for comparison to visualize the extent of damage on the survivor population. For a given group, the extent of damage is reflected by the magnitude of deviation between the red and blue densities.

When comparing panels in Figure 3.1 vertically (i.e., with the same proof-loading mode, such as R20, R40, and R60), groups with higher true values of α and proof-loading levels exhibit larger deviations between the red and blue densities, indicating stronger damage effect. When comparing panels horizontally (i.e., across different modes but with the same proof-loading quantile, such as R40 and T40), the R groups show greater deviations between the densities.

Table 3.3 summarizes the rejection rates of H_0 given by the likelihood ratio test, with data generated using the true values of α . The results align with the findings in Figure 3.1: proof-loading groups with higher rejection rates in the table exhibit greater deviations between the red and blue densities in the figure.

Conditional on the correct model (i.e., knowing that all α have nonzero true values in this case), Table 3.4 reported AE, ASE, 95% CP for all parameters over 200 repetitions. We see that all parameters except ρ are well estimated, with AE values close to the true value and CP near the nominal level of 95%. The estimate of ρ exhibits a slight bias compared to the true value and a higher ASE than those in case (ii) and case (iii) from subsequent subsections. One of the reasons is that, under the correct model where all damage parameters are significant, the information for ρ and α provided by proof-loading groups are intertwined, challenging the estimation of ρ . Our additional experiments show that increasing the sample size to $N = 87 \times 3 = 261$, improves the AE of ρ to 0.707, bringing it closer to the true value.

True value	$\alpha_{R20} = 0.6$	$\alpha_{R40} = 1$	$\alpha_{R60} = 1.4$
Rejection rate	0.24	0.97	1
True value	$\alpha_{T20} = 0.6$	$\alpha_{T40} = 1$	$\alpha_{T60} = 1.4$
Rejection rate	0.21	0.805	1

Table 3.3: Rejection rates for the six proof loading groups given by the likelihood ratio test for the full damage case in the bivariate scenario. Rejection rates are calculated as the proportion of tests with $p < 0.05$ over 200 repetitions for each group.

parameters	AE	ASE	CP
$\mu_X = 45$	45.047	0.695	0.951
$\mu_Y = 5.5$	5.504	0.056	0.958
$\sigma_X = 13$	13.079	0.545	0.939
$\sigma_Y = 1$	1.000	0.041	0.958
$\rho = 0.7$	0.738	0.103	0.940
$\eta = 0.7$	0.696	0.025	0.985
$\alpha_{R20} = 0.6$	0.593	0.215	0.935
$\alpha_{R40} = 1$	0.972	0.254	0.970
$\alpha_{R60} = 1.4$	1.375	0.299	0.975
$\alpha_{T20} = 0.6$	0.627	0.224	0.945
$\alpha_{T40} = 1$	0.994	0.236	0.955
$\alpha_{T60} = 1.4$	1.411	0.287	0.950

Table 3.4: Parameter inference results from the simulation study for the full damage case in the bivariate scenario over 200 repetitions. Conditional on the correct model (i.e., knowing all α have nonzero true values in this case), three performance measures are reported for all parameters: average estimates (AE), average standard errors (ASE), and 95% coverage probabilities (CP).

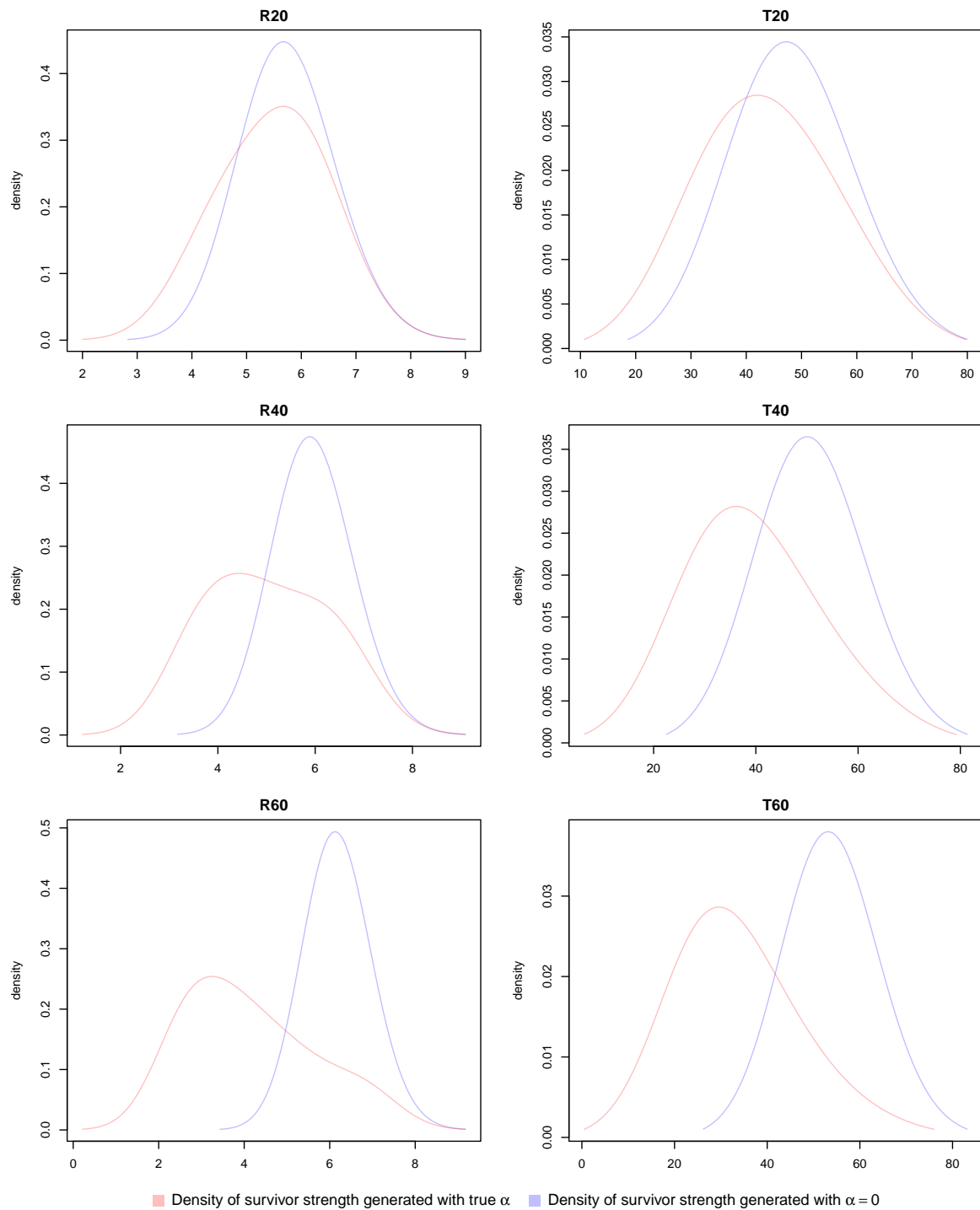


Figure 3.1: The density of survivor strength, generated using Equation (3.9) with the true values of α (red curve), where $\alpha = 0.6, 1, 1.4$ for (R20, T20), (R40, T40), and (R60, T60) groups, respectively. The baseline, where α is set to 0 (blue curve), is included solely for comparison to visualize the extent of damage on the survivor population.

3.5.1.2 Simulation case (ii): no damage effect

In this case, the true values of damage parameters α are set as $\alpha = (\alpha_{R20}, \alpha_{R40}, \alpha_{R60}, \alpha_{T20}, \alpha_{T40}, \alpha_{T60}) = (0, 0, 0, 0, 0, 0)$. In other words, all six proof loading groups are not damaged. Notice that in this case, the threshold parameter $\eta = 0.7$ does not contribute to the data generating process.

Table 3.5 summarizes the rejection rates of H_0 given by likelihood ratio test. It is found that all rejection rates are around the chosen significance level 0.05. By setting $\alpha = 0$, $\eta = 1$ in the joint likelihood (3.17), the simulated datasets are fitted with the parameters $\theta = (\mu_X, \sigma_X, \mu_Y, \sigma_Y, \rho)$. Conditional on the correct model (i.e., fixing all α as 0 in this case), Table 3.6 reports AE, ASE, 95% CP for the remaining parameters across all repetitions. We notice that all five bivariate normal parameters are well estimated.

True value	$\alpha_{R20} = 0$	$\alpha_{R40} = 0$	$\alpha_{R60} = 0$
Rejection rate	0.07	0.065	0.03
True value	$\alpha_{T20} = 0$	$\alpha_{T40} = 0$	$\alpha_{T60} = 0$
Rejection rate	0.035	0.025	0.045

Table 3.5: Rejection rates for the six proof loading groups given by the likelihood ratio test for the no damage case in the bivariate scenario. Rejection rates are calculated as the proportion of tests with $p < 0.05$ over 200 repetitions for each group.

parameters	AE	ASE	CP
$\mu_X = 45$	45.017	0.562	0.960
$\mu_Y = 5.5$	5.509	0.046	0.940
$\sigma_X = 13$	13.009	0.436	0.945
$\sigma_Y = 1$	1.004	0.035	0.945
$\rho = 0.7$	0.701	0.050	0.975

Table 3.6: Parameter inference results from the simulation study for the no damage case in the bivariate scenario over 200 repetitions. Conditional on the correct model (i.e., fixing all α as 0 in this case), three performance measures are reported for the remaining parameters: average estimates (AE), average standard errors (ASE), and 95% coverage probabilities (CP).

3.5.1.3 Simulation case (iii): partial damage effect

In this case, the true values of damage parameters α are set as $\alpha = (\alpha_{R20}, \alpha_{R40}, \alpha_{R60}, \alpha_{T20}, \alpha_{T40}, \alpha_{T60}) = (0.6, 0, 1.4, 0, 1, 1.4)$. In other words, the groups (R40, T20) are not damaged, whereas the other groups experience varying degrees of damage effect.

Table 3.7 summarizes the rejection rates of H_0 given by the likelihood ratio test. For non-damaged groups, the rejection rates are around 0.05, while for damaged groups, the rejection rates increase with α . Conditional on the correct model (i.e., setting $\alpha_{R40} = 0$ and $\alpha_{T20} = 0$ in this case), Table 3.8 reports AE, ASE, CP over all repetitions for the remaining parameters. The correlation coefficient parameter ρ is well estimated compared to the true value. Additionally, the SE of ρ estimate is similar to that in case (ii) (0.05), indicating that ρ is reliably estimated in the presence of both damaged groups and non-damage groups, provided groups with $\alpha = 0$ are correctly identified. For those true $\alpha > 0$ and the threshold parameter η , the proposed damage model also shows good estimates and reasonable coverage probabilities.

True value	$\alpha_{R20} = 0.6$	$\alpha_{R40} = 0$	$\alpha_{R60} = 1.4$
Rejection rate	0.23	0.075	1
True value	$\alpha_{T20} = 0$	$\alpha_{T40} = 1$	$\alpha_{T60} = 1.4$
Rejection rate	0.065	0.81	1

Table 3.7: Rejection rates for the six proof loading groups given for the likelihood ratio test of the partial damage case in the bivariate scenario. Rejection rates are calculated as the proportion of tests with $p < 0.05$ over 200 repetitions for each group.

parameters	AE	ASE	CP
$\mu_X = 45$	45.032	0.687	0.960
$\mu_Y = 5.5$	5.502	0.050	0.940
$\sigma_X = 13$	12.988	0.483	0.945
$\sigma_Y = 1$	0.993	0.039	0.945
$\rho = 0.7$	0.707	0.063	0.975
$\eta = 0.7$	0.701	0.028	0.975
$\alpha_{R20} = 0.6$	0.582	0.227	0.960
$\alpha_{R60} = 1.4$	1.394	0.293	0.980
$\alpha_{T40} = 1$	0.998	0.259	0.975
$\alpha_{R60} = 1.4$	1.373	0.313	0.965

Table 3.8: Parameter inference results from the simulation study for the partial damage case in the bivariate scenario over 200 repetitions. Conditional on the correct model (i.e., setting $\alpha_{R40} = 0$ and $\alpha_{T20} = 0$ in this case), three performance measures are reported for the remaining parameters: average estimates (AE), average standard errors (ASE), and 95% coverage probabilities (CP).

3.5.2 Univariate simulation study

To mimic the experimental dataset, each simulated dataset consists of a shoulder group $G_s = \{R100\}$ and three proof-loading groups $G_p = \{R20_1Y, R20_4Y, R5_4Y\}$. The corresponding parameter vector for the simulated dataset is $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \eta)$, where $\boldsymbol{\beta} = (\mu, \sigma)$ and $\boldsymbol{\alpha} = (\alpha_{R20_1Y}, \alpha_{R20_4Y}, \alpha_{R5_4Y})$. For the generation procedure of survivors in each simulated proof-loading group, we start by generating N^* pieces from $N(\mu, \sigma^2)$. The proposed damage model (3.12) is then applied to these N^* pieces, and the N_g survivors, whose residual strengths exceed the group's corresponding l , are selected for analysis. For each proof-loading group, we set $N^* = 300$ in the simulation, matching the size of the R20_1Y group (i.e., $N^* = 57 + 97 + 146 = 300$ for R20_1Y) in the experimental dataset.

As discussed earlier, the likelihood (3.13) of η exhibits discontinuity except when $\alpha = 0$, and consequently, the joint likelihood $L(\boldsymbol{\theta})$ of η is also discontinuous. Therefore, we estimate the variance and construct confidence intervals for the remaining parameters conditional on the fixed estimate $\hat{\eta}$.

The true value of the population parameters, namely, (μ, σ, η) is set as (51, 18, 0.76), based on the fitted value from the real data analysis. Regarding the damage parameter

α , two cases are considered: (i) All proof-loading groups R20_1Y, R20_4Y, R5_4Y are damaged. (ii) Only R20_1Y is damaged. Case (ii) is designed to mimic the patterns observed in the real data analysis, while case (i) is included to evaluate the robustness of our proposed method.

3.5.2.1 Simulation case (i): full damage effect

In the simulation case (i), the true value of α is set as $(\alpha_{\text{R20_1Y}}, \alpha_{\text{R20_4Y}}, \alpha_{\text{R5_4Y}}) = (5, 2, 2)$, which means that all three proof loading groups – R20_1Y, R20_4Y, and R5_4Y group – are damaged. The choice of the true value of α implies that R20_1Y and R20_4Y groups share the same proof-loading level l but differ in α values, reflecting the their distinct damage effect, caused by distinct load durations (1 year versus 4 years). Meanwhile, R20_4Y and R5_4Y groups share the same α but differ in l . This setup allows us to examine the impact of α (or l) on the extent of damage with other factors constant.

For visual illustration, Figure 3.2 presents the density of survivor strength, generated using Equation (3.12) with the true values of α (red curve). The baseline, where α is set to $\mathbf{0}$ (blue curve), is included solely for comparison to visualize the extent of damage on the survivor population. For a given group, the extent of damage is reflected by the magnitude of deviation between the red and blue densities. Survivors, by definition, have strengths that exceed l . Hence the densities are zero below l with a sharp drop at l . For the red density, survivors with strengths weaker than l/η experience damage, causing an additional sharp drop at l/η . Above l/η , the red and blue densities in each group differ only by a scaling constant, as these survivors are not damaged.

We now compare R20_1Y and R20_4Y groups, differing in true values of α . The deviation between the red and blue densities in R20_1Y (with a higher true α) is more pronounced than that in R20_4Y. This is because a higher α leads to greater damage effect, when other factors are held fixed. Next, we compare the R20_4Y and R5_4Y groups, differing in proof-loading levels l (20% vs. 5% quantiles). The deviation between the red and blue densities in R20_4Y (with a higher l) is greater than that in R5_4Y. This can be attributed to two reasons. First, few pieces experience damage in R5_4Y, as indicated by its lower densities (both red and blue) within $[l, l/\eta]$, due to the lower l . Second, the damage effect to survivors is lower in R5_4Y, as the damage function $h(y) = y - \alpha \cdot (\frac{l}{\eta} - y)$ indicates that, for fixed y, α , and η , a lower l leads to a smaller reduction in strength.

Table 3.9 summarizes the rejection rates for the three proof-loading groups, with data generated using the true values of α , given by the likelihood ratio test with a significance level of 0.05. The results align with the findings in Figure 3.2: proof-loading groups with

higher rejection rates in the table exhibit greater deviations between the red and blue densities in the figure.

True value	$\alpha_{R20.1Y} = 5$	$\alpha_{R20.4Y} = 2$	$\alpha_{R5.4Y} = 2$
Rejection rate	1	0.94	0.41

Table 3.9: Rejection rates for the three proof loading groups given by the likelihood ratio test for the full damage case in the univariate scenario. Rejection rates are calculated as the proportion of tests with $p < 0.05$ over 200 repetitions for each group.

As to parameter inference, conditional on the correct model (i.e., knowing all α have nonzero true values in this case), Table 3.10 reports the AE, ASE, and 95% CP for all parameters over 200 repetitions. As previously noted, $\hat{\eta}$ is excluded from the standard error calculation because the gradient of its likelihood is discontinuous. All CP's are near the nominal level of 95%. The AE's of μ, σ , and η are close to their true values, with reasonable ASE's if calculated. Regarding the damage parameters, the AE's of $\alpha_{R20.1Y}$ and $\alpha_{R5.4Y}$ are slightly higher than their true values and their ASE's are somewhat high. In contrast, the AE of $\alpha_{R20.4Y}$ closely match its true value with a lower ASE. This is because the density of strength of damaged survivors, which corresponds to the interval $[l, l/\eta]$ of the red density in Figure 3.2, is higher in R20.4Y than those in R20.1Y and R5.4Y, providing more samples for estimating its damage parameter.

Parameter	AE	ASE	CP
$\mu = 51$	51.083	0.831	0.960
$\sigma = 18$	17.952	0.502	0.945
$\alpha_{R20.1Y} = 5$	5.913	2.791	0.975
$\alpha_{R20.4Y} = 2$	2.099	0.868	0.945
$\alpha_{R5.4Y} = 2$	2.698	2.056	0.935
$\eta = 0.76$	0.760	/	/

Table 3.10: Parameter inference results from the simulation study for the full damage case in the univariate scenario over 200 repetitions. Conditional on the correct model (knowing that all α have nonzero true values in this case), three performance measures are reported for all parameters: average estimates (AE), average standard errors (ASE), and 95% coverage probabilities (CP).

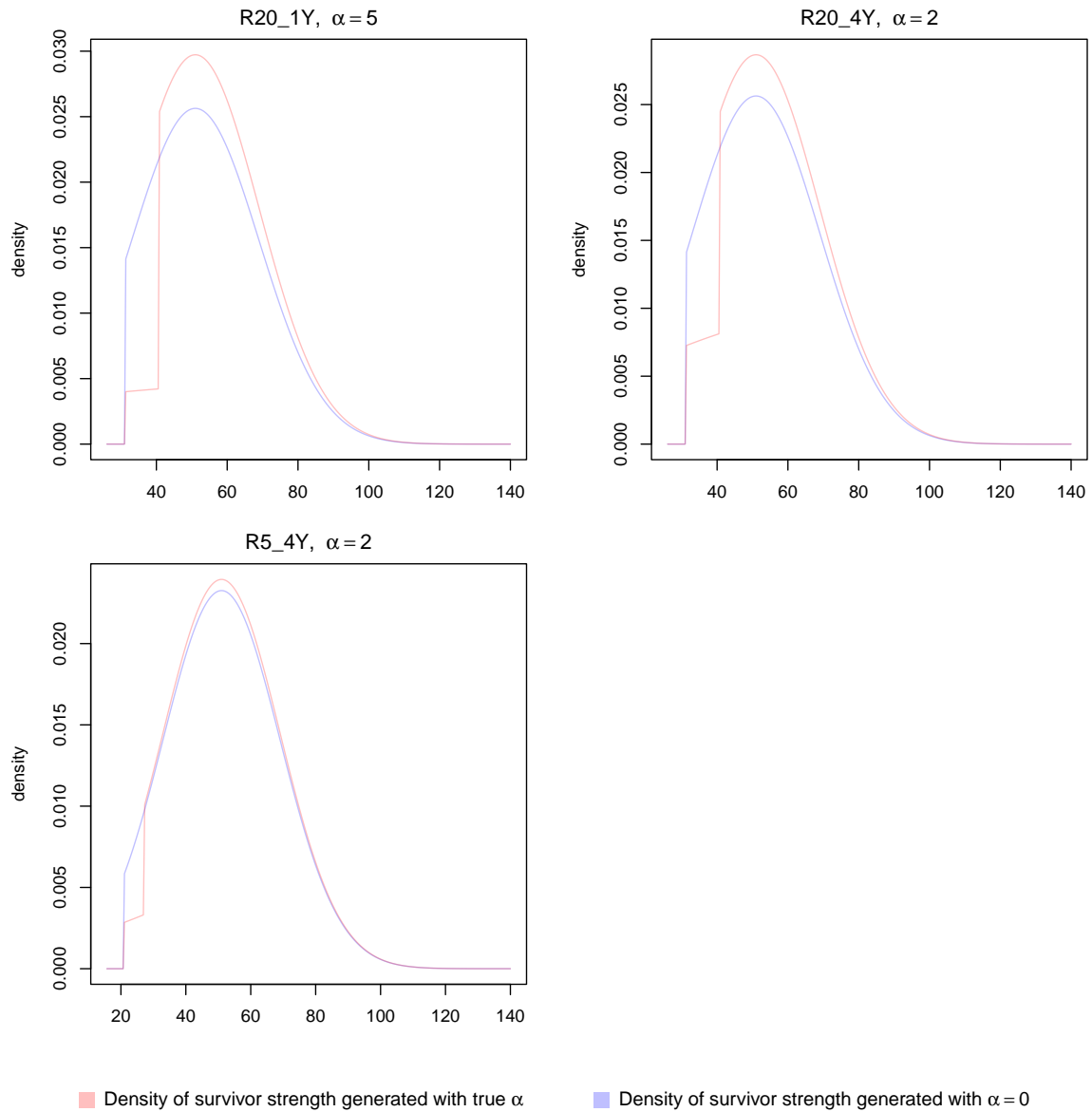


Figure 3.2: The density of survivor strength, generated using Equation (3.12) with the true values of α (red curve), where $\alpha = 5, 2, 2$ for R20_1Y, R20_4Y, R5_4Y groups, respectively. The baseline, where α is set to $\mathbf{0}$ (blue curve), is included solely for comparison to visualize the extent of damage on the survivor population.

3.5.2.2 Simulation case (ii): partial damage effect

In the simulation case (ii), the true value of α is set as $(\alpha_{R20_1Y}, \alpha_{R20_4Y}, \alpha_{R5_4Y}) = (5, 0, 0)$. It means that only R20_1Y group is damaged while R20_4Y group and R5_4Y group are not damaged. The true values are set to resemble the fitted value obtained from the real data analysis in Section 3.6.2.

Table 3.11 summarizes the rejection rates for the three proof-loading groups given by likelihood ratio test. It is found that rejection rates for R20_4Y and R5_4Y are around the chosen significance level 0.05.

Conditional on the correct model (i.e., setting $\alpha_{R20_4Y} = 0$, $\alpha_{R5_4Y} = 0$ in this case), Table 3.12 reports AE, ASE, and 95% CP for the remaining parameters, across all repetitions. All CP's are close to the nominal level (95%). The AE's of μ , σ , and η are close to the corresponding true values with reasonable ASE's. Similar to the full damage case, the ASE and the bias of $\hat{\alpha}_{R20_1Y}$ is slightly high, due to the low density of the damaged survivors in that group.

True value	$\alpha_{R20_1Y} = 5$	$\alpha_{R20_4Y} = 0$	$\alpha_{R5_4Y} = 0$
Rejection rate	1	0.045	0.035

Table 3.11: Rejection rates for the three proof loading groups given by the likelihood ratio test for the partial damage case in the univariate scenario. Rejection rates are calculated as the proportion of tests with $p < 0.05$ over 200 repetitions for each group.

Parameter	AE	ASE	CP
$\mu = 51$	51.077	0.816	0.935
$\sigma = 18$	17.939	0.527	0.955
$\alpha_{R20_1Y} = 5$	5.741	2.685	0.95
$\eta = 0.76$	0.757	/	/

Table 3.12: Parameter inference results from the simulation study for the partial damage case in the univariate scenario over 200 repetitions. Conditional on the correct model (i.e., setting $\alpha_{R20_4Y} = 0$, $\alpha_{R5_4Y} = 0$ in this case), three performance measures are reported for the remaining parameters: average estimates (AE), average standard errors (ASE), and 95% coverage probabilities (CP).

3.6 Real data analysis

In this section, we report the result of real data analysis, where our statistical framework proposed in Section 3.3 is applied to the experimental data described in Section 3.2.

3.6.1 SPF bivariate dataset

Following the exploratory analysis in Cai (2015), we found that MOR and $\sqrt{\text{UTS}}$ (square-root transformation of UTS) each approximately follow the univariate normal distribution in the SPF dataset. We denote MOR as X and $\sqrt{\text{UTS}}$ as Y . The model and fitting process are the same as described in section 3.3 and 3.4.

To determine the significance of damage effect, Table 3.13 presents the P-values from the proposed likelihood ratio test ($H_0 : \alpha = 0$) for each proof-loading group. The numerical distribution under the null hypothesis H_0 is obtained by $N_\lambda = 1000$ simulated datasets. As all P-values are higher than 0.05, we conclude that there is no significant damage effect found in any proof-loading group. Based on the result, we set $\alpha = 0$ and fit the model using 5 parameters $\theta = (\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$. Table 3.14 shows the statistics of estimates, standard error, 95% confidence intervals for these 5 parameters. The correlation coefficient is estimated as $\hat{\rho} = 0.69$ with 95% confidence intervals (0.594, 0.798). Figure 3.3 presents the Quantile-Quantile (Q-Q) plots comparing the experimental data with the simulated data generated from the fitted model. Most points align closely along the 45-degree line in the Q-Q plots, suggesting a good fit to the experimental data.

The proposed framework and analysis yield more comprehensive results than those in Cai (2015). The approach in Cai (2015) could not provide estimates for $\alpha_{\text{T20}}, \alpha_{\text{T60}}$ as the authors assumed no damage effect in T20 group. Moreover, their least squares estimate of the damage parameter was negative for T60, which has unrealistic interpretation. In the other study, Cai et al. (2021) suggested that damage effect due to proof loading may be significant, especially for weaker lumber specimens. However, their conclusion relied on a qualitative graphical method, lacking rigorous numerical characterization and formal hypothesis testing of the damage effect. In contrast, our proposed approach using likelihood ratio test concludes no apparent damage effect in any proof-loading group.

Moving to the correlation estimate, the method in Cai (2015) is restricted to estimating ρ from a single group, yielding $\hat{\rho} = 0.75$ based on the T20 group. In contrast, our proposed joint likelihood approach enables estimation of ρ using data from all six proof-loading groups. As a result, the SE of our $\hat{\rho}$ is only 0.05, substantially smaller than the SE reported in Cai (2015) (0.20). Using copula models, Cai et al. (2021) estimated ρ as 0.789 with the

Group	R20	R40	R60
P-value	0.723	0.649	0.483
Group	T20	T40	T60
P-value	0.234	0.295	0.755

Table 3.13: P-values from the proposed likelihood ratio test ($H_0 : \alpha = 0$) for the six proof-loading groups in the SPF dataset.

Parameters	Estimate	SE	95% CI
μ_X	44.844	0.530	(43.805, 45.883)
μ_Y	5.471	0.041	(5.391, 5.551)
σ_X	12.662	0.424	(11.831, 13.493)
σ_Y	1.051	0.031	(0.990, 1.112)
ρ	0.696	0.052	(0.594, 0.798)

Table 3.14: Estimate, standard error, 95% confidence intervals of the parameters from the proposed model on the SPF dataset.

SE of 0.072, based on the full dataset including all six proof-loading groups. The slight difference between their correlation estimate and ours may arise from the distributional assumption: we assume the strength property follows a normal distribution, whereas [Cai et al. \(2021\)](#) assume a Weibull distribution.

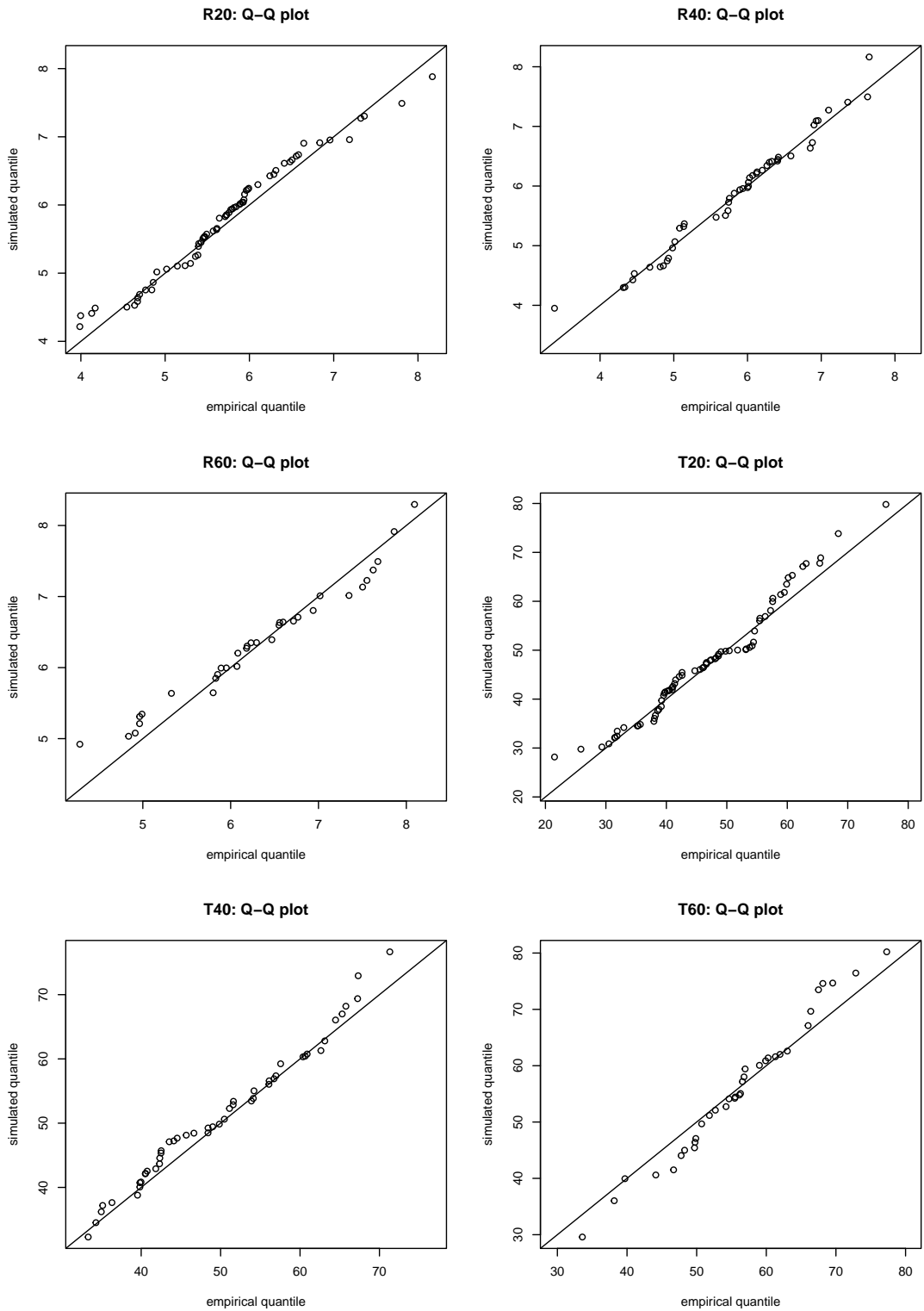


Figure 3.3: The Q-Q plots comparing the empirical quantiles of real data with the quantiles of simulated data from the fitted model in the bivariate real data analysis. The fitted parameters are $\hat{\mu}_X = 44.844$, $\hat{\mu}_Y = 5.471$, $\hat{\sigma}_X = 12.662$, $\hat{\sigma}_Y = 1.051$, $\hat{\rho} = 0.696$. All damage parameter values in each plot are $\hat{\alpha} = 0$.

3.6.2 Hemlock univariate dataset

To determine the significance of damage effect, Table 3.15 summarizes the P-values from the likelihood ratio test ($H_0 : \alpha = 0$) for the three proof loading groups from the Hemlock dataset. The numerical distribution of the test statistics under the null hypothesis is obtained by $N_\lambda = 1000$ simulated datasets. At a significance level of 0.05, the P-values in table 3.15 suggest significant damage effect in R20_1Y while no significant damage effect in R20_4Y and R5_4Y are presented.

Group	R20_1Y	R20_4Y	R5_4Y
P-value	0	0.816	0.433

Table 3.15: P-values from the proposed likelihood ratio test ($H_0 : \alpha = 0$) for the three proof-loading groups in the Hemlock dataset.

Based on the result from the likelihood ratio test, we set $\alpha_{R20_4Y} = 0$ and $\alpha_{R5_4Y} = 0$. And the final model has 4 parameters ($\mu, \sigma, \eta, \alpha_{R20_1Y}$). Table 3.16 displayed the estimate, standard error (SE), and 95% CI for these parameters. The SE of α_{R20_1Y} is slightly high, and the CI of α_{R20_1Y} covers 0 due to the small sample size of damaged pieces among the survivors. Specifically, the fitted model identifies 6 damaged survivors (i.e., with strength lower than $l/\hat{\eta}$).

Parameter	Estimate	SE	95% CI
μ	50.816	0.949	(48.955, 52.676)
σ	17.912	0.584	(16.765, 19.058)
α_{R20_1Y}	4.908	2.563	(-0.116, 9.933)
η	0.76	/	/

Table 3.16: Estimate, standard error, 95% confidence intervals for the parameters from the proposed model on the Hemlock dataset.

The first three panels in Figure 3.4 display Q-Q plots comparing the experimental data with the simulated data from the fitted model, which was generated using the parameter estimates $\hat{\mu} = 50.816$, $\hat{\sigma} = 17.912$, $\hat{\eta} = 0.76$, and the respective damage parameter estimates: $\hat{\alpha}_{R5_4Y} = 0$ for R5_4Y, $\hat{\alpha}_{R20_4Y} = 0$ for R20_4Y, and $\hat{\alpha}_{R20_1Y} = 4.908$ for R20_1Y. The majority of points align along the 45 degree line ($y = x$) in the figure, indicating a good fit to the experimental data. For comparison, the right-bottom panel in figure 3.4

demonstrates a similar Q-Q plot for R20_1Y, where simulated data are generated with the same values of $(\hat{\mu}, \hat{\sigma}, \hat{\eta})$ but with $\alpha_{R20_1Y} = 0$, assuming no damage effect. As most points lie below the 45 degree line, it suggests an under-estimation of survivor strength in R20_1Y, which further supports the presence of damage in this proof loading group.

The duration-of-load (DOL) models described in Chapter 2 also characterize the deterioration of wood strength over a long-term period, by fitting failure time. However, DOL models can only give an indirect way to quantify damage effect, as higher damage effect may lead to shorter failure time. These models lack a parameter that directly represents damage effect. In contrast, our proposed damage model incorporates α , a parameter that explicitly quantifies damage effect, and includes a corresponding likelihood ratio test to assess its significance.

For further comparison with our approach, we apply the method of equal rank and regression from Karacabeyli and Lum (1994) to the experimental data. Following the assumption of equal rank, we assume that pieces from the shoulder group and pieces from proof-loading groups have a 1-to-1 relationship. The strength degradation (SD) per piece is calculated as

$$SD = 1 - \frac{\text{Survivors MOR}}{\text{Shoulders MOR at the same rank}}. \quad (3.20)$$

In Karacabeyli and Lum (1994), only pieces with $SD > 0.05$ were considered to experience significant damage effect, and such pieces were included in the regression analysis. The regression model is

$$SD = d_0 + d_1 \cdot SL + d_2 \cdot TP + \varepsilon, \quad (3.21)$$

where SL is the stress level expressed as the ratio of the proof-loading level to the mean MOR of the shoulder group, TP is the proof-loading time period expressed by months, and ε is the error term assumed to be normally distributed with mean 0 and a constant variance.

The coefficient estimates are $\hat{d}_0 = -0.2924(0.0303)$, $\hat{d}_1 = 0.5061(0.0419)$, and $\hat{d}_2 = 0.0036(0.0003)$, with standard errors reported in the parentheses. The P-values of the t-tests for d_0, d_1, d_2 are all below 0.05, indicating all coefficients are significant. The residual standard error is 0.0420. The coefficient of determination (R square) is $r^2 = 0.6179$, which means that approximately 61.79% of the variation in the SD is accounted for by the models covariates. The low r^2 value suggests that additional factors beyond SL and TP may be required to fully explain the variation of SD.

Figure 3.5 shows the Q-Q plot of the empirical SD and the predicted SD by the linear regression, suggesting clear evidence of under-fitting. In contrast, our proposed method not only uses hypothesis testing as a more rigorous tool to determine the presence of damage

effect, but also provides a better fit to the experimental data, as demonstrated by Q-Q plots.

We also comment that the equal-rank method cannot construct a CI for strength degradation of an individual specimen, as all pieces within a proof-loading group share the same SL and TP. Therefore, using strength degradation > 0.05 is merely a heuristic way to rule out undamaged pieces as the method cannot determine whether an individual strength degradation value is significantly different from zero.

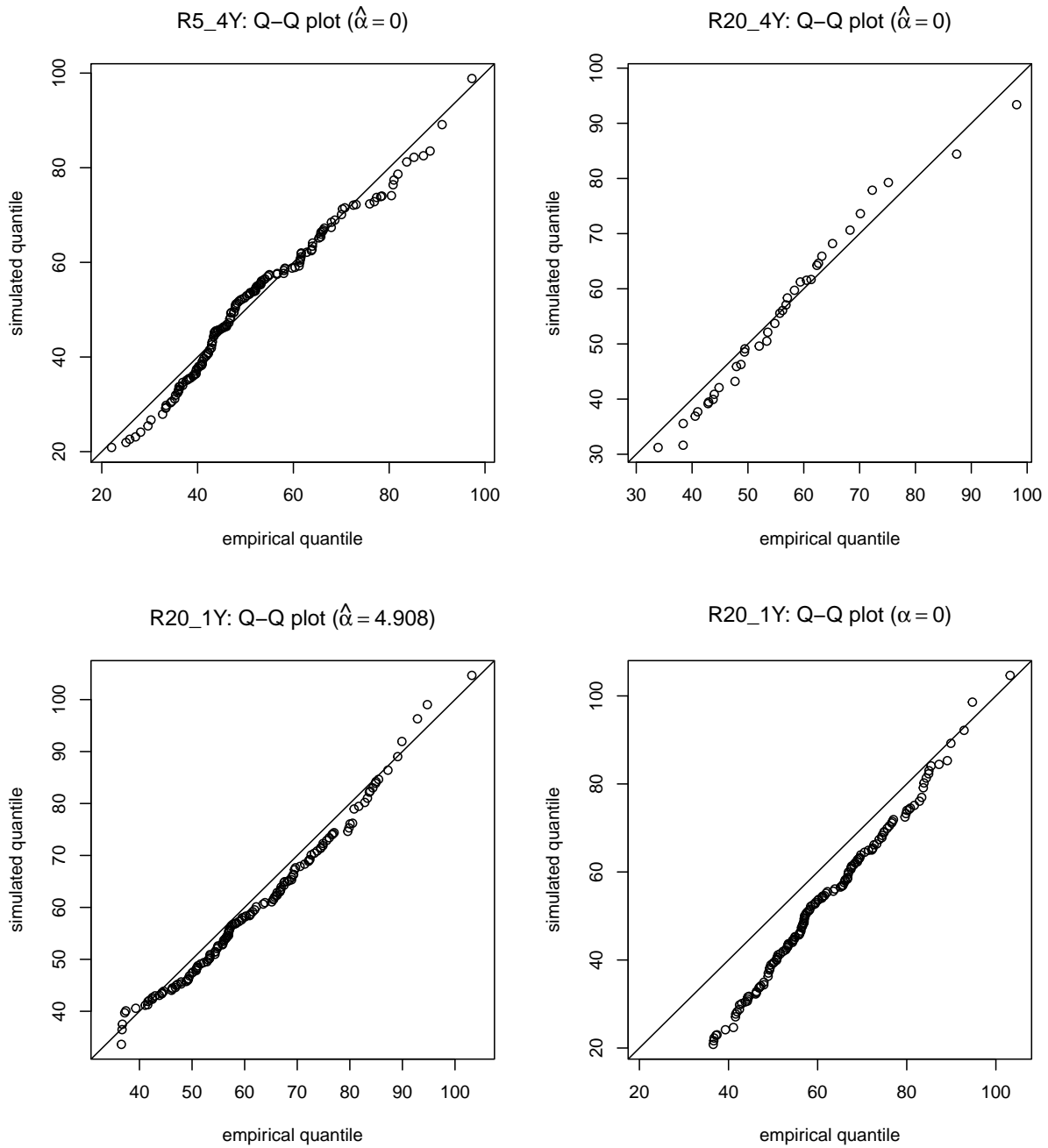


Figure 3.4: The Q-Q plots comparing the empirical quantiles of real data with the quantiles of simulated data from the fitted model in the univariate real data analysis. The fitted parameters are $\hat{\mu} = 50.816$, $\hat{\sigma} = 17.912$, $\hat{\eta} = 0.76$. The top-left and top-right panels evaluate the R5_4Y and R20_4Y groups, using $\hat{\alpha}_{R5_4Y} = 0$ and $\hat{\alpha}_{R20_4Y} = 0$, respectively. The bottom-left panel evaluates the R20_1Y group using the parameter estimate $\hat{\alpha}_{R20_1Y} = 4.908$, while the bottom-right panel evaluates the same group using $\alpha_{R20_1Y} = 0$.

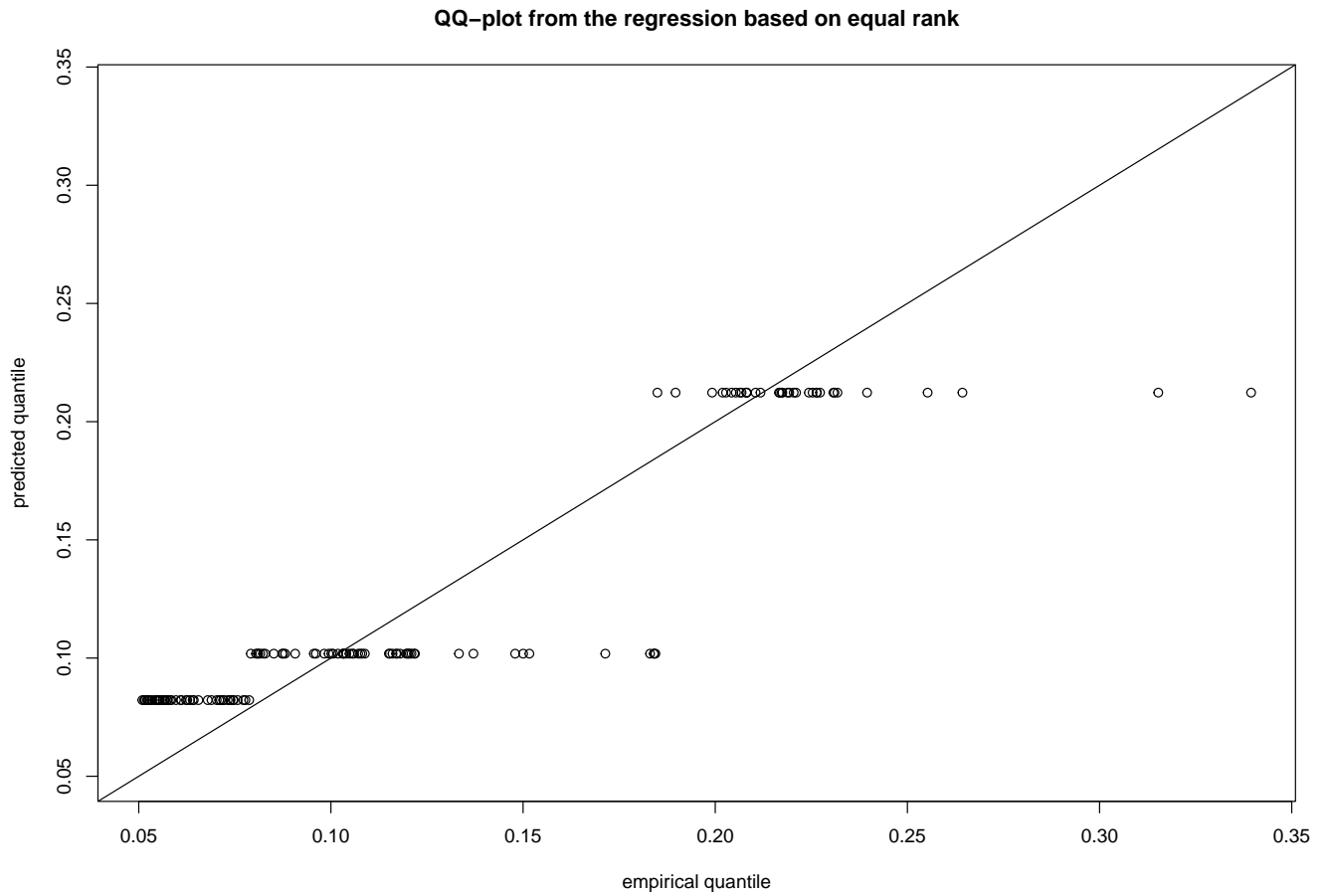


Figure 3.5: The Q-Q plot from the linear regression of the equal rank assumption. The empirical quantile of the strength degradation is against the estimated quantile of the fitted strength degradation. The solid line is the 45-degree line.

3.7 Conclusion

This chapter addresses two key research questions: the correlation between two strength properties in the bivariate scenario and the quantification of damage effect from proof-loading tests in both bivariate and univariate scenarios. To achieve this, we propose a statistical framework applicable to both scenarios, comprising a damage model and a corresponding likelihood ratio test. The damage model incorporates a parameter that directly quantifies the damage effect, allowing both damage and correlation parameters to be estimated together in the bivariate scenario. The likelihood ratio test rigorously evaluates the significance of damage effect in proof-loading groups.

Simulation studies demonstrate the effectiveness of the proposed framework in estimating parameters, particularly the correlation parameter ρ in the bivariate scenario, and accurately detecting significant damage effect in both scenarios. In real data analysis, the framework concludes no significant damage effect in any proof-loading group from the SPF bivariate dataset and provides a precise estimate of $\hat{\rho} = 0.70$ with a much smaller SE (0.05) than the SE (0.2) reported in [Cai \(2015\)](#). For the Hemlock univariate dataset, the likelihood ratio test identifies significant damage effect in the R20_1Y group but finds no apparent damage effect in the R20_4Y or R5_4Y groups. Compared to existing methods such as the DOL model and equal-rank method, the proposed approach directly quantifies damage effect while offering a better fit to the survivor strengths in each group.

Our work contributes to areas such as the reliability analysis of wood products, the design of proof-loading experiments, and the cost reduction of wood strength monitoring. Future research directions include exploring Bayesian inference, which can incorporate prior knowledge of the threshold parameter and better quantify its uncertainty, particularly given the challenges posed by the jagged likelihood for the Hessian method. Additionally, for the univariate scenario, the damage function could be extended to additionally account for pieces that fail during the proof-loading period, which can be interpreted as having their strengths reduced below the proof-loading level.

Chapter 4

Synthetic Empirical Likelihood: Asymptotic Properties and an Extension with Exponential Tilting

4.1 Introduction

The likelihood function is a fundamental component of model-based statistical inference, as it captures the probability of observing a particular set of random variable realizations given specific values of the statistical parameters. While an analytical formula for the likelihood function is typically available for simple models, it can be computationally intractable for more complex models, particularly for those requiring intractable integration over latent variables.

These challenges have led to the development of so-called *likelihood-free* inference methods, which merely assume that it is feasible to simulate data from the forward model (i.e., conditioned on the model parameters). In the absence of a likelihood function, likelihood-free methods approximate the likelihood by using a distance metric to compare observed data and simulated data generated from the model. [Drovandi and Frazier \(2022\)](#) reviewed many such methods. Two widely used approaches are approximate Bayesian computation (ABC) and Bayesian synthetic likelihood (BSL), which differ in their methods of likelihood approximation.

ABC uses nonparametric approaches to approximate the intractable likelihood. The basic idea of ABC is to repeatedly sample parameters from a prior distribution and use

these parameters to simulate data through a known data-generating model. The sampled parameters are then accepted if the observed and simulated data – or more commonly, a problem-dependent set of summary statistics of each – are within a given distance. The method was first introduced by [Pritchard et al. \(1999\)](#) and has since gained popularity due to its ability to handle complex models and its flexibility in accommodating various data types. ABC has been applied in various fields, including medicine ([Drovandi and Pettitt, 2011](#)), engineering ([Yang et al., 2019](#)), ecology ([van der Vaart et al., 2015](#)), biology ([Quach et al., 2009](#)), demography ([Chan et al., 2014](#)), and epidemiology ([Marcondes, 2020](#)).

In contrast, Bayesian synthetic likelihood methods ([Wood, 2010](#); [Price et al., 2018](#)) approximate the likelihood parametrically, assuming the summary statistics of the observed data follow a multivariate normal distribution. The mean and covariance matrix of this multivariate normal distribution are estimated from the simulated data through the data-generating model. Conventionally implemented via Markov chain Monte Carlo (MCMC), BSL methods have been tested successfully and show great potential in a wide range of application areas such as epidemiology and ecology ([Barbu et al., 2018](#); [Price et al., 2018](#)).

However, both ABC and BSL suffer from certain limitations. ABC methods can be computationally intensive and require a large number of simulations, especially for high-dimensional data. To address these issues, several variants of ABC have been introduced, such as Markov Chain Monte Carlo ABC ([Marjoram et al., 2003](#), MCMC-ABC) and sequential Monte Carlo ABC ([Del Moral et al., 2012](#), SMC-ABC), which can improve sampling efficiency. As to BSL methods, the synthetic likelihood approach sometimes performs poorly when the normal approximation of the distribution of the summary statistics is inaccurate. Many extensions of BSL focus on improving its robustness to the normality assumption ([Sisson et al., 2018](#); [An et al., 2020](#); [Priddle et al., 2022](#)).

For models in which the full likelihood is intractable but certain moments involving the parameters of interest are not, empirical likelihood (EL) was suggested by [Mengersen et al. \(2013\)](#) as another method of likelihood-free inference. EL, originally proposed by [Owen \(1988\)](#), only relies on moment conditions involving the parameter of interest, without the need for a parametric likelihood. [Newey and Smith \(2004\)](#) noted that the asymptotic bias of EL remains stable as the number of moment conditions increases, whereas the bias of the generalized method of moments (GMM) often grows, although both are moment-based methods. Bayesian EL (BayesEL), a pseudo-posterior combining EL with prior, has been considered by [Lazar \(2003\)](#), [Rao and Wu \(2010\)](#), [Yang and He \(2012\)](#), [Mengersen et al. \(2013\)](#), and [Chaudhuri et al. \(2017\)](#). [Mengersen et al. \(2013\)](#) first pointed out that BayesEL is much faster than ABC approaches and circumvents the need to simulate data or tune ABC-specified parameters such as distance and tolerance. Nonetheless, BayesEL is limited to cases where moment conditions are analytically available. In addition, the

posterior support of BayesEL can be extremely irregular, requiring highly sophisticated sampling methods to work with, such as sequential sampling (Mengersen et al., 2013) and Hamiltonian Monte Carlo (Chaudhuri et al., 2017).

To address this limitation, Chaudhuri et al. (2024) proposed a likelihood-free method combining appealing properties of ABC/BSL methods and EL. The method, which we refer to as synthetic empirical likelihood (SEL), incorporates simulated data without requiring analytic moment conditions, allowing the use of any summary statistics. By construction, SEL lies somewhere between nonparametric ABC and parametric BSL, and thus stands to be more computationally efficient than the former and more robust to the normality assumption than the latter. However, the original work of Chaudhuri et al. (2024) has notable limitations. First, while it is shown that their proposed SEL estimator is consistent as both the number of (iid) observations N and the number of simulated datasets R increase, no theory for uncertainty quantification is provided. Moreover, the estimator proposed by Chaudhuri et al. (2024) is a Bayesian SEL method which requires samples from the exact pseudo-posterior distribution. Not only does this involve time-consuming MCMC; it also targets an estimator whose frequentist coverage properties deteriorate as R increases.

The limitations above motivate our contributions. First, we develop a framework that leverages the reparametrization trick (Kingma and Welling, 2013) to define the SEL estimator conditional on R sets of noise variables, instead of integrating over all possible draws. This framework allows us to investigate Frequentist and Bayesian SEL estimators through an explicit function of parameters. Second, using the framework, the asymptotic property of the SEL estimator is derived under *just-identification* (equal number of summary statistics and parameters), thereby yielding several accompanying variance estimators. Not only are our estimators significantly faster than conventional MCMC, but also have correct coverage as R increases. Finally, in the case of *over-identification* (more summary statistics than parameters), we study the breakdown of the SEL estimator due to a well-known form of model misspecification (Schennach, 2007). For this setting, we extend SEL with exponential tilting, following the approach from the simulation-free context in Schennach (2005, 2007). Indeed, we demonstrate via examples that our synthetic exponentially-tilted EL (SETEL) estimator outperforms SEL in the over-identified setting.

This chapter is organized as follows. Section 4.2 provides a review of relevant background materials such as EL, BSL, and the SEL method of Chaudhuri et al. (2024). In Section 4.3, we study the asymptotic properties of SEL under just-identification and propose SETEL that empirically retains favorable properties under over-identification. Several methods of standard error calculation are also introduced. Section 4.4 presents simulation studies to evaluate the performance of our proposed methods under various scenarios. Last,

Section 4.5 concludes the chapter.

4.2 Background

4.2.1 Empirical likelihood

Let $\mathbf{x}_{1:N} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ be the observed data or samples from an unknown distribution $F(\mathbf{x})$, where \mathbf{x}_i denotes an N_x -dimensional data vector, and $\boldsymbol{\theta}$ is the p -dimensional parameter vector associated with the distribution F . [Qin and Lawless \(1994\)](#) assumed that information of distribution F and parameter vector $\boldsymbol{\theta}$, is available in an m -dimensional moment condition,

$$\mathbb{E}[\mathbf{g}(\mathbf{x}, \boldsymbol{\theta})] = \mathbf{0}, \quad (4.1)$$

where $\mathbf{g}(\mathbf{x}, \boldsymbol{\theta}) = (g_1(\mathbf{x}, \boldsymbol{\theta}), \dots, g_m(\mathbf{x}, \boldsymbol{\theta}))$, with $p \leq m$. The term ‘moment condition’ is also mentioned as ‘moment constraint’, ‘moment restriction’ in some other literature.

For a given $n \times d$ matrix \mathbf{G} , consider the function \mathcal{C} that maps \mathbf{G} to the $n-1$ dimensional simplex, such that

$$\mathcal{C}(\mathbf{G}) = \{\boldsymbol{\omega} = (\omega_1, \dots, \omega_n) : \sum_{i=1}^n \omega_i = 1, \omega_i \geq 0, \mathbf{G}'\boldsymbol{\omega} = \mathbf{0}\}. \quad (4.2)$$

The empirical likelihood (EL) function, proposed by [Owen \(1988\)](#) and [Qin and Lawless \(1994\)](#), is defined as

$$\text{EL}(\boldsymbol{\theta}) = \prod_{i=1}^N \hat{\omega}_i(\boldsymbol{\theta}), \quad (4.3)$$

where $\hat{\boldsymbol{\omega}}(\boldsymbol{\theta}) = (\hat{\omega}_1(\boldsymbol{\theta}), \dots, \hat{\omega}_N(\boldsymbol{\theta}))$ solves the optimization problem,

$$\hat{\boldsymbol{\omega}}(\boldsymbol{\theta}) = \underset{\boldsymbol{\omega} \in \mathcal{C}(\mathbf{G}_\theta)}{\text{argmax}} \sum_{i=1}^N \log \omega_i, \quad (4.4)$$

where \mathbf{G}_θ is $N \times m$ with rows $\mathbf{g}(\mathbf{x}_1, \boldsymbol{\theta}), \dots, \mathbf{g}(\mathbf{x}_N, \boldsymbol{\theta})$.

Essentially, EL tries to assign weights to the data so that the moment condition is satisfied. Equivalently, the weights in (4.4) can be obtained by solving the unconstrained dual problem ([Owen, 2001](#)),

$$\hat{\boldsymbol{\omega}}(\boldsymbol{\theta}) = \frac{1}{N} \cdot \frac{1}{1 + \mathbf{G}_\theta \hat{\boldsymbol{\lambda}}(\boldsymbol{\theta})}, \quad (4.5)$$

where

$$\hat{\lambda}(\boldsymbol{\theta}) = \operatorname{argmax}_{\lambda} \mathbf{1}' \log^*(1 - \mathbf{G}_{\boldsymbol{\theta}} \hat{\lambda}(\boldsymbol{\theta})), \quad (4.6)$$

and

$$\log^*(x) = \begin{cases} \log(x) & x \geq \frac{1}{N} \\ -\frac{1}{2}N^2x^2 + 2Nx - \frac{3}{2} - \log(N) & x < \frac{1}{N}. \end{cases} \quad (4.7)$$

Note that Equation (4.6) is a convex optimization problem, which can readily be solved using the Newton-Raphson algorithm (when the solution exists). Unless explicitly stated otherwise, we use the convention that functions defined as scalar-to-scalar apply element-wise to vectors, matrices, etc.

The EL estimator is defined as

$$\hat{\boldsymbol{\theta}}_{\text{EL}} = \operatorname{argmax}_{\boldsymbol{\theta}} \text{EL}(\boldsymbol{\theta}). \quad (4.8)$$

Qin and Lawless (1994) derived the asymptotic property of EL estimator. With true value $\boldsymbol{\theta}_0$, under standard regularity conditions, we have

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{\text{EL}} - \boldsymbol{\theta}_0) \xrightarrow{d} \text{Normal}(\mathbf{0}, \mathbf{V}_{\text{EL}}), \quad (4.9)$$

where

$$\mathbf{V}_{\text{EL}} = \left[\text{E} \left(\frac{\partial \mathbf{g}(\mathbf{x}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right)' \text{E} (\mathbf{g}(\mathbf{x}, \boldsymbol{\theta}_0) \mathbf{g}(\mathbf{x}, \boldsymbol{\theta}_0)')^{-1} \text{E} \left(\frac{\partial \mathbf{g}(\mathbf{x}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right) \right]^{-1}. \quad (4.10)$$

4.2.2 Support adjustment

One of the challenges in EL is that, in some cases, the optimization problem (4.4), or equivalently, (4.6), may have no solution, resulting in the so-called *empty-set* problem. To solve (4.4), a prerequisite is that the convex hull of $\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}), i = 1, \dots, N$ must contain the zero vector $\mathbf{0}$ as an interior point. If the prerequisite cannot be met for the given parameter value $\boldsymbol{\theta}$, we define $\text{EL}(\boldsymbol{\theta}) = 0$. The support of EL is $\{\boldsymbol{\theta} \in \mathbb{R}^p : \text{EL}(\boldsymbol{\theta}) > 0\}$, where \mathbb{R}^p is the parameter space.

This prerequisite may not always be satisfied, particularly when the parameter value is far from the mode of the EL function, leading to considerable difficulties with initializing mode-finding algorithms for computing the EL estimator (4.8). The EL parameter support issue has been studied by many researchers. For instance, Chen et al. (2008), Emerson and

Owen (2009), and Liu and Liang (2022) proposed adapted likelihood methods that easily satisfy the condition for the existence of a solution while retaining the critical asymptotic properties of the original EL method. In this chapter, we briefly introduce the adjusted empirical likelihood (AEL), which implements support adjustment to EL (i.e., adjust the support of EL), proposed in Chen et al. (2008).

By adding a pseudo-observation to the EL inner optimization problem (4.4), the AEL function is

$$\text{AEL}(\boldsymbol{\theta}) = \prod_{i=1}^{N+1} \hat{\omega}_i(\boldsymbol{\theta}), \quad (4.11)$$

where $\hat{\boldsymbol{\omega}}(\boldsymbol{\theta}) = (\hat{\omega}_1(\boldsymbol{\theta}), \dots, \hat{\omega}_N(\boldsymbol{\theta}), \hat{\omega}_{N+1}(\boldsymbol{\theta}))$ solves the optimization problem

$$\hat{\boldsymbol{\omega}}(\boldsymbol{\theta}) = \operatorname{argmax}_{\boldsymbol{\omega} \in \mathcal{C}(\mathbf{G}_{\boldsymbol{\theta}}^+)} \sum_{i=1}^{N+1} \log \omega_i, \quad (4.12)$$

where $\mathbf{G}_{\boldsymbol{\theta}}^+$ is $(N+1) \times m$ with rows $\mathbf{g}(\mathbf{x}_1, \boldsymbol{\theta}), \dots, \mathbf{g}(\mathbf{x}_N, \boldsymbol{\theta}), -\frac{a_N}{N} \sum_{i=1}^N \mathbf{g}_i(\mathbf{x}_i, \boldsymbol{\theta})$.

In other words, $\mathbf{G}_{\boldsymbol{\theta}}^+$ of AEL has an additional row $-\frac{a_N}{N} \sum_{i=1}^N \mathbf{g}_i(\mathbf{x}_i, \boldsymbol{\theta})$, compared to $\mathbf{G}_{\boldsymbol{\theta}}$ of EL. In application, Chen et al. (2008) recommended that $a_N = \max(1, \log(N)/2)$ so that when the sample size N is small, the support adjustment effectively improves the coverage probability of the confidence region. The upshot is that the AEL optimization (4.12) satisfies the prerequisite for any given value of $\boldsymbol{\theta}$. Chen et al. (2008) proved that the AEL function behaves similarly to the original EL function around the mode.

4.2.3 Bayesian EL

In the Bayesian paradigm, the Bayesian empirical likelihood (BayesEL) method combines the prior information on parameters and the data-driven likelihood to draw inference on the estimand based on moment conditions. One of the notable advantages of BayesEL is that it retains the desirable properties of frequentist EL while also allowing for the incorporation of random effects (e.g., Chaudhuri and Ghosh, 2011; Chaudhuri et al., 2017) and prior information. The BayesEL pseudo-posterior based on the empirical likelihood $\text{EL}(\boldsymbol{\theta})$ is defined as:

$$p_{\text{EL}}(\boldsymbol{\theta} \mid \mathbf{x}_{1:N}) = \frac{\text{EL}(\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int \text{EL}(\boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\boldsymbol{\theta}} \propto \text{EL}(\boldsymbol{\theta}) \cdot \pi(\boldsymbol{\theta}), \quad (4.13)$$

where $\pi(\boldsymbol{\theta})$ is the prior of $\boldsymbol{\theta}$.

In Bayesian analysis, various tools are available for posterior inference. One widely used method is Markov Chain Monte Carlo (MCMC). Another common approach is Bayesian normal approximation (BNA, e.g., [Gelman et al. 2013](#)). The basic idea of BNA is to use a normal distribution to approximate the posterior, which significantly reduces computational complexity and improves inference efficiency.

To obtain such normal distribution, the BNA relies on a second-order Taylor expansion around the mode of the posterior, which is the BayesEL pseudo-posterior (4.13) in BayesEL. It gives

$$\boldsymbol{\theta} \mid \mathbf{x}_{1:N} \sim \text{Normal}(\hat{\boldsymbol{\theta}}_{\text{BEL}}, \left[-\text{H}(\hat{\boldsymbol{\theta}}_{\text{BEL}})\right]^{-1}), \quad (4.14)$$

where $\hat{\boldsymbol{\theta}}_{\text{BEL}}$ is the mode of $\log p_{\text{EL}}(\boldsymbol{\theta} \mid \mathbf{x}_{1:N})$, and $\text{H}(\boldsymbol{\theta})$, based on the Hessian matrix of the BayesEL pseudo-posterior, is given by

$$\text{H}(\boldsymbol{\theta}) = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \log p_{\text{EL}}(\boldsymbol{\theta} \mid \mathbf{x}_{1:N}). \quad (4.15)$$

The BNA for BayesEL is asymptotically justified via the Bernstein-von Mises theorem. Specifically, under both BayesEL and its BNA, the distribution of $\sqrt{N}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ converge to a normal distribution with mean $\mathbf{0}$ and variance \mathbf{V}_{EL} ([Lazar, 2003](#); [Sueishi, 2022](#)).

For BayesEL, the empty-set issue can induce a highly irregular parameter support region, making it very difficult to design effective MCMC algorithms ([Mengersen et al., 2013](#); [Chaudhuri et al., 2017](#)). To the best of our knowledge, the implementation of support adjustment in BayesEL has not been noted in any literature. The following example illustrates an important caveat in such scenario.

Consider observations y_1, \dots, y_n such that $y_i \stackrel{\text{ind}}{\sim} \text{Binomial}(m_i, \rho_i)$ and

$$\text{logit } \rho_i = \beta_1 + \beta_2 \cdot x_i, \quad (4.16)$$

where x_i is a covariate, (β_1, β_2) are unknown parameters. As an alternative to estimating (β_1, β_2) using the true likelihood, instead, we consider an empirical likelihood approach based on the Bartlett moment conditions

$$E[y_i - m_i \rho_i] = 0, \quad E\left[\frac{(y_i - m_i \rho_i)^2}{m_i \rho_i (1 - \rho_i)} - 1\right] = 0. \quad (4.17)$$

Then in this problem, the EL inner optimization is constructed by incorporating (4.17) into (4.4). To provide a numerical illustration, we generate $N = 25$ observations with

$m_i = 4$, $x_i \sim N(0, 1)$, and $(\beta_1, \beta_2) = (1, 2)$. BayesEL inference is then conducted using the prior of $\beta_1, \beta_2 \stackrel{\text{iid}}{\sim} N(0, 10^2)$.

Figure 4.1 presents the (pseudo) posterior densities of β_1 and β_2 using five methods: true posterior, (exact) unadjusted EL, (exact) adjusted EL, BNA of unadjusted EL, BNA of adjusted EL. The true posterior combines the parametric likelihood and the prior. The first three posteriors are obtained via grid approximation (Gelman et al., 2013). This figure reveals several key insights. First, the unadjusted and adjusted EL pseudo-posteriors differ substantially across their entire support. Second, both BNA methods yield results that closely resemble the unadjusted EL and the true posterior, regardless of whether an adjustment is applied, suggesting that EL adjustments have minimal impact when using BNA. In contrast, the exact AEL pseudo-posterior places significant mass away from the correct mode, indicating it should be used with caution. Finally, given that the adjustment alters the posterior across its entire support, MCMC methods should be restricted to the unadjusted EL.

4.2.4 Bayesian synthetic likelihood

Synthetic likelihood (SL) methods (Wood, 2010; Price et al., 2018; Frazier et al., 2023), similar to ABC, are popular likelihood-free approaches, used when the likelihood is intractable but it is easy to simulate from the model.

Suppose the observed data $\mathbf{x}_{1:N} \sim F(\mathbf{x}_{1:N} \mid \boldsymbol{\theta})$ is generated from a parametric model $F(\mathbf{x}_{1:N} \mid \boldsymbol{\theta})$, from which we can sample for any $\boldsymbol{\theta}$ but cannot evaluate the likelihood in closed form. The m -dimensional function of the summary statistics is denoted as $\mathbf{s}(\mathbf{x}_{1:N}) = (s_1(\mathbf{x}_{1:N}), \dots, s_m(\mathbf{x}_{1:N}))$.

SL methods assume that N is large enough for a central limit theorem to apply. In other words, the summary statistics of the observed data, referred to as the observed summary statistics and denoted by $\mathbf{s}_N = \mathbf{s}(\mathbf{x}_{1:N})$, are assumed to follow a multivariate normal distribution, that is

$$\mathbf{s}_N \sim \text{Normal}(\boldsymbol{\mu}_N(\boldsymbol{\theta}), \boldsymbol{\Sigma}_N(\boldsymbol{\theta})), \quad (4.18)$$

where $\boldsymbol{\mu}_N(\boldsymbol{\theta}) = \text{E}[\mathbf{s}_N \mid \boldsymbol{\theta}]$ and $\boldsymbol{\Sigma}_N(\boldsymbol{\theta}) = \text{var}(\mathbf{s}_N \mid \boldsymbol{\theta})$.

As $\boldsymbol{\mu}_N(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_N(\boldsymbol{\theta})$ are generally not available in closed forms, they are estimated via simulations. That is, for given $\boldsymbol{\theta}$, let $\mathbf{x}_{1:N}^{(1:R)}(\boldsymbol{\theta}) = (\mathbf{x}_{1:N}^{(1)}(\boldsymbol{\theta}), \dots, \mathbf{x}_{1:N}^{(R)}(\boldsymbol{\theta}))$ denote R iid (independent and identically distributed) simulated datasets from the parametric model where $\mathbf{x}_{1:N}^{(r)}(\boldsymbol{\theta}) \stackrel{\text{iid}}{\sim} F(\mathbf{x}_{1:N} \mid \boldsymbol{\theta})$, $r = 1, \dots, R$. The corresponding summary statistics,

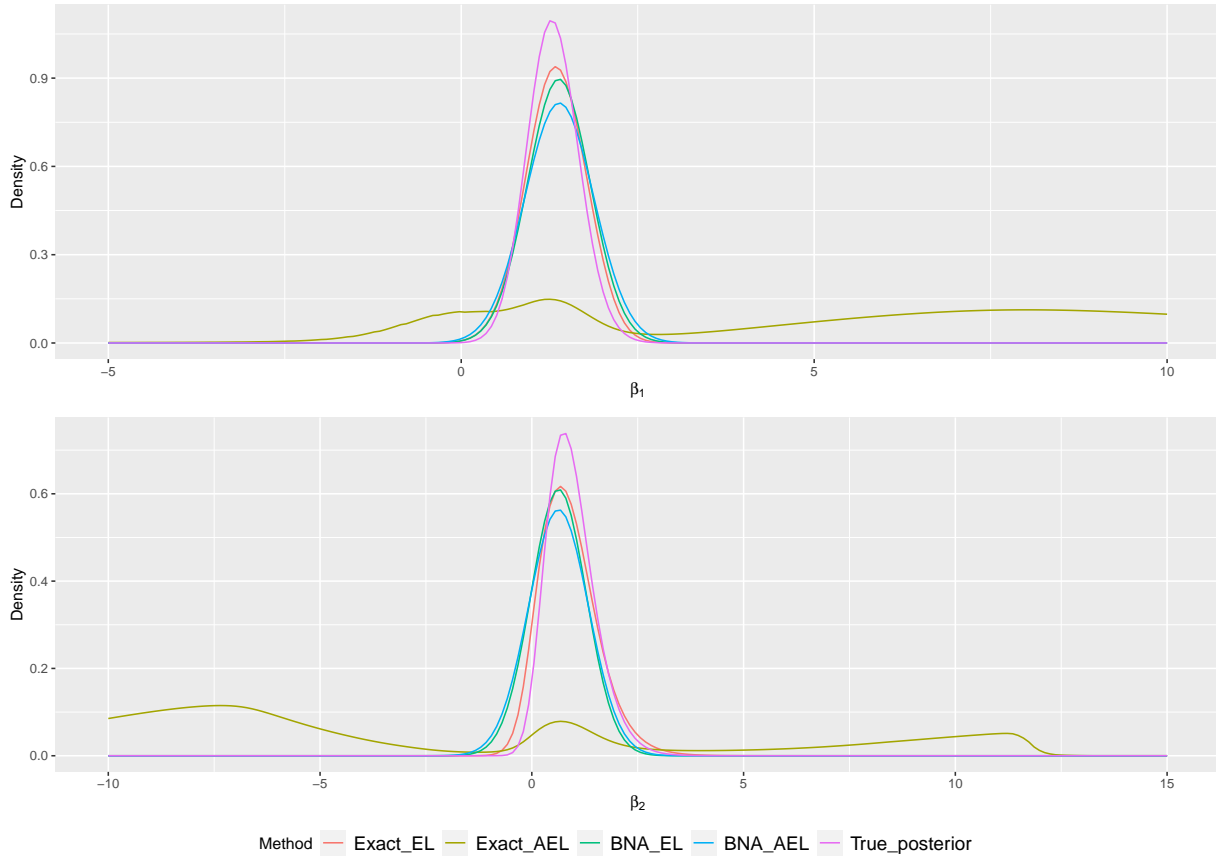


Figure 4.1: Marginal posterior densities of β_1, β_2 in binomial logistic regression model by the true posterior, (exact) EL posterior, (exact) AEL posterior, BNA of EL posterior, BNA of AEL posterior, and the true posterior.

referred to as the simulated summary statistics, are $\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}) = (\mathbf{s}_N^{(1)}(\boldsymbol{\theta}), \dots, \mathbf{s}_N^{(R)}(\boldsymbol{\theta}))$, where $\mathbf{s}_N^{(r)}(\boldsymbol{\theta}) = \mathbf{s}(\mathbf{x}_{1:N}^{(r)}(\boldsymbol{\theta}))$.

Then $\boldsymbol{\mu}_N(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_N(\boldsymbol{\theta})$ are estimated by $\hat{\boldsymbol{\mu}}_N^R(\boldsymbol{\theta})$ and $\hat{\boldsymbol{\Sigma}}_N^R(\boldsymbol{\theta})$, which are

$$\begin{aligned}\hat{\boldsymbol{\mu}}_N^R(\boldsymbol{\theta}) &= \frac{1}{R} \sum_{r=1}^R \mathbf{s}_N^{(r)}(\boldsymbol{\theta}), \\ \hat{\boldsymbol{\Sigma}}_N^R(\boldsymbol{\theta}) &= \frac{1}{R-1} \sum_{r=1}^R (\mathbf{s}_N^{(r)}(\boldsymbol{\theta}) - \hat{\boldsymbol{\mu}}_N^R(\boldsymbol{\theta}))(\mathbf{s}_N^{(r)}(\boldsymbol{\theta}) - \hat{\boldsymbol{\mu}}_N^R(\boldsymbol{\theta}))'.\end{aligned}\quad (4.19)$$

Given the observed summary statistics \mathbf{s}_N , the parameter $\boldsymbol{\theta}$, and the simulated summary statistics $\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta})$, let

$$\text{SL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta})) = \text{Normal}(\mathbf{s}_N; \hat{\boldsymbol{\mu}}_N^R(\boldsymbol{\theta}), \hat{\boldsymbol{\Sigma}}_N^R(\boldsymbol{\theta})), \quad (4.20)$$

where $\text{Normal}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density function of multivariate normal at point \mathbf{x} with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$. Then, the unknown distribution of \mathbf{s}_N , $p(\mathbf{s}_N | \boldsymbol{\theta})$, can be approximated by

$$p_{\text{SL}}(\mathbf{s}_N | \boldsymbol{\theta}) = \int \text{SL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta})) p(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}) | \boldsymbol{\theta}) d\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}). \quad (4.21)$$

The Bayesian SL (BSL) posterior (Price et al., 2018; Frazier et al., 2023) takes the form

$$p_{\text{SL}}(\boldsymbol{\theta} | \mathbf{s}_N) \propto p_{\text{SL}}(\mathbf{s}_N | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}), \quad (4.22)$$

where $\pi(\boldsymbol{\theta})$ is the prior of $\boldsymbol{\theta}$. Price et al. (2018) developed a pseudo-marginal MCMC algorithm (Andrieu and Roberts, 2009), of which the stationary distribution is the BSL posterior (4.22), as outlined in Algorithm 1. We call this method BSL-MCMC.

4.2.5 Synthetic empirical likelihood

Chaudhuri et al. (2024) proposed a likelihood-free method based on EL. With the same notation in SL, we further assume that the observed data $\mathbf{x}_{1:N}$ is generated from the parametric model $\mathbf{x}_{1:N} \sim F(\mathbf{x}_{1:N} | \boldsymbol{\theta} = \boldsymbol{\theta}_0)$ at the true parameter value $\boldsymbol{\theta}_0$.

When the given parameter equals the true value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, the construction of the simulated dataset $\mathbf{x}_{1:N}^{(1:R)}(\boldsymbol{\theta})$ indicates that $\mathbf{x}_{1:N}, \mathbf{x}_{1:N}^{(r)}(\boldsymbol{\theta}_0)$ are identically distributed. And thus we have moment conditions,

$$\text{E}[\mathbf{s}_N^{(r)}(\boldsymbol{\theta}_0) - \mathbf{s}_N] = \mathbf{0}, \quad (4.23)$$

Algorithm 1 BSL-MCMC algorithm

Input: the observed data $\mathbf{x}_{1:N}$, a prior $\pi(\boldsymbol{\theta})$, an initial value $\boldsymbol{\theta}_0$, a proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$, a number of iterations T

- 1: Calculate the observed summary statistics $\mathbf{s}_N = \mathbf{s}(\mathbf{x}_{1:N})$
- 2: Initialize the term $\text{SL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}_0))\pi(\boldsymbol{\theta}_0)$
- 3: **for** $i = 1$ to T **do**
- 4: Draw a candidate $\boldsymbol{\theta}^*$ from the proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{i-1})$
- 5: Simulate R datasets $\mathbf{x}_{1:N}^{(1:R)}(\boldsymbol{\theta}^*) = (\mathbf{x}_{1:N}^{(1)}(\boldsymbol{\theta}^*), \dots, \mathbf{x}_{1:N}^{(R)}(\boldsymbol{\theta}^*))$
- 6: Calculate the simulated summary statistics, $\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}^*) = (\mathbf{s}(\mathbf{x}_{1:N}^{(1)}), \dots, \mathbf{s}(\mathbf{x}_{1:N}^{(R)}))$
- 7: Calculate the SL function $\text{SL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}^*))$ based on the observed and simulated summary statistics
- 8: Calculate the acceptance ratio,

$$\alpha = \min \left(1, \frac{\text{SL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}^*))\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^{i-1}|\boldsymbol{\theta}^*)}{\text{SL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}^{i-1}))\pi(\boldsymbol{\theta}^{i-1})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{i-1})} \right)$$

- 9: Draw $u \sim \text{Uniform}(0, 1)$
 - 10: **if** $u < \alpha$ **then**
 - 11: Accept the candidate and set $\boldsymbol{\theta}^i = \boldsymbol{\theta}^*$
 - 12: Save $\text{SL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}^*))\pi(\boldsymbol{\theta}^*)$ as $\text{SL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}^i))\pi(\boldsymbol{\theta}^i)$ for the next iteration
 - 13: **else**
 - 14: Reject the candidate and set $\boldsymbol{\theta}^i = \boldsymbol{\theta}^{i-1}$
 - 15: Retain $\text{SL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}^{i-1}))\pi(\boldsymbol{\theta}^{i-1})$ as $\text{SL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}^i))\pi(\boldsymbol{\theta}^i)$ for the next iteration
 - 16: **end if**
 - 17: **end for**
- Output:** A sequence of samples $\{\boldsymbol{\theta}^i\}_{i=1}^T$
-

$r = 1, \dots, R$.

Then based on the moment conditions (4.23), Chaudhuri et al. (2024) proposed what we shall refer to as the *synthetic empirical likelihood* (SEL) function,

$$\text{SEL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta})) = \exp \left[\frac{1}{R} \cdot \sum_{i=1}^R \log \hat{\omega}_i(\boldsymbol{\theta}) \right], \quad (4.24)$$

where $\hat{\boldsymbol{\omega}}(\boldsymbol{\theta}) = (\hat{\omega}_1(\boldsymbol{\theta}), \dots, \hat{\omega}_R(\boldsymbol{\theta}))$ solves the optimization problem,

$$\hat{\boldsymbol{\omega}}(\boldsymbol{\theta}) = \underset{\boldsymbol{\omega} \in \mathcal{C}(\mathbf{G}_\theta)}{\text{argmax}} \sum_{r=1}^R \log \omega_r, \quad (4.25)$$

where \mathbf{G}_θ is $R \times m$ with rows $\mathbf{s}_N^{(1)}(\boldsymbol{\theta}) - \mathbf{s}_N, \dots, \mathbf{s}_N^{(R)}(\boldsymbol{\theta}) - \mathbf{s}_N$.

Similar to BSL, $p(\mathbf{s}_N | \boldsymbol{\theta})$ can be approximated by

$$p_{\text{SEL}}(\mathbf{s}_N | \boldsymbol{\theta}) = \int \text{SEL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta})) p(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}) | \boldsymbol{\theta}) d\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}). \quad (4.26)$$

Unlike $p_{\text{SL}}(\mathbf{s}_N | \boldsymbol{\theta})$, which is a probability distribution, $p_{\text{SEL}}(\mathbf{s}_N | \boldsymbol{\theta})$ does not define a valid probability distribution. Therefore, we use the terms pseudo-likelihood and pseudo-posterior in SEL.

Then, the Bayesian SEL pseudo-posterior takes the form

$$p_{\text{SEL}}(\boldsymbol{\theta} | \mathbf{s}_N) \propto p_{\text{SEL}}(\mathbf{s}_N | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}). \quad (4.27)$$

Since $p_{\text{SEL}}(\mathbf{s}_N | \boldsymbol{\theta})$ cannot be calculated in closed form, Chaudhuri et al. (2024) also employed the pseudo-marginal MCMC to explore the SEL pseudo-posterior (4.27), by replacing $\text{SL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}))$ with $\text{SEL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}))$ in Algorithm 1. Chaudhuri et al. (2024) used the term *ABCel* to refer both to the SEL likelihood approximation (4.26) and the pseudo-marginal MCMC algorithm used to perform full Bayesian inference on $p_{\text{SEL}}(\boldsymbol{\theta} | \mathbf{s}_N)$. In contrast, we draw a distinction here between the likelihood approximation and the resulting estimator, referring to the MCMC approach of Chaudhuri et al. (2024) as SEL-MCMC.

Theorem 4 in Chaudhuri et al. (2024) demonstrated that the SEL pseudo-posterior converges in probability to a point mass at $\boldsymbol{\theta}_0$ as $N \rightarrow \infty$, but left the asymptotic variance of the posterior distribution unexplored.

4.3 Methodology

Chaudhuri et al. (2024) used MCMC to sample from full SEL pseudo-posterior. However, for the SEL pseudo-posterior, MCMC is not only time-consuming but may also exhibit deteriorating coverage properties as R increases. Specifically, as R grows, the SEL pseudo-posterior becomes flatter, leading to higher variance, which is considered an undesirable property. An example for graphical illustration is provided in Section 4.4.1. The theoretical investigation for this phenomenon can be found in Section 4.2 of Chaudhuri et al. (2024).

In addition to the drawbacks from MCMC, the function $\text{SEL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}))$, introduced in Chaudhuri et al. (2024), depends on $\boldsymbol{\theta}$ implicitly, through the simulated value $\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta})$. We refer to it as the implicit SEL function and it may pose challenges such as differentiation or optimization involving this function with respect to $\boldsymbol{\theta}$.

The issues above motivate us to develop new methods to better quantify uncertainties in SEL methods. We consider the *reparametrization trick* method, proposed by Kingma and Welling (2013). Suppose that we can reparametrize the model $\mathbf{x}_{1:N}(\boldsymbol{\theta}) \sim F(\mathbf{x}_{1:N} | \boldsymbol{\theta})$ such that

$$\mathbf{x}_{1:N}(\boldsymbol{\theta}) = h(\boldsymbol{\varepsilon}_{1:N}, \boldsymbol{\theta}), \quad \boldsymbol{\varepsilon}_{1:N} \sim p(\boldsymbol{\varepsilon}_{1:N}), \quad (4.28)$$

where $\boldsymbol{\varepsilon}_{1:N} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_N)$ is a noise variable (or noise term), whose distribution $p(\boldsymbol{\varepsilon}_{1:N})$ does not depend on $\boldsymbol{\theta}$. By pre-generating R iid sets of noise variables $\boldsymbol{\varepsilon}_{1:N}^{(1:R)} = (\boldsymbol{\varepsilon}_{1:N}^{(1)}, \dots, \boldsymbol{\varepsilon}_{1:N}^{(R)})$, where $\boldsymbol{\varepsilon}_{1:N}^{(r)} \stackrel{\text{iid}}{\sim} p(\boldsymbol{\varepsilon}_{1:N})$, $r = 1, \dots, R$, we transform the implicit SEL function $\text{SEL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}))$ into the explicit SEL function $\text{SEL}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)})$, which explicitly depends on $\boldsymbol{\theta}$. For fixed observed summary statistics \mathbf{s}_N , the moment condition in SEL is then expressed as $\text{E}[\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \boldsymbol{\theta})] - \mathbf{s}_N = \mathbf{0}$, where $\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \boldsymbol{\theta})$ is the reparametrized form of $\mathbf{s}_N^{(r)}(\boldsymbol{\theta})$.

Our proposed Bayesian SEL pseudo-posterior is then formulated as

$$p_{\text{SEL}}(\boldsymbol{\theta} | \mathbf{s}_N, \boldsymbol{\varepsilon}_{1:N}^{(1:R)}) \propto \text{SEL}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)})\pi(\boldsymbol{\theta}). \quad (4.29)$$

We emphasize that the main difference between our proposed explicit pseudo-posterior $p_{\text{SEL}}(\boldsymbol{\theta} | \mathbf{s}_N, \boldsymbol{\varepsilon}_{1:N}^{(1:R)})$ and the implicit pseudo-posterior $p_{\text{SEL}}(\boldsymbol{\theta} | \mathbf{s}_N)$ introduced by Chaudhuri et al. (2024), lies in the approach to likelihood approximation. Our likelihood approximation, $\text{SEL}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)})$, is evaluated using a fixed value of $\boldsymbol{\varepsilon}_{1:N}^{(1:R)}$. In contrast, the likelihood approximation in Chaudhuri et al. (2024) takes the form $\int \text{SEL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}))p(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}) | \boldsymbol{\theta}) d\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta})$, which integrates over all possible values of the noise variable.

Conditional on the noise variables, our approach naturally induces a BNA estimator for the pseudo-posterior (4.29). Analogously, in the frequentist framework, the SEL estimator

is defined as

$$\hat{\boldsymbol{\theta}}_{\text{SEL}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \operatorname{SEL}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)}). \quad (4.30)$$

Importantly, both the BNA estimator and the frequentist SEL estimator can be obtained by solving a deterministic optimization problem, which is orders of magnitude faster than MCMC which requires integrating over $\boldsymbol{\varepsilon}_{1:N}^{(1:R)}$.

Similar rules apply to SL by converting the implicit SL function $\operatorname{SL}(\mathbf{s}_N^{(1:R)}(\boldsymbol{\theta}))$ to the explicit SL function $\operatorname{SL}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)})$ using the reparametrization trick, which yields our proposed BSL posterior and frequentist SL estimator for a fixed value of $\boldsymbol{\varepsilon}_{1:N}^{(1:R)}$. Unless stated otherwise, all references to the SEL (SL) function hereafter in this chapter refer to the explicit SEL (SL) function based on the reparametrization trick.

4.3.1 Just-identified moment condition

Section 4.2.5 introduces the moment condition of SEL where \mathbf{s}_N is treated as random. However, when working with real data, N is fixed, along with $\mathbf{x}_{1:N}$ and \mathbf{s}_N . When \mathbf{s}_N is fixed, SEL reduces to a classical EL problem for iid noise draws $\boldsymbol{\varepsilon}_{1:N}^{(r)} \stackrel{\text{iid}}{\sim} p(\boldsymbol{\varepsilon}_{1:N})$, $r = 1, \dots, R$. In this case, we distinguish between two different settings: (1) when $m = p$, termed just-identification, and (2) when $m > p$, termed over-identification.

In this subsection, within the frequentist framework, we explore the asymptotic properties of $\hat{\boldsymbol{\theta}}_{\text{SEL}}$ under just-identified moment conditions, with the following assumptions.

Assumption 1. *The data-generating mechanism is $\mathbf{x}_i(\boldsymbol{\theta}) \stackrel{\text{iid}}{\sim} F(\mathbf{x}|\boldsymbol{\theta})$, with $\mathbf{x}_i(\boldsymbol{\theta}) = \mathbf{h}(\boldsymbol{\varepsilon}_i, \boldsymbol{\theta})$ and $\boldsymbol{\varepsilon}_i \stackrel{\text{iid}}{\sim} p(\boldsymbol{\varepsilon})$.*

Assumption 2. *The summary statistics are formed as the sample mean of the transformed data, i.e., $\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}, \boldsymbol{\theta}) = (N^{-1} \sum_{i=1}^N f_1(\mathbf{h}(\boldsymbol{\varepsilon}_i, \boldsymbol{\theta})), \dots, N^{-1} \sum_{i=1}^N f_m(\mathbf{h}(\boldsymbol{\varepsilon}_i, \boldsymbol{\theta})))$.*

Assumption 3. *For any fixed and observable value of the summary statistic, $\mathbf{s}_N = \mathbf{s}(\boldsymbol{\varepsilon}_{1:N}, \boldsymbol{\theta}_0)$, there exists a unique $\tilde{\boldsymbol{\theta}}$, the idealized SEL estimator, that satisfies*

$$E \left[\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \tilde{\boldsymbol{\theta}}) \right] - \mathbf{s}_N = \mathbf{0}. \quad (4.31)$$

Assumption 4. *For any value of $\boldsymbol{\theta}$, we have the central limit theorem*

$$\sqrt{N}(\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}, \boldsymbol{\theta}) - \boldsymbol{\mu}(\boldsymbol{\theta})) \xrightarrow{d} \operatorname{Normal}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta})), \quad (4.32)$$

where $\boldsymbol{\mu}(\boldsymbol{\theta}) = E[\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}, \boldsymbol{\theta})]$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = N \cdot \operatorname{Var}(\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}, \boldsymbol{\theta}))$.

Assumption 5. For any $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$, (i) the function $\boldsymbol{\mu}(\boldsymbol{\theta})$ is a continuous and bijective function; (ii) the derivative $\nabla \boldsymbol{\mu}(\boldsymbol{\theta}) = \partial \boldsymbol{\mu}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ exists and is continuous and invertible.

Assumptions 1–2 are introduced to simplify the asymptotic problem. However, it is important to note that Theorem 1 below is not necessarily compromised if Assumptions 1–2 are not satisfied. Under Assumptions 1–2, $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ are independent of N , which facilitate Assumption 4. Based on the assumptions, we obtain the asymptotic property of the SEL estimator under just identification, as stated in Theorem 1.

Theorem 1. If Assumptions 1–5 are satisfied, let $\mathbf{Z} \sim \text{Normal}(\mathbf{0}, \mathbf{V}_{EL})$, with $\mathbf{g}(\mathbf{x}, \boldsymbol{\theta})$ from \mathbf{V}_{EL} (4.10) given by $\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}, \boldsymbol{\theta}) - \mathbf{s}_N$. Then for B a continuity set of \mathbf{Z} , and any $\epsilon > 0$, there exist N_0 and a function $R(N)$ of N , such that for all $N > N_0$, $R > R(N)$,

$$\left| P\left(\sqrt{N}(\hat{\boldsymbol{\theta}}_{SEL} - \boldsymbol{\theta}_0) \in B\right) - P(\mathbf{Z} \in B) \right| < \epsilon. \quad (4.33)$$

Proof. The moment condition in SEL defines that $E\left[\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \tilde{\boldsymbol{\theta}})\right] - \mathbf{s}_N = \mathbf{0}$, where $\tilde{\boldsymbol{\theta}}$ is the idealized estimator of $\boldsymbol{\theta}_0$ and $\boldsymbol{\varepsilon}_{1:N}^{(r)}$ is the simulated noise. Thus we have $\mathbf{s}_N = \mathbf{s}(\boldsymbol{\varepsilon}_{1:N}, \boldsymbol{\theta}_0) = E[\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \tilde{\boldsymbol{\theta}})] = \boldsymbol{\mu}(\tilde{\boldsymbol{\theta}})$, yielding $\tilde{\boldsymbol{\theta}} = \boldsymbol{\mu}^{-1}(\mathbf{s}_N)$. As $\hat{\boldsymbol{\theta}}_{SEL}$ is a deterministic function of $\boldsymbol{\varepsilon}_{1:N}^{(1:R)}$ and \mathbf{s}_N , we can write it as $\hat{\boldsymbol{\theta}}_{SEL} = \hat{\boldsymbol{\theta}}(\boldsymbol{\varepsilon}_{1:N}^{(1:R)}, \mathbf{s}_N)$.

Note that for any fixed N and \mathbf{s}_N , the proposed SEL estimator $\hat{\boldsymbol{\theta}}_{SEL}$ is just a classical EL estimator about the parameter value $\tilde{\boldsymbol{\theta}} = \boldsymbol{\mu}^{-1}(\mathbf{s}_N)$. As $R \rightarrow \infty$, the EL convergence property (Qin and Lawless, 1994; Liang and Chen, 2023) gives

$$\hat{\boldsymbol{\theta}}_{SEL} \xrightarrow{a.s.} \tilde{\boldsymbol{\theta}}. \quad (4.34)$$

This means

$$P\left(\omega \in \Omega : \lim_{R \rightarrow \infty} \hat{\boldsymbol{\theta}}(\boldsymbol{\varepsilon}_{1:N}^{(1:R)}(\omega), \boldsymbol{\mu}(\tilde{\boldsymbol{\theta}})) = \tilde{\boldsymbol{\theta}}\right) = 1, \quad (4.35)$$

where ω represents any possible random infinite sequence $\left\{\boldsymbol{\varepsilon}_{1:N}^{(1:R)}\right\}_{R=1}^{\infty}$, and Ω is the sample space of all such sequences. Note that Eq. (4.35) holds for any random \mathbf{s}_N , which implies

$$P\left((\omega, \pi) \in (\Omega, \Pi) : \lim_{R \rightarrow \infty} \hat{\boldsymbol{\theta}}(\boldsymbol{\varepsilon}_{1:N}^{(1:R)}(\omega), \boldsymbol{\mu}(\tilde{\boldsymbol{\theta}}(\pi))) = \tilde{\boldsymbol{\theta}}(\pi)\right) = 1 \quad (4.36)$$

as $R \rightarrow \infty$ for any random \mathbf{s}_N , where π is any possible value of \mathbf{s}_N and Π is the sample space of all such value.

On the other hand, treating $\mathbf{s}_N = \boldsymbol{\mu}(\tilde{\boldsymbol{\theta}}) = \mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \boldsymbol{\theta}_0)$ as a random variable and plugging it into (4.32) yields

$$\sqrt{N}(\mathbf{s}_N - \boldsymbol{\mu}(\boldsymbol{\theta}_0)) \xrightarrow{d} \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)).$$

We may apply the delta method to get the asymptotic distribution of $\tilde{\boldsymbol{\theta}} = \boldsymbol{\mu}^{-1}(\mathbf{s}_N)$, which is

$$\sqrt{N}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \text{Normal}(\mathbf{0}, \mathbf{V}_{\text{SEL}}), \quad (4.37)$$

where $\mathbf{V}_{\text{SEL}} = \nabla [\boldsymbol{\mu}^{-1}(\boldsymbol{\theta}_0)] \cdot \boldsymbol{\Sigma}(\boldsymbol{\theta}_0) \cdot \nabla [\boldsymbol{\mu}^{-1}(\boldsymbol{\theta}_0)]'$.

As Assumption 5 implies the interchangeability of differentiation and inversion for $\boldsymbol{\mu}(\boldsymbol{\theta})$, we have

$$\begin{aligned} \mathbf{V}_{\text{SEL}} &= [\nabla \boldsymbol{\mu}(\boldsymbol{\theta}_0)]^{-1} \cdot \boldsymbol{\Sigma}(\boldsymbol{\theta}_0) \cdot \{[\nabla \boldsymbol{\mu}(\boldsymbol{\theta}_0)]^{-1}\}' \\ &= [\nabla \boldsymbol{\mu}(\boldsymbol{\theta}_0)]' \cdot \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1} \cdot \nabla \boldsymbol{\mu}(\boldsymbol{\theta}_0)^{-1}. \end{aligned} \quad (4.38)$$

With $\mathbf{g}(\mathbf{x}, \boldsymbol{\theta})$ from \mathbf{V}_{EL} (4.10) given by $\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}, \boldsymbol{\theta}) - \mathbf{s}_N$, we immediately have $\mathbf{V}_{\text{SEL}} = \mathbf{V}_{EL}$ in (4.37).

For B a continuity set of \mathbf{Z} , we have

$$\begin{aligned} \left| P(\sqrt{N}(\hat{\boldsymbol{\theta}}_{\text{SEL}} - \boldsymbol{\theta}_0) \in B) - P(\mathbf{Z} \in B) \right| &\leq \left| P(\sqrt{N}(\hat{\boldsymbol{\theta}}_{\text{SEL}} - \boldsymbol{\theta}_0) \in B) - P(\sqrt{N}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \in B) \right| \quad (\text{I}) \\ &\quad + \left| P(\sqrt{N}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \in B) - P(\mathbf{Z} \in B) \right| \quad (\text{II}). \end{aligned}$$

For (II), according to (4.37) which states $\sqrt{N}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution to $\text{Normal}(\mathbf{0}, \mathbf{V}_{EL})$, it follows that for any continuity set B of \mathbf{Z} ,

$$P(\sqrt{N}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \in B) \rightarrow P(\mathbf{Z} \in B) \quad \text{as } N \rightarrow \infty. \quad (4.39)$$

Thus, for any $\epsilon > 0$, there exists N_0 such that (II) $< \epsilon/2$, for all $N > N_0$.

For (I), according to (4.36), for any fixed N , we have $\sqrt{N}(\hat{\boldsymbol{\theta}}_{\text{SEL}} - \boldsymbol{\theta}_0) \xrightarrow{a.s.} \sqrt{N}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ as $R \rightarrow \infty$, which implies convergence in distribution $\sqrt{N}(\hat{\boldsymbol{\theta}}_{\text{SEL}} - \boldsymbol{\theta}_0) \xrightarrow{d} \sqrt{N}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ as well. That is, there exists $R(N)$ such that (I) $< \epsilon/2$, for $R > R(N)$.

Combining (I) and (II) gives the desired result. □

Theorem 1 implies that, if Assumptions 1–5 hold, as $R, N \rightarrow \infty$ at an appropriate rate,

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{\text{SEL}} - \boldsymbol{\theta}_0) \xrightarrow{d} \text{Normal}(\mathbf{0}, \mathbf{V}_{EL}). \quad (4.40)$$

Also, Frazier et al. (2023) showed that the synthetic likelihood estimator has the same limiting distribution as $\text{Normal}(\mathbf{0}, \mathbf{V}_{EL})$ under Assumptions 1–5.

4.3.2 Over-identified moment condition

In the presence of over-identification, where the number of summary statistics exceeds the number of parameters ($m > p$), for almost every fixed \mathbf{s}_N , there does not exist any value of $\boldsymbol{\theta}$ that satisfies the moment condition $\mathbb{E}[\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}, \boldsymbol{\theta})] - \mathbf{s}_N = \mathbf{0}$. The reason is that $\boldsymbol{\mu}(\boldsymbol{\theta})$ lies in a lower dimensional manifold of the space where lies \mathbf{s}_N . A toy example for illustration is provided in Section 4.4.4.

In the classic EL, it is also common to encounter situations where no value of $\boldsymbol{\theta}$ solves the moment condition $\mathbb{E}[\mathbf{g}(\mathbf{x}, \boldsymbol{\theta})] = \mathbf{0}$ (Maasoumi and Phillips, 1982; Imbens, 1997; Hall and Inoue, 2003). Although the EL estimator $\hat{\boldsymbol{\theta}}_{\text{EL}}$ may still converge to some limiting value, this convergence does not occur at the typical \sqrt{N} rate (Schennach, 2007). To address this issue, Schennach (2007) proposed Exponentially Tilted Empirical Likelihood (ETEL), a variant of EL that ensures the associated estimator achieves \sqrt{N} -consistency, restoring the desirable convergence rate.

Building on these developments, we introduce a similar variant in the context of simulation-based methods, which we refer to as synthetic ETEL (SETEL). This approach is detailed in Section 4.3.2.2, following a review of the classical ETEL in Section 4.3.2.1.

4.3.2.1 ETEL

Schennach (2007) proposed the exponentially tilted empirical likelihood (ETEL) estimator, which is defined as

$$\hat{\boldsymbol{\theta}}_{\text{ETEL}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{i=1}^N \log \omega_i(\boldsymbol{\theta}), \quad (4.41)$$

where $\hat{\boldsymbol{\omega}}(\boldsymbol{\theta}) = (\hat{\omega}_1(\boldsymbol{\theta}), \dots, \hat{\omega}_N(\boldsymbol{\theta}))$ solves the optimization problem,

$$\hat{\boldsymbol{\omega}}(\boldsymbol{\theta}) = \underset{\boldsymbol{\omega} \in \mathcal{C}(\mathbf{G}_{\boldsymbol{\theta}})}{\operatorname{argmax}} \sum_{i=1}^N -\omega_i \log \omega_i, \quad (4.42)$$

and $\mathbf{G}_{\boldsymbol{\theta}}$ is the same as that in the EL method (4.4).

Schennach (2007) also showed that, the calculation of $\omega_i(\boldsymbol{\theta})$ of (4.42) is equivalent to

$$\omega_i(\boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\lambda}(\boldsymbol{\theta})' \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})\}}{\sum_{j=1}^N \exp\{\boldsymbol{\lambda}(\boldsymbol{\theta})' \mathbf{g}(\mathbf{x}_j, \boldsymbol{\theta})\}}, \quad (4.43)$$

where

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) = \underset{\boldsymbol{\lambda}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^N \exp\{\boldsymbol{\lambda}'\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})\}. \quad (4.44)$$

Similar to EL, Equation (4.44) is a convex optimization problem, which can be effectively solved by Newton-Raphson (when the solution exists).

We next review the asymptotic properties of the ETEL estimator. When there exists a value of $\boldsymbol{\theta}$ that satisfies the moment condition, the ETEL estimator $\hat{\boldsymbol{\theta}}_{\text{ETEL}}$ shares the same limiting distribution as the EL estimator $\hat{\boldsymbol{\theta}}_{\text{EL}}$, with identical asymptotic variance \mathbf{V}_{EL} . In contrast, when no value of $\boldsymbol{\theta}$ satisfies the moment condition (i.e., $\mathbb{E}[\mathbf{g}(\mathbf{x}, \boldsymbol{\theta})] \neq \mathbf{0}$ for any $\boldsymbol{\theta}$), $\hat{\boldsymbol{\theta}}_{\text{ETEL}}$ has favorable asymptotic properties that $\hat{\boldsymbol{\theta}}_{\text{EL}}$ does not possess.

In the ETEL approach for such cases, [Schennach \(2007\)](#) and [Chib et al. \(2018\)](#) defined $\boldsymbol{\lambda}_0(\boldsymbol{\theta})$ as the unique solution to

$$\boldsymbol{\lambda}_0(\boldsymbol{\theta}) = \underset{\boldsymbol{\lambda}}{\operatorname{argmin}} \mathbb{E}[\exp\{\boldsymbol{\lambda}'(\boldsymbol{\theta})\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})\}]. \quad (4.45)$$

Then the object of interest $\boldsymbol{\theta}_0$ is defined as the unique value that maximizes the function

$$L(\boldsymbol{\theta}) = \log(\mathbb{E}[\exp(\boldsymbol{\lambda}'_0(\boldsymbol{\theta})(\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}) - \mathbb{E}[\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})]))]). \quad (4.46)$$

Note that $\boldsymbol{\theta}_0$ also maximizes the function $L(\boldsymbol{\theta})$ when $\boldsymbol{\theta}_0$ solves the moment condition.

[Schennach \(2007\)](#) derived the asymptotic property of the ETEL estimator by using Ψ -type M estimators ([Huber, 1992](#)). By Lemma 9 of [Schennach \(2007\)](#), the ETEL estimator $\hat{\boldsymbol{\theta}}_{\text{ETEL}}$ corresponds to the $\hat{\boldsymbol{\theta}}$ component of $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\kappa}}, \hat{\tau})$, which solves

$$\frac{1}{N} \sum_{i=1}^N \boldsymbol{\psi}(\mathbf{x}_i, \hat{\boldsymbol{\beta}}) = \mathbf{0}, \quad (4.47)$$

where

$$\boldsymbol{\psi}(\mathbf{x}_i, \boldsymbol{\beta}) = \begin{bmatrix} \exp(\boldsymbol{\lambda}'\mathbf{g}_i)\nabla\mathbf{g}'_i(\boldsymbol{\kappa} + \boldsymbol{\lambda}\mathbf{g}'_i\boldsymbol{\kappa} - \boldsymbol{\lambda}) + \tau\nabla\mathbf{g}'_i\boldsymbol{\lambda} \\ (\tau - \exp(\boldsymbol{\lambda}'\mathbf{g}_i))\mathbf{g}_i + \exp(\boldsymbol{\lambda}'\mathbf{g}_i)\mathbf{g}_i\mathbf{g}'_i\boldsymbol{\kappa} \\ \exp(\boldsymbol{\lambda}'\mathbf{g}_i)\mathbf{g}_i \\ \exp(\boldsymbol{\lambda}'\mathbf{g}_i) - \tau \end{bmatrix}, \quad (4.48)$$

where $\mathbf{g}_i = \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})$, $\nabla\mathbf{g}_i = \frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}) \in \mathbb{R}^{m \times m}$. The dimensions of $\boldsymbol{\beta} \in \mathbb{R}^{1+2m+p}$ are given by subvectors $\boldsymbol{\theta} \in \mathbb{R}^{p \times 1}$, $\boldsymbol{\lambda} \in \mathbb{R}^{m \times 1}$, $\boldsymbol{\kappa} \in \mathbb{R}^{m \times 1}$, and scalar $\tau \in \mathbb{R}$. [Lee \(2016\)](#) gave the

expression of $\hat{\tau}$ and $\hat{\kappa}$ given by

$$\begin{aligned}\hat{\tau} &= \frac{1}{N} \sum_{i=1}^N \exp(\hat{\lambda}' \hat{\mathbf{g}}_i), \\ \hat{\kappa} &= - \left(\frac{1}{N} \sum_{i=1}^N \frac{\exp(\hat{\lambda}' \hat{\mathbf{g}}_i)}{\hat{\tau}} \hat{\mathbf{g}}_i \hat{\mathbf{g}}_i' \right)^{-1} \hat{\mathbf{g}}_N,\end{aligned}\tag{4.49}$$

where $\hat{\mathbf{g}}_i = \mathbf{g}(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$, $\hat{\mathbf{g}}_N = \frac{1}{N} \sum_{i=1}^N \mathbf{g}(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$.

Then Theorem 10 in [Schennach \(2007\)](#) gave

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \rightarrow \text{Normal}(\mathbf{0}, \boldsymbol{\Gamma}^{-1} \boldsymbol{\Psi} (\boldsymbol{\Gamma}')^{-1}),\tag{4.50}$$

where $\boldsymbol{\beta}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\lambda}_0, \boldsymbol{\kappa}_0, \tau_0)$, $\boldsymbol{\Gamma} = \text{E}[\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\psi}(\mathbf{x}_i, \boldsymbol{\beta}) |_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}]$, and $\boldsymbol{\Psi} = \text{E}[\boldsymbol{\psi}(\mathbf{x}_i, \boldsymbol{\beta}_0) \boldsymbol{\psi}(\mathbf{x}_i, \boldsymbol{\beta}_0)']$.

Denote the upper left $p \times p$ submatrix of $\boldsymbol{\Gamma}^{-1} \boldsymbol{\Psi} (\boldsymbol{\Gamma}')^{-1}$ as \mathbf{V}_{ETEL} . Then we have

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{\text{ETEL}} - \boldsymbol{\theta}_0) \rightarrow \text{Normal}(\mathbf{0}, \mathbf{V}_{\text{ETEL}}).\tag{4.51}$$

[Lee \(2016\)](#) remarked that, when no $\boldsymbol{\theta}$ value satisfies the moment conditions, \mathbf{V}_{ETEL} in (4.51) is different from the classic EL variance \mathbf{V}_{EL} in (4.9). In contrast, when there exists a $\boldsymbol{\theta}$ value which satisfies the moment conditions, \mathbf{V}_{ETEL} is identical to \mathbf{V}_{EL} .

4.3.2.2 SETEL

Analogously to EL, in SEL, there does not exist any value of $\boldsymbol{\theta}$ to satisfy the moment condition $\text{E}[\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}, \boldsymbol{\theta})] - \mathbf{s}_N = \mathbf{0}$ for almost every fixed \mathbf{s}_N under over-identifications. This motivates us to combine SEL and ETEL, and we suspect that the new method may inherit the favorable property of ETEL when the moment condition cannot be satisfied.

Based on [Schennach \(2007\)](#), the synthetic exponentially tilted empirical likelihood function (SETEL) is defined as:

$$\text{SETEL}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)}) = \exp \left[\frac{1}{R} \cdot \sum_{r=1}^R \log \hat{\omega}_r(\boldsymbol{\theta}) \right],\tag{4.52}$$

where $\hat{\boldsymbol{\omega}}(\boldsymbol{\theta}) = (\hat{\omega}_1(\boldsymbol{\theta}), \dots, \hat{\omega}_R(\boldsymbol{\theta}))$ solves the optimization problem,

$$\hat{\boldsymbol{\omega}}(\boldsymbol{\theta}) = \underset{\boldsymbol{\omega} \in \mathcal{C}(\mathbf{G}_{\boldsymbol{\theta}})}{\text{argmax}} \sum_{r=1}^R -\omega_r \log \omega_r,\tag{4.53}$$

where \mathbf{G}_θ is the same as that in the SEL method (4.25).

The SETEL estimator $\hat{\boldsymbol{\theta}}_{\text{SETEL}}$ is defined as $\hat{\boldsymbol{\theta}}_{\text{SETEL}} = \operatorname{argmax}_\theta \text{SETEL}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)})$. Under just-identification, $\hat{\boldsymbol{\theta}}_{\text{SETEL}}$ shares the same limiting distribution as $\hat{\boldsymbol{\theta}}_{\text{SEL}}$ in Theorem 1, because EL and ETEL share identical asymptotic properties when the moment condition has a solution for some parameter value. For the over-identified setting, deriving the asymptotic theory of the SETEL estimator is the subject of further research. However, the simulation study in Section 4.4 empirically supports the effectiveness of the proposed SETEL estimator $\hat{\boldsymbol{\theta}}_{\text{SETEL}}$ under such case.

4.3.3 Standard error calculation

In the previous sections, we have introduced three point estimators based on reparametrization trick: the SEL, SETEL, and SL estimators. In this section, we propose several variance estimators for constructing standard errors and confidence intervals for the SEL estimator. The methods extend naturally to the SETEL and SL estimators and have counterparts in Bayesian inference via BNA.

The key insight underlying our approach is that when \mathbf{s}_N is fixed under just-identification, the SEL problem reduces to a classic EL problem. Consequently, the EL property (4.9), together with Assumptions 1–5, implying $\boldsymbol{\theta} = \boldsymbol{\mu}^{-1}(\mathbf{s}_N)$, give

$$\sqrt{R}(\hat{\boldsymbol{\theta}}_{\text{SEL}} - \tilde{\boldsymbol{\theta}}) \rightarrow N(\mathbf{0}, \mathbf{A}^{-1} \mathbf{B}_N (\mathbf{A}^{-1})'), \quad (4.54)$$

where

$$\begin{aligned} \mathbf{A} &= \frac{\partial}{\partial \boldsymbol{\theta}} \mathbb{E} \left[\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \tilde{\boldsymbol{\theta}}) - \mathbf{s}_N \right] = \frac{\partial}{\partial \boldsymbol{\theta}} \mathbb{E} \left[\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \tilde{\boldsymbol{\theta}}) \right] = \nabla \boldsymbol{\mu}(\tilde{\boldsymbol{\theta}}), \\ \mathbf{B}_N &= \mathbb{E}[(\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \tilde{\boldsymbol{\theta}}) - \mathbf{s}_N)(\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \tilde{\boldsymbol{\theta}}) - \mathbf{s}_N)'] \\ &= \mathbb{E}[(\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \tilde{\boldsymbol{\theta}}) - \boldsymbol{\mu}(\tilde{\boldsymbol{\theta}}))(\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \tilde{\boldsymbol{\theta}}) - \boldsymbol{\mu}(\tilde{\boldsymbol{\theta}}))'] = \boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}})/N. \end{aligned} \quad (4.55)$$

As theorem 1 shows that $\tilde{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta}_0$, we have $N\mathbf{A}^{-1}\mathbf{B}_N(\mathbf{A}^{-1})' \rightarrow \mathbf{V}_{\text{EL}}$ by the continuity mapping theorem. Therefore, any variance estimator from the classic EL method (as Equation (4.54) is a classic EL problem) that consistently estimates $N\mathbf{A}^{-1}\mathbf{B}_N(\mathbf{A}^{-1})'$, can be adapted, with an appropriate scaling factor involving R , to a consistent estimator of \mathbf{V}_{EL} in the SEL problem.

Based on this relationship, we consider four primary types of variance estimators—**QL-type**, **Hessian-type**, **Jacobian-type**, **Sandwich-type**, adapted from classic EL methods. Since the SEL, SETEL, SL estimators share the identical asymptotic distribution

under just-identification, these variance estimators are broadly flexible: the first two apply to SL, SEL, and SETEL, while all four apply to SEL and SETEL. We use the acronym ‘SEL-QL’ to represent a method where the SEL point estimator is implemented, and its variance is estimated using the QL-type variance estimator. Similar rules apply to other point and variance estimators.

Given the observed summary statistics \mathbf{s}_N , the SEL estimate $\hat{\boldsymbol{\theta}}_{\text{SEL}}$, and R sets of noise variables $\boldsymbol{\varepsilon}_{1:N}^{(1:R)}$, the steps for calculating the estimated covariance $\hat{\mathbf{V}}$ of $\hat{\boldsymbol{\theta}}_{\text{SEL}}$ using these four variance estimators are summarized below.

- For SEL-QL, compute $\hat{\mathbf{V}}_{\text{SEL-QL}}$ by

$$\begin{aligned} \hat{\mathbf{V}}_{\text{SEL-QL}} = & \left\{ \sum_{r=1}^R \hat{\omega}_r \cdot \left. \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \boldsymbol{\theta}) \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{SEL}}} \right\}^{-1} \times \\ & \left\{ \sum_{r=1}^R \hat{\omega}_r \cdot (\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \hat{\boldsymbol{\theta}}_{\text{SEL}}) - \mathbf{s}_N)(\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \hat{\boldsymbol{\theta}}_{\text{SEL}}) - \mathbf{s}_N)' \right\} \times \\ & \left(\left\{ \sum_{r=1}^R \hat{\omega}_r \cdot \left. \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \hat{\boldsymbol{\theta}}) \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{SEL}}} \right\}' \right)^{-1}, \end{aligned} \quad (4.56)$$

where $\hat{\omega}_r = \omega_r(\hat{\boldsymbol{\theta}}_{\text{SEL}}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)})$ comes from $\log \text{SEL}(\hat{\boldsymbol{\theta}}_{\text{SEL}}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)}) = \sum_{r=1}^R \log \omega_r(\hat{\boldsymbol{\theta}}_{\text{SEL}}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)})$.

- For SEL-Hess, compute $\hat{\mathbf{V}}_{\text{SEL-Hess}}$ by

$$\hat{\mathbf{V}}_{\text{SEL-Hess}} = \left[- \left. \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log \text{SEL}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)}) \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{SEL}}} \right]^{-1} \quad (4.57)$$

- For SEL-Jac, compute $\hat{\mathbf{V}}_{\text{SEL-jac}}$ by

$$\hat{\mathbf{V}}_{\text{SEL-jac}} = \left[\frac{1}{R} \sum_{r=1}^R \left. \frac{\partial}{\partial \boldsymbol{\theta}} \log \omega_r(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)}) \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{SEL}}} \cdot \left. \frac{\partial}{\partial \boldsymbol{\theta}} \log \omega_r(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)})' \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{SEL}}} \right]^{-1} \quad (4.58)$$

- For SEL-Sand, compute $\hat{\mathbf{V}}_{\text{SEL-Sand}}$ by

$$\begin{aligned} \hat{\mathbf{V}}_{\text{SEL-Sand}} = & \left[-\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log \text{SEL}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)}) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{SEL}}} \right]^{-1} \times \\ & \left[\frac{1}{R} \sum_{r=1}^R \frac{\partial}{\partial \boldsymbol{\theta}} \log \omega_r(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)}) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{SEL}}} \cdot \frac{\partial}{\partial \boldsymbol{\theta}} \log \omega_r(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)})' \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{SEL}}} \right] \times \\ & \left[-\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log \text{SEL}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)}) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{SEL}}} \right]^{-1}. \end{aligned} \quad (4.59)$$

This method is an adaption from the asymptotic covariance of the EL and ETEL estimator, introduced in [Lee \(2016\)](#) and [Chib et al. \(2018\)](#).

By construction, all these four variance estimators, when multiplied by N , are consistent estimators of \mathbf{V}_{EL} in Theorem 1. However, under over-identification, their consistency remains uncertain. Nonetheless, these four estimators can still be computed, and their performance is assessed in the simulation study of Section 4.4.

All these four estimators are readily defined for SETEL as well, by replacing the SEL function and point estimate with the SETEL function and point estimate. However, SETEL enjoys another estimator, SETEL-M, adapted from the ψ -type M-estimator in [Schenach \(2007\)](#). For SETEL-M, first compute the $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\theta}}_{\text{SETEL}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\kappa}}, \hat{\tau})$ and $\boldsymbol{\psi}_{\text{SETEL}}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \hat{\boldsymbol{\beta}})$, $r = 1, \dots, R$, by letting $\mathbf{g}_r = \mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \boldsymbol{\theta}) - \mathbf{s}_N$ in (4.48). Then compute $\hat{\boldsymbol{\Sigma}}_{\text{SETEL}}$ by

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{\text{SETEL}} = & \left\{ \frac{1}{R} \sum_{r=1}^R \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\psi}_{\text{SETEL}}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right\}^{-1} \times \\ & \left\{ \frac{1}{R} \sum_{r=1}^R \boldsymbol{\psi}_{\text{SETEL}}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \hat{\boldsymbol{\beta}}) \boldsymbol{\psi}_{\text{SETEL}}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \hat{\boldsymbol{\beta}})' \right\} \\ & \left\{ \left[\frac{1}{R} \sum_{r=1}^R \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\psi}_{\text{SETEL}}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right]' \right\}^{-1}. \end{aligned} \quad (4.60)$$

Finally, $\hat{\mathbf{V}}_{\text{SETEL-M}}$ is calculated as the upper left $p \times p$ submatrix of $\hat{\boldsymbol{\Sigma}}_{\text{SETEL}}$. Under just-identification, $N \cdot \hat{\mathbf{V}}_{\text{SETEL-M}}$ also consistently estimates \mathbf{V}_{EL} .

SL has analogous QL-type and Hessian-type variance estimators, using the SL function and point estimate, with the QL-type estimator additionally replacing ω_r with R^{-1} . To

the best of our knowledge, this is the first application of SL which does not involve MCMC methods, and is thus a contribution of this Chapter in its own right.

In addition to frequentist variance estimation, we extend the discussion to Bayesian variance estimation by incorporating prior information directly into the variance calculation. Taking SEL as an example, within the BNA framework, the Bayesian variance estimator is calculated as:

$$\hat{\mathbf{V}}_{\text{Bayes}} = \left[\hat{\mathbf{V}}^{-1}(\hat{\boldsymbol{\theta}}_{\text{Bayes}}) + \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log \pi(\hat{\boldsymbol{\theta}}_{\text{Bayes}}) \right]^{-1}, \quad (4.61)$$

where $\pi(\boldsymbol{\theta})$ is the prior of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_{\text{Bayes}}$ is the mode of the posterior, given by

$$\hat{\boldsymbol{\theta}}_{\text{Bayes}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} [\log \text{SEL}(\boldsymbol{\theta}) + \log \pi(\boldsymbol{\theta})], \quad (4.62)$$

and $\hat{\mathbf{V}}$ is the frequentist variance estimator introduced above.

As to the construction of confidence interval, with any variance estimate $\hat{\mathbf{V}}$ obtained above, the $1-\alpha$ level confidence interval of the i -th element of $\boldsymbol{\theta}$ is $\left[\left(\hat{\boldsymbol{\theta}}_{\text{SEL}} \right)_i \pm |z_{\frac{\alpha}{2}}| \cdot \sqrt{\hat{\mathbf{V}}_{ii}} \right]$, where $\alpha \in (0, 1)$, z_{α} is the α quantile value of the standard normal distribution, $\left(\hat{\boldsymbol{\theta}}_{\text{SEL}} \right)_i$ is the i -th element of $\hat{\boldsymbol{\theta}}_{\text{SEL}}$, and $\hat{\mathbf{V}}_{ii}$ represents the i -th diagonal element of $\hat{\mathbf{V}}$. In the Bayesian framework, the credible interval is obtained by replacing $\hat{\boldsymbol{\theta}}_{\text{SEL}}$ and $\hat{\mathbf{V}}$ with $\hat{\boldsymbol{\theta}}_{\text{Bayes}}$ and $\hat{\mathbf{V}}_{\text{Bayes}}$.

A Similar calculation for both Bayesian variance estimation and interval construction applies to SETEL and SL, with the corresponding point and variance estimates substituted accordingly.

4.4 Simulation Study

The overall goal of the simulation study is to evaluate the performance of our proposed methods, which are point and variance estimators based on SEL, SETEL, and SL. Specifically, nine of our proposed methods in Section 4.3 are examined in the simulation study: three SEL-class methods (SEL-QL, SEL-Hess, SEL-Jac), five SETEL-class methods (SETEL-QL, SETEL-Hess, SETEL-Jac, SETEL-Sand, SETEL-M), and one SL method (SL-Hess).

To achieve this, we investigate two scenarios: the just-identified scenario through the first three examples, and the over-identified scenario through the last two examples. The

first example is implemented with Bayesian inference to compare our proposed methods to MCMC, while the remaining four examples focus on frequentist inference to evaluate other aspects of the proposed methods. The three just-identified examples involve models with increasing complexity, allowing us to assess the robustness of our proposed methods. If not specifically mentioned, automatic differentiation (e.g., Margossian, 2019) is used in the section for high-performance computations of the Hessian and gradient, using the Python library JAX (Bradbury et al., 2018).

As discussed in Section 4.2.2, it is well-known that EL may suffer from the empty-set issue. SEL and SETEL are expected to encounter the same issue, which poses challenges for initializing mode-finding algorithms for computing the SEL and SETEL estimators. To address this, we consider the adjusted SEL (ASEL) that adjusts the support of SEL. Specifically, the adjusted SEL function is given by

$$\text{ASEL}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)}) = \exp \left[\frac{1}{R+1} \cdot \sum_{r=1}^{R+1} \log \hat{\omega}_r(\boldsymbol{\theta}) \right], \quad (4.63)$$

where $\hat{\boldsymbol{\omega}}(\boldsymbol{\theta}) = (\hat{\omega}_1(\boldsymbol{\theta}), \dots, \hat{\omega}_R(\boldsymbol{\theta}), \hat{\omega}_{R+1}(\boldsymbol{\theta}))$ solves the optimization problem,

$$\hat{\boldsymbol{\omega}}(\boldsymbol{\theta}) = \underset{\boldsymbol{\omega} \in \mathcal{L}(\mathbf{G}_{\boldsymbol{\theta}}^+)}{\operatorname{argmax}} \sum_{r=1}^{R+1} \log \hat{\omega}_r, \quad (4.64)$$

where $\mathbf{G}_{\boldsymbol{\theta}}^+$ is $(R+1) \times m$ with rows $\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(1)}, \boldsymbol{\theta}) - \mathbf{s}_N, \dots, \mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(R)}, \boldsymbol{\theta}) - \mathbf{s}_N, -\frac{a_R}{R} \sum_{r=1}^R [\mathbf{s}(\boldsymbol{\varepsilon}_{1:N}^{(r)}, \boldsymbol{\theta}) - \mathbf{s}_N]$, and $a_R = \max(1, \log(R)/2)$.

The steps for calculating $\hat{\boldsymbol{\theta}}_{\text{SEL}}$ are then summarized as follows.

1. Given the observed data $\mathbf{x}_{1:N}$, find the observed summary statistics \mathbf{s}_N .
2. Pre-generated R sets of noise variables, $\boldsymbol{\varepsilon}_{1:N}^{(1:R)}$, according to the model and the reparametrization trick.
3. Given the noise $\boldsymbol{\varepsilon}_{1:N}^{(1:R)}$ and the observed summary statistics \mathbf{s}_N , find the initial value by adjusted SEL, $\hat{\boldsymbol{\theta}}_{\text{init}} = \max_{\boldsymbol{\theta}} \text{ASEL}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)})$.
4. Given the noise $\boldsymbol{\varepsilon}_{1:N}^{(1:R)}$, the observed summary statistics \mathbf{s}_N , and the initial value $\hat{\boldsymbol{\theta}}_{\text{init}}$, calculate the SEL estimate $\hat{\boldsymbol{\theta}}_{\text{SEL}} = \operatorname{argmax}_{\boldsymbol{\theta}} \text{SEL}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:N}^{(1:R)})$.

Note that the ASEL function is merely used to provide an initial value for computing $\hat{\theta}_{\text{SEL}}$, by maximizing the ASEL function. The calculations of $\hat{\theta}_{\text{SETEL}}$ and $\hat{\theta}_{\text{SL}}$ follow a similar procedure, except that step 4 maximizes the SETEL and SL functions, respectively, instead of the SEL function. In addition, $\hat{\theta}_{\text{SL}}$ does not require step 3 for initialization, as the SL function $\text{SL}(\theta, \varepsilon_{1:N}^{(1:R)})$, inherently a multivariate normal density, is well defined for any parameter value and thus does not encounter the empty-set issue.

4.4.1 Univariate normal distribution under just-identification

In the first example, we consider the univariate normal distribution, a toy example chosen for its simplicity and tractable likelihood, also used in Chaudhuri et al. (2024). The normal data $x \sim \text{Normal}(\mu, \sigma^2)$ can be reparametrized as $x = \mu + \sigma\varepsilon$, where the noise $\varepsilon \sim \text{Normal}(0, 1)$ is independent of $\theta = (\mu, \sigma)$. Suppose in the simulation study, we have the observed data $x_{1:N}$ from the normal distribution $\text{Normal}(\theta, 1)$, where the true parameter value is $\theta_0 = 0$ and the prior is $\theta \propto 1$.

We first investigate the just-identification, in which the number of summary statistics equals to the number of parameters (i.e., $m = p = 1$). The summary statistic is the mean of $x_{1:N}$, that is $s_N = N^{-1} \sum_{i=1}^N x_i$. Different numbers of simulated datasets are considered: $R = 25, 100, 1000, 10000$. For each R , we simulate 1000 datasets (i.e., $N_{\text{rep}} = 1000$ repetition times).

Several methods of Bayesian point and variance estimators are considered for comparison. First, we include the true posterior which combines the parametric likelihood and the prior, using the Hessian matrix to calculate the variance via BNA. Next, We examine two MCMC methods, BSL-MCMC and SEL-MCMC, which are traditionally used to construct point and interval estimates for BSL and SEL. These MCMC methods use the same algorithms as Price et al. (2018) for BSL-MCMC and Chaudhuri et al. (2024) for SEL-MCMC, both of which are based on Algorithm 1. Finally, we investigate nine of our proposed methods mentioned at the beginning of this section, using BNA.

The true posterior is regarded as the ground truth. The two MCMC methods, BSL-MCMC and SEL-MCMC, use Metropolis-Hastings to sample from the implicit SEL and BSL posterior based on the implicit SEL and SL function, respectively, with 10,000 sampling and 10,000 burn-in iterations. Point estimates are summarized by the posterior mean, and the 95% credible intervals are constructed by the 2.5% and 97.5% quantiles. A key difference between MCMC methods and our proposed approach lies in how they handle simulated noise. MCMC methods generate new noise at each iteration to compute the posterior, whereas our method simulates the noise only once for a given observed dataset.

Figure 4.2 present three performance metrics for the methods over $N_{\text{rep}} = 1000$ simulation repetitions: average posterior standard deviation (APSD), root mean square error (RMSE), and 95% coverage probability (CP), which is the proportion of samples for which the true value is enclosed by the credible interval. Since all proposed methods yielded identical results to four decimal places, we present only the SEL-Hess results in the figure. These results align with the expected asymptotic behaviors: under just-identification and standard regularity conditions, the SEL, SETEL, and SL estimators share the same limiting distribution, which is normal with mean θ_0 and variance \mathbf{V}_{EL}/N . Furthermore, all our proposed variance estimators, when multiplied by N , consistently estimate \mathbf{V}_{EL} . Notably, as R increases, APSD, RMSE, CP of these proposed methods converge toward the true posterior results, with the additional cost of computation. In contrast, for SEL-MCMC, there is an observed increase of APSD and CP as R grows, indicating an undesirable property.

To assess whether two methods produce significantly different values of a performance metric (e.g., ASE), we compute the standard error of the metric from one method as the error bar, summarized in the caption of Figure 4.2, using nonparametric bootstrap resampling with 10,000 replications. A 95% interval is then constructed by adding and subtracting 1.96 times the error bar from the metric. If the corresponding value from the other method falls outside this interval, we conclude that the difference is statistically significant at the 5% level. The error bars suggest that RMSE values are not significantly different across methods at any R . When $R > 1000$, SEL-MCMC has significantly different ASE and CP from other methods, whereas the remaining methods do not differ significantly from each other in these two metrics.

We then compare the computational speed of our proposed methods with traditional MCMC methods. Figure 4.3 presents the average execution time of our proposed methods at $R = 1000$ for computing point and variance (or interval) estimates for a specific observed dataset, with time reported in seconds. For MCMC methods, the speeds are obtained from a single repetition with 10,000 iterations, as this number is typically required to ensure good approximation and convergence for MCMC, while for the other methods, the speeds are averaged over 100 repetitions.

Figure 4.3 highlights a few points. First and foremost, our proposed methods are significantly faster than the two MCMC methods. For instance, although both are based on SL, SL-Hess computes point and variance estimates in just 0.078 seconds, compared to the 29 seconds required by BSL-MCMC. Second, SEL-Hess stands out as the fastest method overall, for this specific example. Third, SEL-MCMC is noticeably slower than BSL-MCMC as the inner optimization of SEL requires substantial additional steps when evaluated further from the posterior mode, resulting in increased computational overhead.

We next examine the behavior of the densities of SEL, SETEL, SL posterior when R grows. Figure 4.4 displays the densities of SEL, SETEL, SL posterior, along with the corresponding BNA posterior densities based on SEL-Hess, SETEL-Hess, and SL-Hess, for a particular observed dataset under the just-identification with a variety of R values. The true posterior is added as the ground truth. For easier shape comparison, all densities are scaled so that the integral over the entire space equals one.

Notably, as R increases, the density of SEL pseudo-posterior flattens and moves further away from the density of true posterior, a phenomenon theoretically explored in Section 4.2 of Chaudhuri et al. (2024). This explains why the variance of MCMC samples explodes as R increases. In contrast, the BNA density of SEL pseudo-posterior moves closer to the true posterior by increased R , since it only evaluates the SEL pseudo-posterior around the mode.

Further, Figure 4.4 shows that, unlike SEL, the density of SETEL pseudo-posterior converges to that of the true posterior as R increases. This indicates that MCMC could be potentially applicable to SETEL. However, the SETEL pseudo-posterior requires $R = 30000$ to closely match the true posterior, whereas BNA via SETEL-Hess achieves similar accuracy with only $R = 1000$. Regarding SL, both SL and its BNA via SL-Hess match the true posterior when $R = 1000$. The strong performance of SL may be attributed to the fact that the mean of normal data naturally follows a normal distribution.

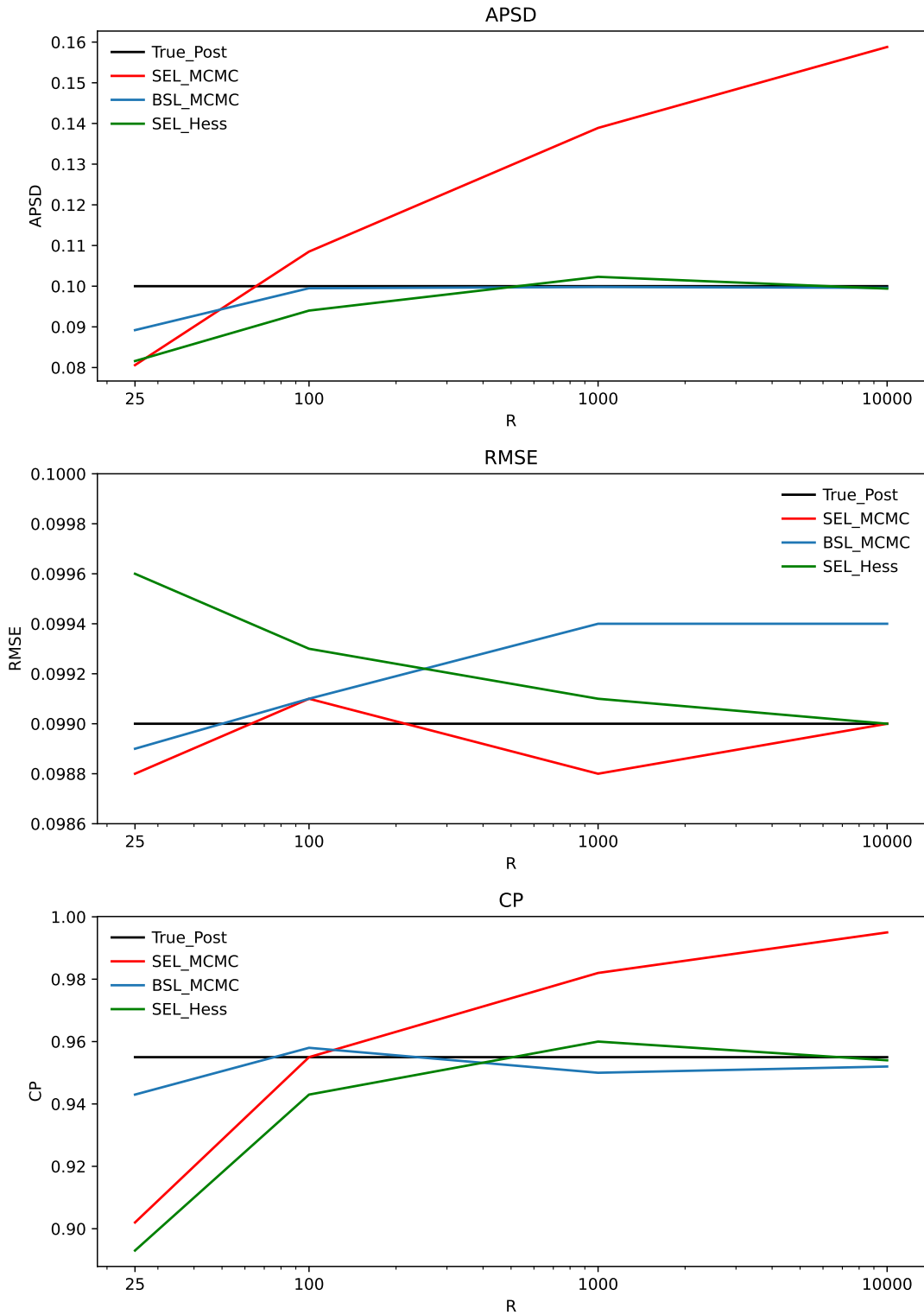


Figure 4.2: The average posterior standard deviation (APSD), root mean square error (RMSE), and 95% coverage probability (CP) for the just-identified normal example, with 1000 repetitions and the number of the simulated datasets $R = 25, 100, 1000, 10000$. Error bars per R per method are approximately 6×10^{-5} for ASE, 0.003 for RMSE, and 0.007 for CP.

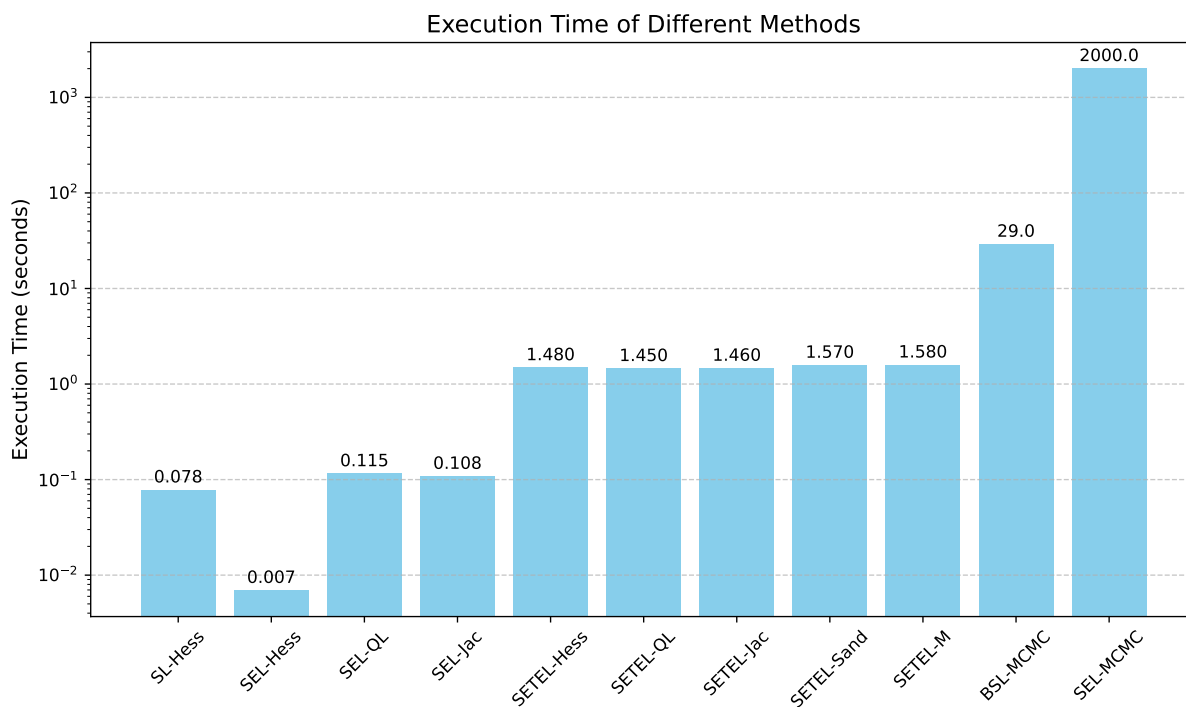


Figure 4.3: The average execution time of various methods to compute point and variance (or interval) estimates for a specific observed dataset in the just-identified normal example with $R = 1000$. The time unit is second. For MCMC methods, the speeds are obtained from a single repetition with 10000 iterations, excluding the burnin periods, while for the other methods, the speeds are averaged over 100 repetitions.

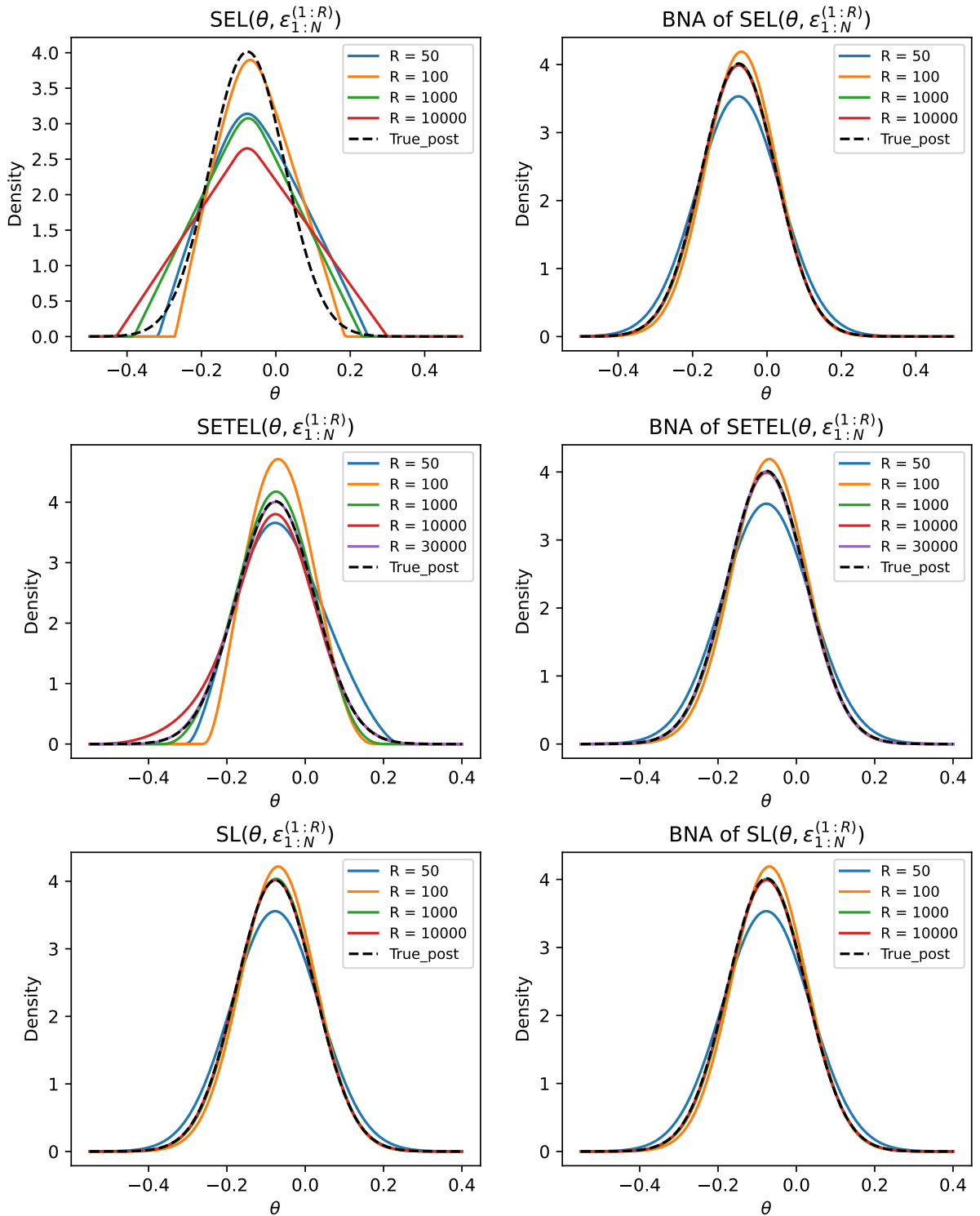


Figure 4.4: The left three panels display the densities of SEL, SETEL, and SL posterior with the prior $\theta \propto 1$ in the just-identified normal example, while the right three panels present the corresponding BNA densities based on SEL-Hess, SETEL-Hess, and SL-Hess. The true posterior is added as the ground truth. All densities are scaled to ensure their integration equals 1.

4.4.2 g-and-k distribution

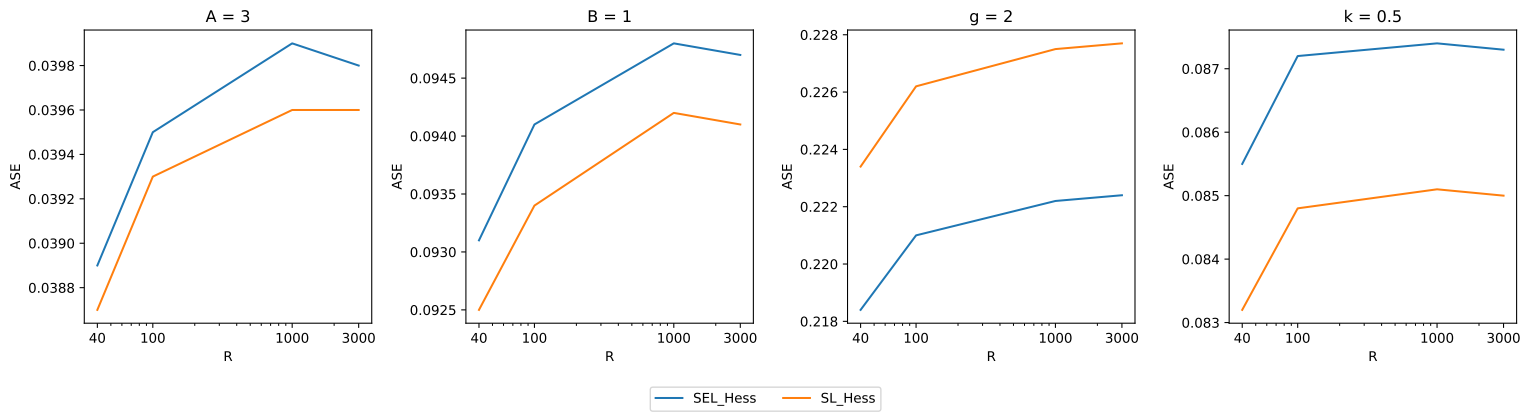
In this subsection, we assess the performance of our proposed methods in the g-and-k distribution (Haynes et al., 1997), which is more complex than the univariate normal distribution discussed earlier. With its typically intractable likelihood often requiring the use of ABC algorithms, the g-and-k model is a widely used benchmark in many ABC studies. The data from the g-and-k distribution x can be generated as

$$x = A + B \cdot \left[1 + \frac{4}{5} \times \frac{1 - \exp(-g \cdot z)}{1 + \exp(-g \cdot z)} \right] \cdot (1 + z^2)^k \cdot z, \quad (4.65)$$

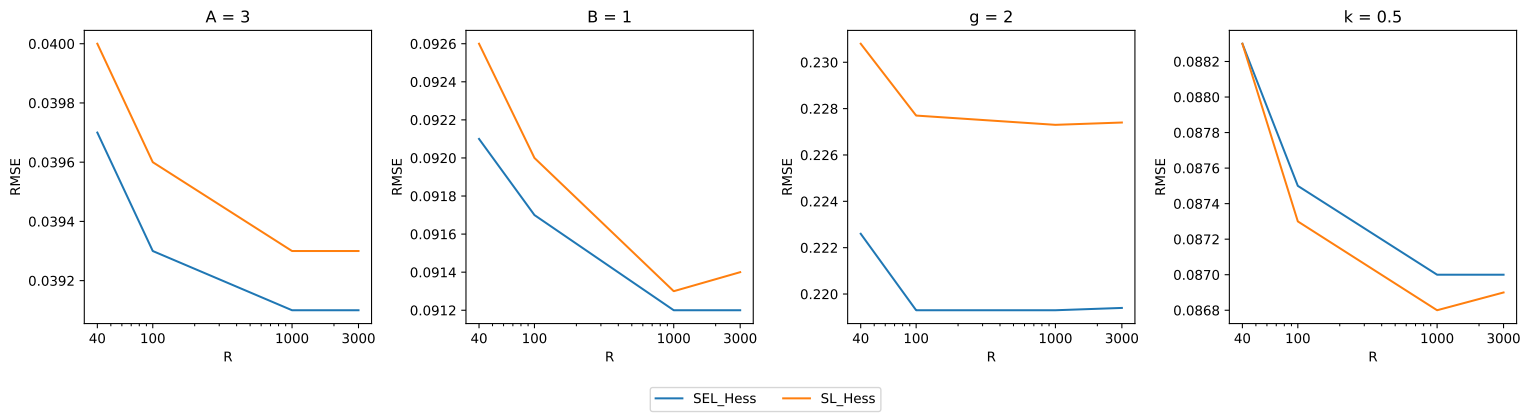
where $\boldsymbol{\theta} = (A, B, g, k)$ is the parameter vector related to location, scale, skewness and kurtosis of the distribution, z is a random variable $z \sim N(0, 1)$. An observed dataset of size $N = 1000$ was simulated from the distribution with $\boldsymbol{\theta}_0 = (A, B, g, k) = (3, 1, 2, 0.5)$. Four summary statistics are considered: mean, 25%, 50%, and 75% quantiles.

We assess nine of our proposed methods, also used in the just-identified normal example: three SEL-class methods, five SETEL-class methods, and one SL-Hess method, all implemented using frequentist inference. Figure 4.5 presents the average standard error (ASE), RMSE, and 95% CP for SEL-Hess and SL-Hess across $R = 40, 100, 1000, 3000$, based on $N_{\text{rep}} = 1000$ repetitions for each R . We only report SEL-Hess and SL-Hess in the figure, as all results from SEL-class and SETEL-class methods are identical to four decimal places. Overall, the SEL-Hess and SL-Hess methods perform similarly and reasonably across these three metrics. For a fixed method, when $R = 1000$ and $R = 3000$, the results are almost indistinguishable, as indicated by the y-axis values, showing evidence of convergence.

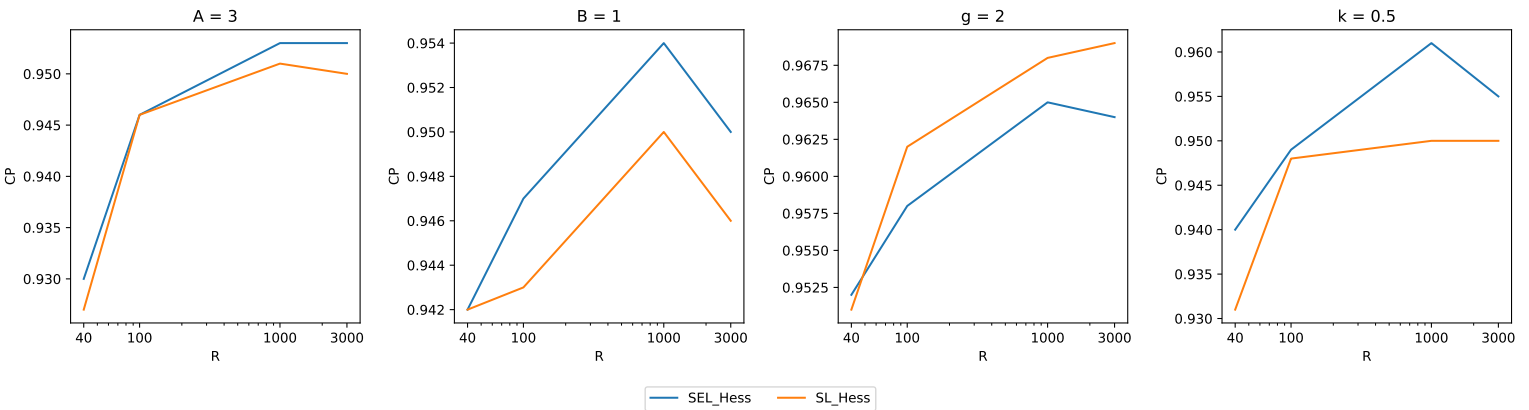
As to the speed comparison, SL-Hess and SEL-Hess requires approximately 1.5 minutes and 3 minutes, respectively, for the calculation of point and variance estimates at $R = 3000$. In contrast, BSL-MCMC needs around 40 minutes for 10,000 MCMC iterations, considerably slower than our proposed SL-Hess.



(a) ASE for $R = 40, 100, 1000, 3000$.



(b) RMSE for $R = 40, 100, 1000, 3000$.



(c) CP for $R = 40, 100, 1000, 3000$.

Figure 4.5: The ASE, RMSE, and 95% CP of SEL-Hess and SL-Hess for the g -and- k distribution across $R = 40, 100, 1000, 3000$, based on $N_{\text{rep}} = 1000$ repetitions for each R . The true values of parameters used to generate the observed data are $(A, B, g, k) = (3, 1, 2, 0.5)$.

4.4.3 A stochastic volatility model

While stochastic volatility (SV) models offer numerous attractive attributes, parameter inference on SV models poses challenges due to the inherent complexity of assessing its likelihood function. Comprehensive reviews of SV models are found in [Taylor \(1994\)](#) and [Shephard and Andersen \(2009\)](#). Several ABC methods are considered for SV models, including ABC auxiliary particle filter ([Vankov and Ensor, 2014](#)), auxiliary likelihood-based ABC ([Martin et al., 2019](#)), ABC with the Wasserstein distance ([Bernton et al., 2019](#)). Compared to the first two examples, the SV model involves a far more complex data-generating mechanism and summary statistics. We evaluate the performance of our proposed methods on this challenging model.

The SV model with 5 parameters $\boldsymbol{\theta} = (\alpha, \beta, \phi, \sigma, \rho)$ assumes that the observations $y_{1:N} = (y_1, \dots, y_N)$ come from

$$\begin{aligned} y_t &= \alpha + \beta \exp\left(\frac{z_t}{2}\right)w_t, \\ z_t &= \phi z_{t-1} + \sigma v_t, \end{aligned} \tag{4.66}$$

where $t = 1, \dots, N$, $w_t, v_t \sim N(0, 1)$ are noises with the correlation ρ ($\text{Corr}(w_t, v_t) = \rho$), z_t is the latent variable which is unobserved. The stationary assumption requires $|\phi| < 1$ and $z_0 \sim N(0, \sigma^2/(1 - \phi^2))$. The correlation ρ is commonly negative as an increase in predicted volatility tends to be associated with falls in the stock price ([Harvey and Shephard, 1996](#)). To the best of our knowledge, no prior literature has investigated the estimation and coverage of all these five parameters together via a simulation study.

The sample size of $y_{1:N}$ is set as $N = 2000$. The true values of $(\alpha, \beta, \phi, \sigma, \rho)$ are $(0, 0.2, 0.95, 0.9, -0.5)$ based on [Ahsan \(2021\)](#) and [Harvey and Shephard \(1996\)](#). The 5 summary statistics $(\tilde{\alpha}, \tilde{\beta}, \tilde{\phi}, (\tilde{\sigma})^2, \tilde{\rho})$ are calculated as follows:

1. $\tilde{\alpha}$: The median of the observed data, $\tilde{\alpha} = \text{median}(y_{1:N})$.
2. $\tilde{\beta}$: Defined as $\tilde{\beta} = \exp(\bar{r} + 1.2704)$, where $\bar{r} = \frac{1}{N} \sum_{t=1}^N r_t$ is the mean of r_t and $r_t = \log [(y_t - \tilde{\alpha})^2]$.
3. $\tilde{\phi}$: Defined as $\tilde{\phi} = \frac{\sum_{j=1}^J \hat{c}_r(j)\hat{c}_r(j+1)}{\sum_{j=1}^J \hat{c}_r(j)^2}$, where $J = 15$ and $\hat{c}_r(j)$ is the lag j autocovariance of r_t , given by $\hat{c}_r(j) = \frac{1}{N-j} \sum_1^{N-j} (r_t - \bar{r})(r_{t+j} - \bar{r})$.
4. $(\tilde{\sigma})^2$: Defined as $(\tilde{\sigma})^2 = \left[1 - (\tilde{\phi})^2\right] \cdot \left[\hat{c}_r(0) - \frac{\pi^2}{2}\right]$.

5. $\tilde{\rho}$: Defined as $\tilde{\rho} = \tilde{b}/\tilde{\sigma}$, where \tilde{b} is the ordinary least square (OLS) estimator of the slope term b in the simple linear regression $\eta_t = a + b\sigma_t$. Here $\eta_t = r_t - \phi r_{t-1}$, with ϕ being potentially replaced with $\tilde{\phi}$, and s_t represents the sign of y_t , with $s_t = 1$ if $y_t > 0$ and $s_t = -1$ if $y_t < 0$.

The choice of these five summary statistics is motivated by their ability to capture key parameters of interest. Specifically, the median of the data y_t contains information of α . [Ahsan \(2021\)](#) showed that $(\tilde{\beta}, \tilde{\phi}, (\tilde{\sigma})^2)$ are consistent estimators of (β, ϕ, σ^2) . Intuitively, higher J in $\tilde{\phi}$ provides a more accurate point estimate of ϕ , but requires more computational resources. The empirical experiments in [Ahsan \(2021\)](#) recommended $J = 15$ in $\tilde{\phi}$ for a balance between accuracy and computational cost. [Harvey and Shephard \(1996\)](#) suggested that $\tilde{\rho}$ carries sufficient information of ρ as the authors proved that $b = \rho\sigma\sqrt{2/\pi}$.

Unlike the univariate normal and the g-and-k-distribution examples, numerical differentiation ([Richard and Burden, 2001](#)) is employed in the SV model to compute the Hessian and gradient values.

4.4.3.1 Automatic Differentiation vs Numerical Differentiation

In the SV model, unlike the univariate normal and g-and-k distribution examples, we employ numerical differentiation to compute Hessian and gradient values. Before presenting the SV model results, we first discuss the choice between two derivative-computation tools: automatic differentiation (AD) and numerical differentiation (ND).

To illustrate this, we begin by examining the logSL function with respect to the parameters, given simulated noise and an observed dataset. Figure 4.6 displays projection plots of the logSL functions for the transformed parameter vector $\boldsymbol{\xi}$ across 100 equally-spaced grid points, within the interval $[\hat{\boldsymbol{\xi}}_{\text{SL}} \pm \delta \cdot \text{SE}(\hat{\boldsymbol{\xi}}_{\text{SL}})]$. Here, $\hat{\boldsymbol{\xi}}_{\text{SL}}$ is the SL point estimate, with $\boldsymbol{\xi}$ being a suitable transformation of $\boldsymbol{\theta}$, and $\text{SE}(\hat{\boldsymbol{\xi}})$ is the standard error given by SL-Hess using ND. The transformation $\boldsymbol{\xi} = (\alpha^*, \beta^*, \phi^*, \sigma^*, \rho^*) = (1000 \cdot \alpha, \log(\beta), \text{logit}(\rho), \log(\sigma), \tan(\rho \cdot \pi/2))$ is to facilitate unconstrained optimization. Each panel displays the SL function of one specific element of $\boldsymbol{\xi}$ with the red vertical line marking its SL point estimate, while all other elements are fixed at their respective SL point estimates. Subfigures 4.6a, 4.6b, and 4.6c correspond to $\delta = 2, 0.5, 0.1$, respectively.

From Figure 4.6, it is found that when the grid spacing, the space between two nearby grid points, is wide (i.e., $\delta = 2$), the logSL function appears to be continuous. However,

when the grid spacing is narrow (i.e., $\delta = 0.5, 0.1$), the logSL function is indeed discontinuous. This behavior is due to the discontinuity in one of the summary statistics, the median function $\text{median}(y_{1:N})$.

The visualization of the logSL function helps the discussion of AD and ND. AD is a highly accurate and efficient technique, that systematically applies the chain rule to compute exact derivatives of functions by breaking them down into elementary operations. When applied to the discontinuous logSL function, AD provides an accurate Hessian matrix, but the result is not positive definite, making it unsuitable and unstable for standard error calculation.

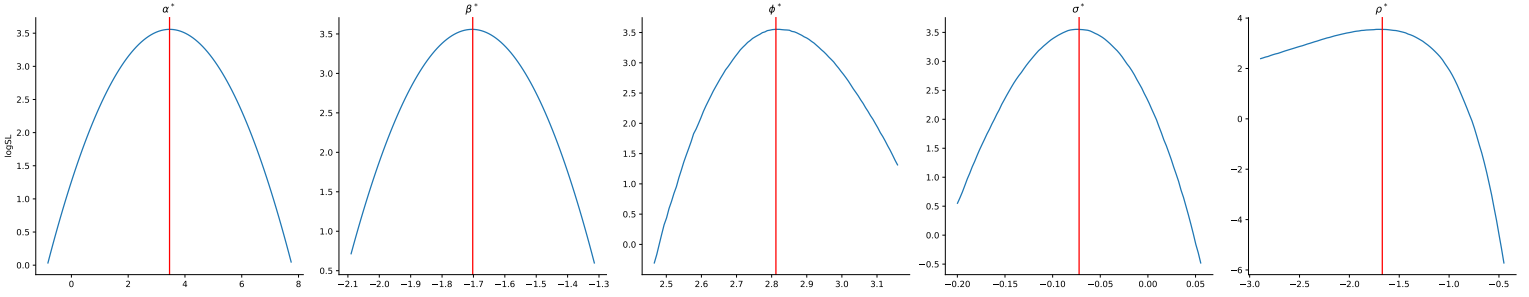
In contrast, ND computes derivatives using finite difference approximations, estimating the derivative of a function by evaluating it at nearby points. For the discontinuous logSL function, ND evaluates the function at discrete points, making it more robust in non-smooth regions and capable of producing a positive definite Hessian matrix. Using the Python package `numdifftools` (Brodtkorb and D’Errico, 2015), our simulation results show that ND yields reliable and stable results in the SV model example.

Last, we comment that the SELEL, SEL, and SL functions in this example all exhibit varying degrees of irregularities (i.e., unsmoothness and discontinuity). Among them, we choose the SL function for illustration as it shows the most pronounced issues. For comparison, in the simpler g-and-k model, the SL function uses quantile-based summary statistics, which are also discontinuous. However, no noticeable irregularities in the logSL function were observed, even with narrow grid spacing. Therefore, we suspect that the combination of a complex data-generating model and discontinuous summary statistics may lead to irregularities in the SETEL, SEL, and SL functions. However, the exact level of complexity in the data-generating model required for this to occur remains uncertain.

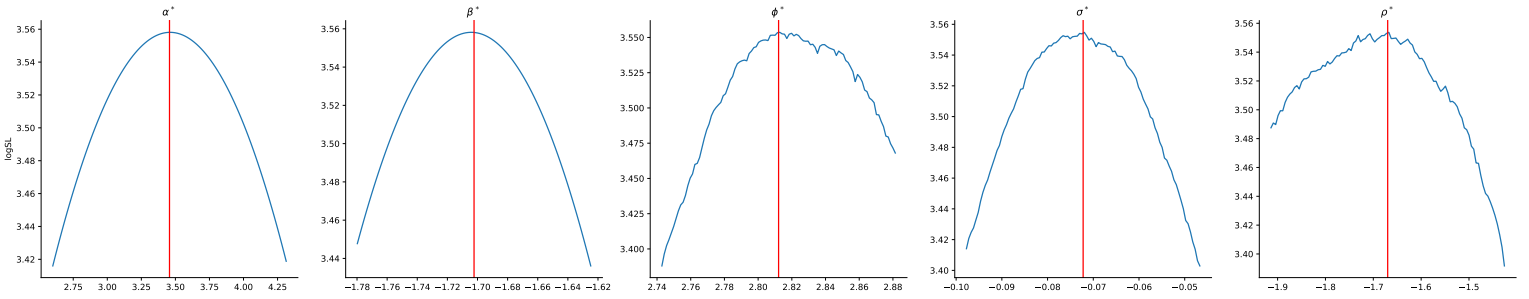
To summarize, when irregularities are absent, both AD and ND are suitable tools for implementing our methods, with AD being faster and more accurate. However, when irregularities are present, ND is more robust and delivers stable results, whereas AD struggles with unreliable standard error calculations. A practical approach is to start with AD due to its speed. If the Hessian matrix at the point estimate given by AD is not positive definite, ND may be used instead for the reasons discussed.

4.4.3.2 SV model results

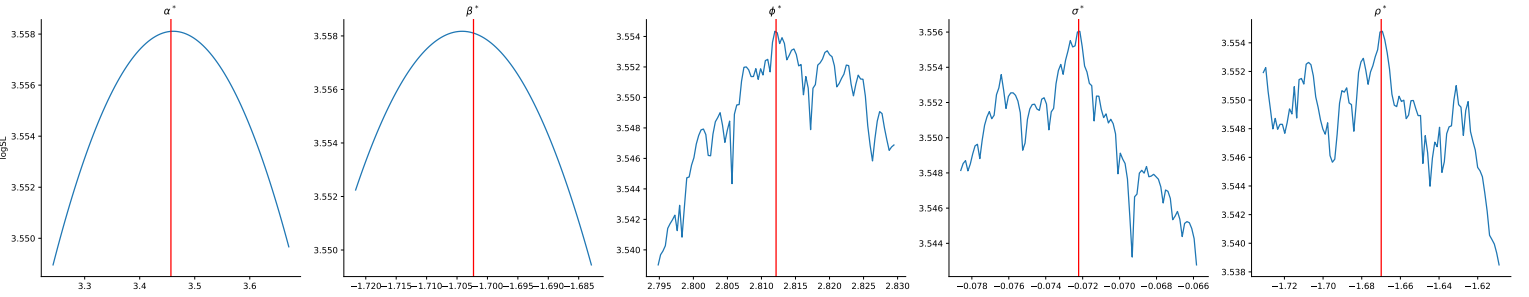
Figure 4.7, 4.8, 4.9 display the ASE, RMSE, 95% CP of the same methods used in the g-and-k distribution, applied to the SV model across $R = 100, 500, 1000$, based on $N_{\text{rep}} = 1000$ for each R . Overall, the SEL-class, SETEL-class methods, and SL-Hess have reasonable



(a) The logSL function evaluated within the interval $[\hat{\xi}_{\text{SL}} \pm \delta \cdot \text{SE}(\hat{\xi}_{\text{SL}})]$, where $\delta = 2$.



(b) The logSL function evaluated within the interval $[\hat{\xi}_{\text{SL}} \pm \delta \cdot \text{SE}(\hat{\xi}_{\text{SL}})]$, where $\delta = 0.5$.



(c) The logSL function evaluated within the interval $[\hat{\xi}_{\text{SL}} \pm \delta \cdot \text{SE}(\hat{\xi}_{\text{SL}})]$, where $\delta = 0.1$.

Figure 4.6: The logSL function with respect to $\xi = (\alpha^*, \beta^*, \phi^*, \sigma^*, \rho^*)$ in the SV model, generated by 100 equally-spaced grid points, within the interval $[\hat{\xi}_{\text{SL}} \pm \delta \cdot \text{SE}(\hat{\xi}_{\text{SL}})]$ via SL-Hess and ND. Here ξ is a suitable transformation of θ and $\delta = 2, 0.5, 0.1$. Each panel displays the SL function of one specific element of ξ with the red vertical line marking its SL point estimate, while all other elements are fixed at their respective SL point estimates.

and comparable performance, as indicated by the narrow range of the y-axis. At $R = 1000$, all ASEs show convergence and all CPs are close to the nominal level of 95%. In the figure of RMSE, the lines for SETEL-Hess (red lines) are invisible in the figure because they overlap with the lines for SEL-Hess. Regarding the parameters β and ρ , SL-Hess achieves slightly lower ASE, RMSE, and CP than SEL and SETEL methods.

The RMSE of the SEL estimates is comparable to the RMSE reported in [Ahsan \(2021\)](#), despite being based on different observed datasets. For a sample size $N = 2000$, known parameters ($\alpha = 0, \rho = 0$), and the same true values ($\beta = 0.2, \phi = 0.95, \sigma = 0.9$), the simulation study in [Ahsan \(2021\)](#) reported RMSEs of (0.0415, 0.0103, 0.0602) for parameters (β, ϕ, σ) , respectively. However, [Ahsan \(2021\)](#) did not provide the variance estimates and coverage assessments, whereas our proposed methods are able to quantify uncertainties of the parameter estimates.

We then compare the computational speed of the traditional MCMC method and our proposed method. At $R = 1000$, SL-Hess requires approximately 15 minutes, involving around 1000 function evaluations (250 for optimization and 750 for the numerical Hessian calculation). In contrast, BSL-MCMC requires approximately 150 minutes for 10,000 MCMC iterations, making it roughly 10 times slower.

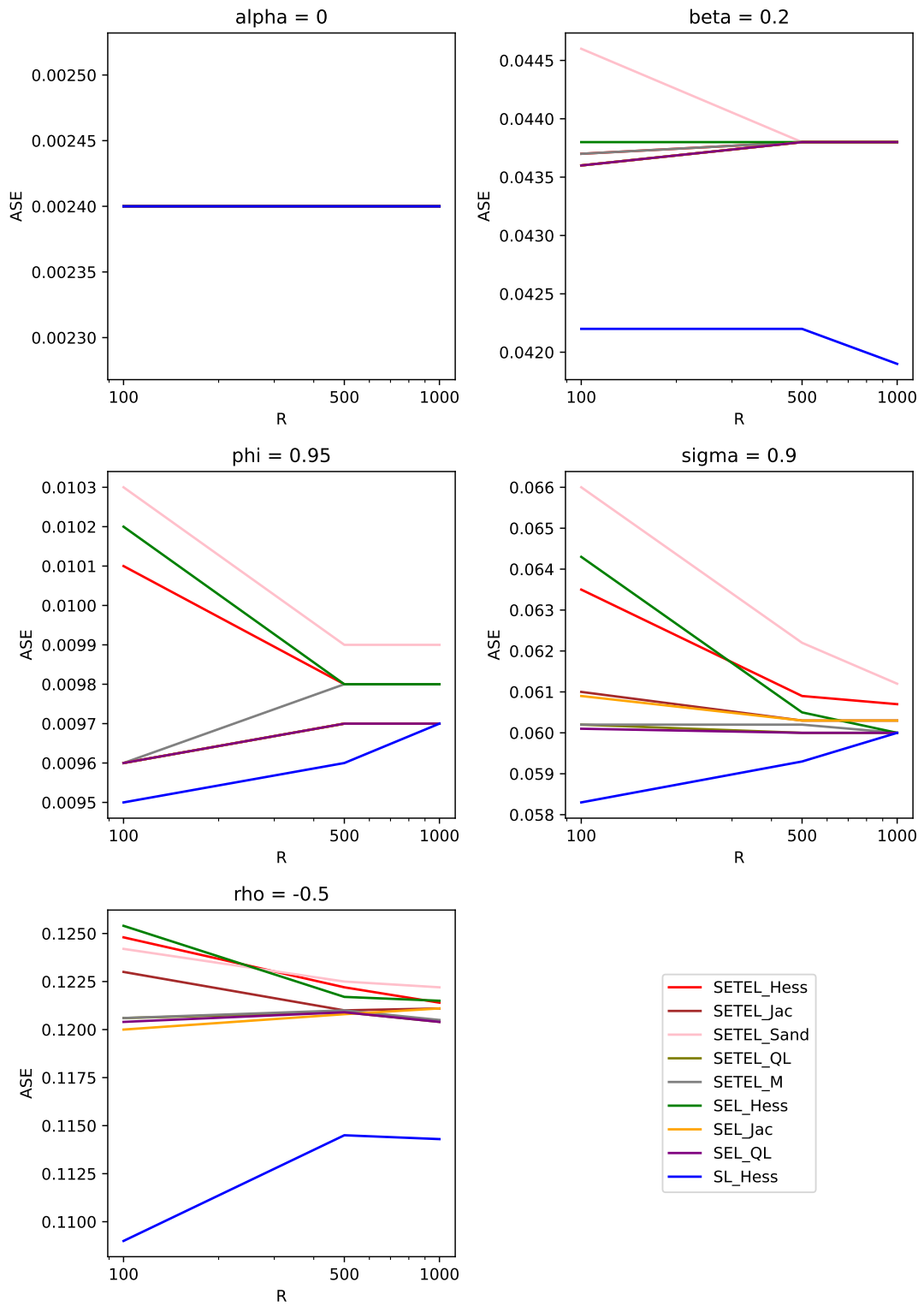


Figure 4.7: The ASE of various proposed methods of the SV model at $R = 100, 500, 1000$ in 1000 repetitions.

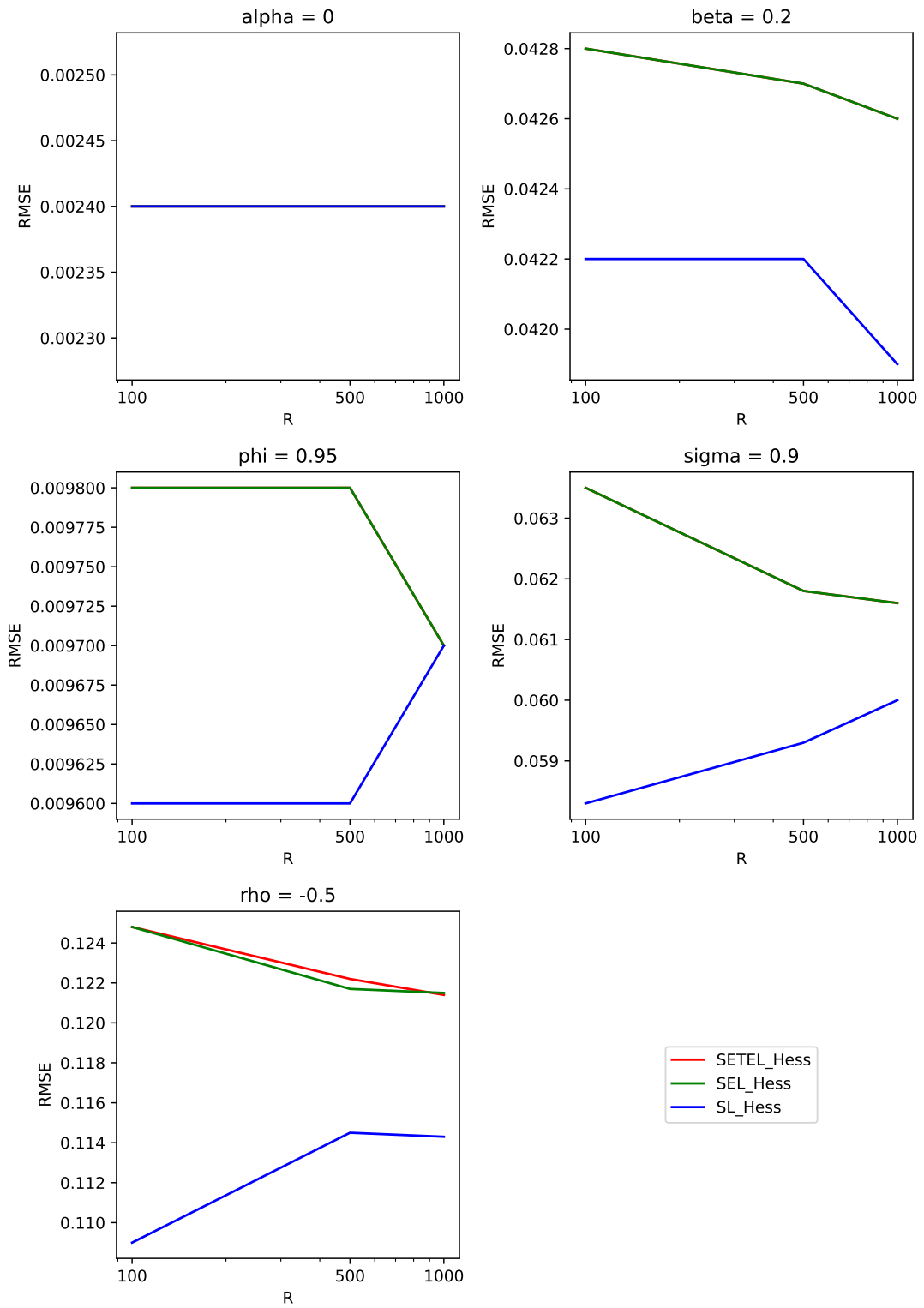


Figure 4.8: The RMSE of various proposed methods of the SV model at $R = 100, 500, 1000$ in 1000 repetitions. Where the red lines (SETEL-Hess) are not visible in the figure, they coincide exactly with the green lines (SEL-Hess).

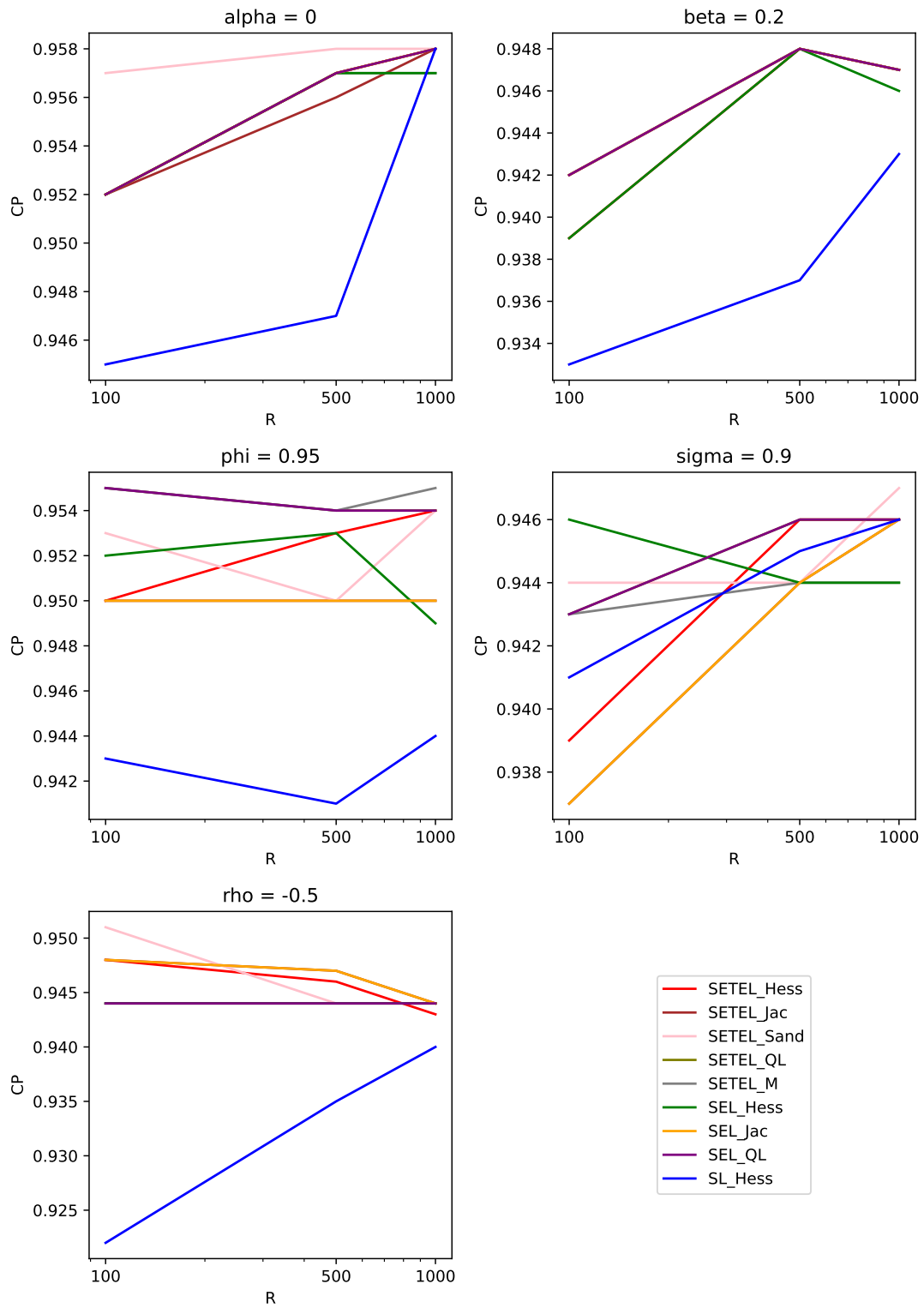


Figure 4.9: The 95% CP of various proposed methods of the SV model at $R = 100, 500, 1000$ in 1000 repetitions.

4.4.4 Univariate normal distribution under over-identification

All previous examples evaluate the just-identification where the number of summary statistics m equals the number of the parameter p . From this subsection onward, we investigate the over-identification, in which summary statistics outnumber parameters (i.e., $m > p$). Under this circumstance, the moment condition of SEL cannot be satisfied by any θ value for almost every fixed \mathbf{s}_N . The goal of this example is to assess the performance of our proposed methods under over-identification, using a simple model with a tractable likelihood. It is worth noting that, unlike in the just-identified case, the asymptotic properties of the SETEL estimator under over-identification remain unknown.

We use a toy example for better illustration of over-identification. Consider an observed dataset $x_{1:N}$ from $N(\theta, 1)$ with an unknown parameter θ . Under just-identification, supposing the mean of the observed dataset is 1.1 (i.e., $\text{mean}(x_{1:N}) = 1.1$), it is easily found that $\theta = 1.1$ satisfies the moment condition (4.31). However, under over-identification, supposing the mean of the observed dataset is still 1.1 and the median of the observed dataset is 0.8 (i.e., $\text{mean}(x_{1:N}) = 1.1$, $\text{median}(x_{1:N}) = 0.8$), there does not exist any θ value to let the mean of $N(\theta, 1)$ equal to 1.1 and the median of $N(\theta, 1)$ equal 0.8 simultaneously, as the normal distribution is symmetric. This is how the moment condition in the over-identification for a fixed observed dataset cannot be satisfied by any parameter value.

The simulation settings closely resemble those in the just-identified case, based on the univariate normal $x \sim \text{Normal}(\theta, 1)$ and the true value $\theta_0 = 0$. The differences are (i) three summary statistics are used here: mean, 25% quantile, and 75% quantile values of the observed data $x_{1:N}$ ($m = 3 > p = 1$), (ii) the sample sizes $N = 25$ and $N = 50$ of the observed dataset are both assessed, (iii) methods are implemented using frequentist inference as MCMC is not needed for comprehensive comparison. The motivation for exploring different sample sizes is that the moment condition is satisfied when $N \rightarrow \infty$. So the performance discrepancy between SETEL and SEL may increase when N decreases.

A variety of methods are considered, including the parametric method, our proposed SEL-class methods, SETEL-class methods, and SL-Hess method. The parametric method uses maximum likelihood estimator (MLE) with the Hessian to compute standard errors and is regarded as the ground truth. Compared to the just-identified case, SETEL-QL is not included here because the QL method is not valid for ETEL when the moment condition cannot be satisfied. Since SETEL is based on ETEL, we suspect that SETEL-QL is also invalid in the over-identified case.

Figure 4.10 displays the ASE, RMSE, 95% CP of various methods with the sample size of $N = 25, 50$ respectively. The numbers of the simulated datasets considered are

$R = 25, 100, 1000, 10000, 20000$, with repetitions $N_{\text{rep}} = 10,000$ for each R value. For $N = 25$, the ASE plot shows that the ASEs of SEL-Hess, SEL-Jac, and SETEL-M deviate increasingly from the ASE of the parametric method as R increases, indicating a problematic property. The RMSE plot reports that all SEL, SETEL, SL point estimates converge to the parametric point estimate (i.e., MLE), with SL-Hess converging the fastest. The possible reason is that the chosen summary statistics of normally distributed data easily satisfy the normality assumption of SL. Regarding CP, at $R = 20000$, SETEL-M suffers from over-coverage while SEL-Jac and SEL-QL exhibits under-coverage. All other methods show convergence to the parametric CP, around the nominal level of 95%.

For $N = 50$, the results are similar to those for $N = 25$, but the performance discrepancy between SETEL-class and SEL-class methods is less pronounced at $N = 50$, especially for CP. Intuitively, under over-identification, as N increases, the performance of SETEL-class and SEL-class methods becomes closer. When $N \rightarrow \infty$, the moment condition can be satisfied and their performance becomes identical.

Overall, all SEL methods are unsuitable for the over-identified case. Specifically, SEL-Hess and SEL-Jac display increasing ASE as R grows, while SEL-Jac and SEL-QL show under-coverage as R increases. In contrast, the proposed SL-Hess and all SETEL-class methods, except SETEL-M, converge to the parametric results as the computational resource R increases, suggesting that they are appropriate statistical tools for parameter inference in the presence of over-identification.

To show the speed advantage of our proposed methods over MCMC, we consider an observed dataset with $N = 50$ and $R = 10000$. SL-Hess requires approximately 1.5 seconds for the calculation of point and interval estimates while it takes BSL-MCMC 92 seconds to obtain 10,000 MCMC iterations, demonstrating a significantly slower performance.

Similar to Section 4.4.1, we now implement Bayesian inference to study the posterior densities obtained from these methods using the prior $\boldsymbol{\theta} \propto 1$. For a particular observed dataset of $N = 50$, Figure 4.11 presents the density plots of, the true posterior based on the parametric likelihood, the SEL (SETEL, SL) posterior, and the BNA via SEL-Hess (SETEL-Hess, SL-Hess), at $R = 50, 100, 1000, 10000$. All densities are scaled so that the integral over the entire space equals one. As R increases, the densities of both SEL-Hess and its BNA flatten and deviate further from the true posterior density. This phenomenon aligns with the observed increase in ASE given by SEL-Hess in Figure 4.10. In contrast, the densities of both the SETEL and SL posteriors, along with their BNA, converge toward the true posterior density as R grows. For this particular observed dataset, the SETEL and its BNA require around $R = 10000$ to closely approximate the true posterior, whereas SL and its BNA achieve similar accuracy with only $R = 1000$.

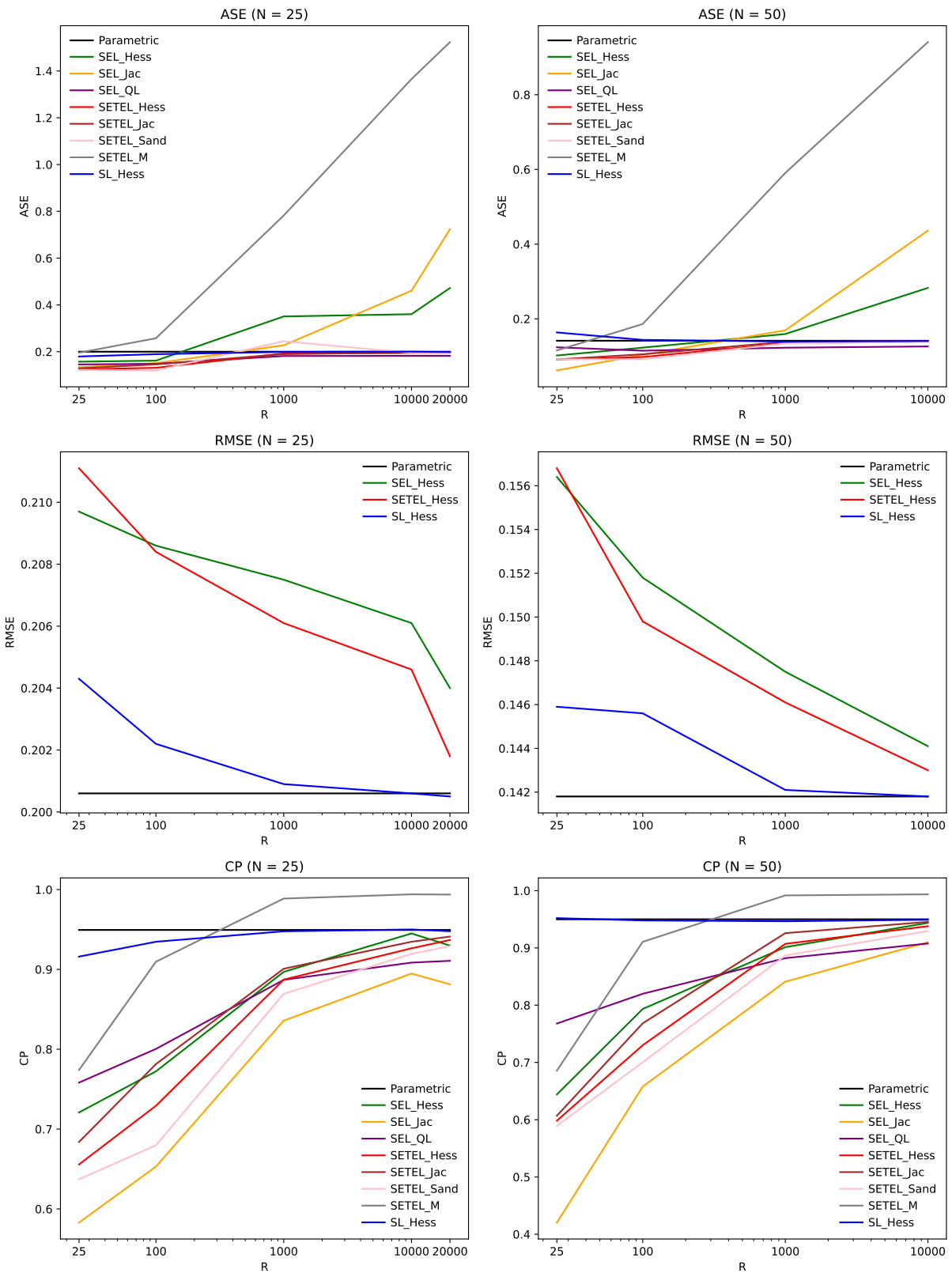


Figure 4.10: The ASE, RMSE and 95% CP of various methods in the over-identified normal example and 10,000 repetitions. The data comes from $N(\theta, 1)$ with the true value $\theta_0 = 0$ and the sample size $N = 25, 50$.

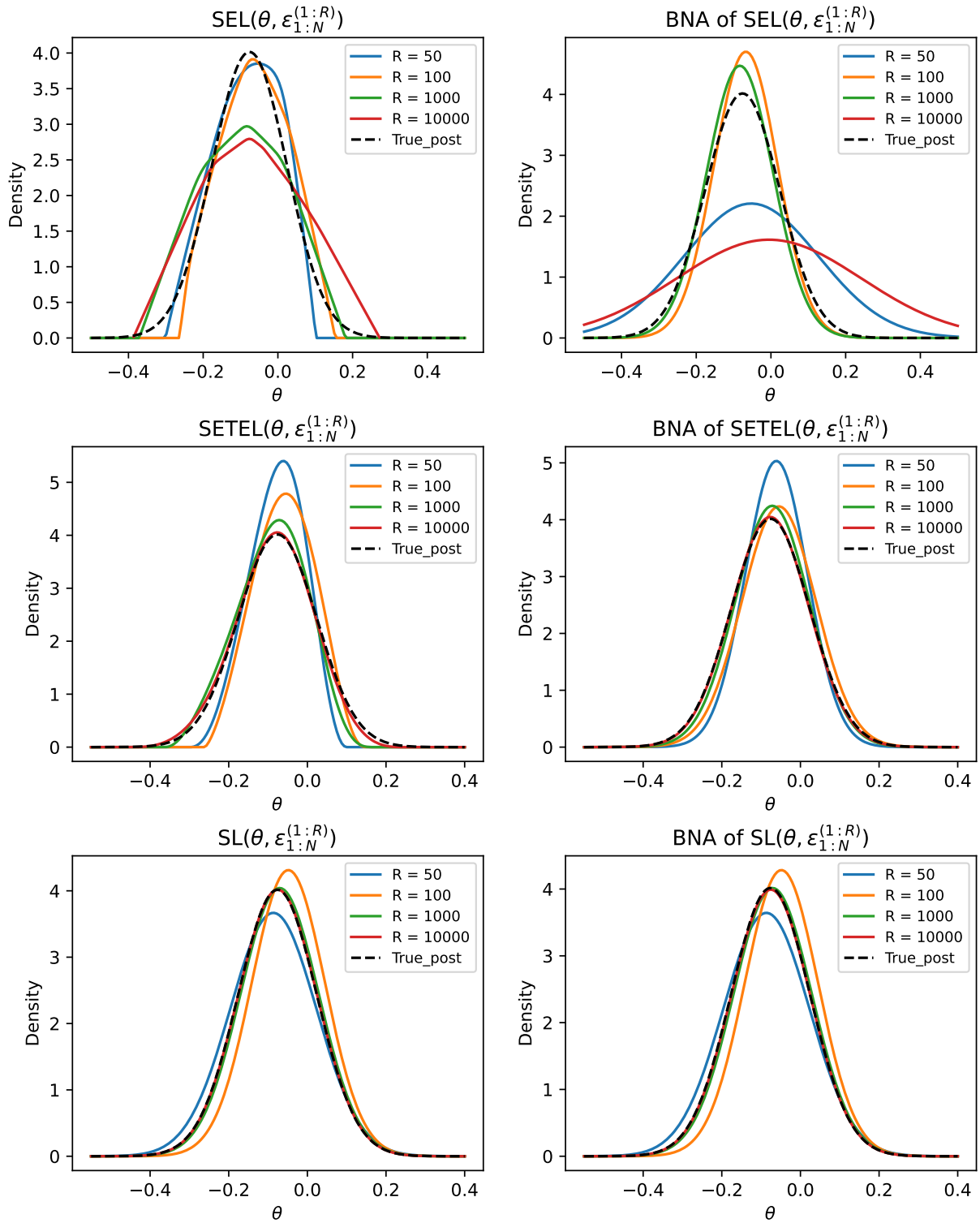


Figure 4.11: The left three panels display the SEL, SETEL, and SL posterior densities in the over-identified normal example, while the right three panels present the corresponding BNA densities based on SEL-Hess, SETEL-Hess, and SL-Hess. The true posterior density is added as the ground truth. All densities are scaled to ensure their integration over whole space equals 1.

4.4.5 An M/G/1 queuing model

The M/G/1 queuing model also has intractable likelihood but is easy to simulate from, which was previously explored by several ABC literature (Blum and François, 2010; Fearnhead and Prangle, 2012; Bernton et al., 2019; An et al., 2020). An et al. (2020) pointed out that SL has poor performance in this model due to its non-normally distributed summary statistics. The aim of this simulation study is to compare the performance of SETEL and SL when the distribution of the summary statistics is non-normal.

The M/G/1 model assumes that customers arrive at a server with the independent inter-arrival times $w_t \sim \exp(\lambda)$, exponentially distributed with rate λ . The service time for each customer u_t independently follows a uniform distribution in the interval $[\theta_1, \theta_2]$, $u_t \sim \text{Unif}(\theta_1, \theta_2)$. The observations are the inter-departure times $y_{1:N}$, which are calculated/generated as

$$y_t = u_t + \max\left(0, \sum_{j=1}^t w_j - \sum_{j=1}^{t-1} y_j\right), \quad t = 1, \dots, N. \quad (4.67)$$

As used in Blum and François (2010), we set the sample size $N = 50$, the true value of parameters $(\theta_1, \theta_2, \lambda) = (1, 5, 0.2)$, and 5 summary statistics, which are 25% quantile, 50% quantile, 75% quantile, minimum, maximum of $y_{1:N}$. Figure 4.12 presents the densities of the 5 summary statistics based on 10000 simulated datasets, which indicates that 75% quantile value of $y_{1:N}$ is extremely non-normal. Reparametrization trick is applied to exponential distributions by using inverse transform sampling. That is, if $\epsilon \sim \text{Unif}(0, 1)$, then x follows an exponential distribution with rate λ , $x = -\lambda^{-1} \log(1 - \epsilon) \sim \exp(\lambda)$. Four methods which performed well in the over-identified normal example are assessed: SETEL-Hess, SETEL-Jac, SETEL-Sand, and SL-Hess. Similar to the SV model example, numerical differentiation is employed in the M/G/1 model to compute the Hessian and gradient values.

Figure 4.13 summarizes the ASE, RMSE, and 95% CP, respectively, for the four considered methods, at $R = 20000, 40000, 80000$. The repetitions are $N_{\text{rep}} = 1000$ for each R value. At $R = 80000$, for θ_1 , the RMSE of the SETEL point estimate is slightly higher than that of the SL point estimate. However, for θ_2 and λ , the RMSEs of the SETEL point estimates are significantly lower than those of SL, indicated by the range of the y-axis of the panels. Regarding CP, SETEL-Hess achieves CP close to the nominal level of 95% for all three parameters. In contrast, SL-Hess shows clear under-coverage for θ_2 and λ . Additionally, SETEL-Sand exhibits high ASE and corresponding over-coverage, while SETEL-Jac has low ASE and under-coverage. Overall, when the summary statistics are not normally

distributed, SETEL yields more accurate point estimates than SL. Also, SETEL-Hess provides reasonable ASE and CP, whereas SL-Hess suffers from under-coverage.

Similar to the SV model, the logSETEL function in the M/G/1 queuing model also exhibits unsmoothness and sharp transitions (i.e., irregularities) when evaluated with very narrow grid spacing. This behavior is potentially caused by the complexity of the data-generating process and the use of discontinuous quantile summary statistics. Figure 4.14 displays the logEDEL functions $\boldsymbol{\xi} = (\theta_1, \theta_2, \log(\lambda))$ across 100 equally-spaced grid points within the interval of $\left[\hat{\boldsymbol{\xi}}_{\text{SETEL}} \pm \delta \cdot \text{SE}(\hat{\boldsymbol{\xi}}_{\text{SETEL}}) \right]$, where $\hat{\boldsymbol{\xi}}_{\text{SETEL}}$ is the SETEL point estimate, and $\text{SE}(\hat{\boldsymbol{\xi}}_{\text{SETEL}})$ is the standard error given by SETEL-Hess. Subfigures 4.14a, 4.14b correspond to $\delta = 0.2, 0.02$, respectively. Each panel varies one parameter while holding all others constant at their respective SETEL point estimates. The red vertical line indicates the SETEL point estimate of the parameter being evaluated.

When the grid spacing is relatively wide ($\delta = 0.2$), the logSETEL function of each parameter appears smooth and regular. However, with a narrower grid spacing ($\delta = 0.02$), the logSETEL function exhibits unsmoothness and sharp transition. This behavior can result in high gradients, corresponding to the low ASE observed for SETEL-Jac and high ASE observed for SETEL-Sand, as shown in Figure 4.13. Nevertheless, our simulation study results suggest that SETEL-Hess is more robust and stable for standard error calculation under such situation and across all five evaluated examples.

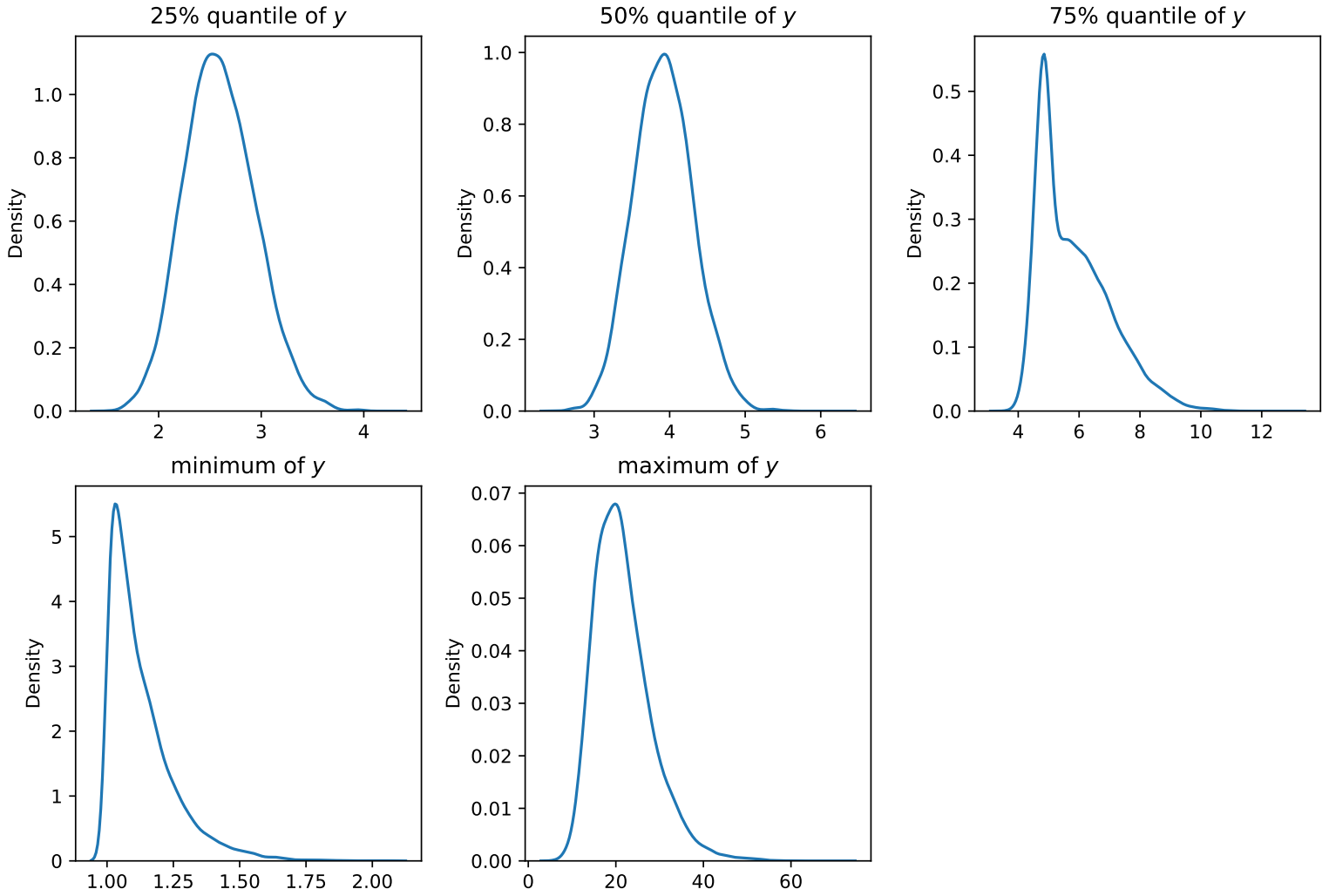


Figure 4.12: Densities of 5 summary statistics from the M/G/1 model with $(\theta_1, \theta_2, \lambda) = (1, 5, 0.2)$ based on 10000 simulated datasets: 25% quantile, 50% quantile, 75% quantile, minimum, maximum of $y_{1:N}$.

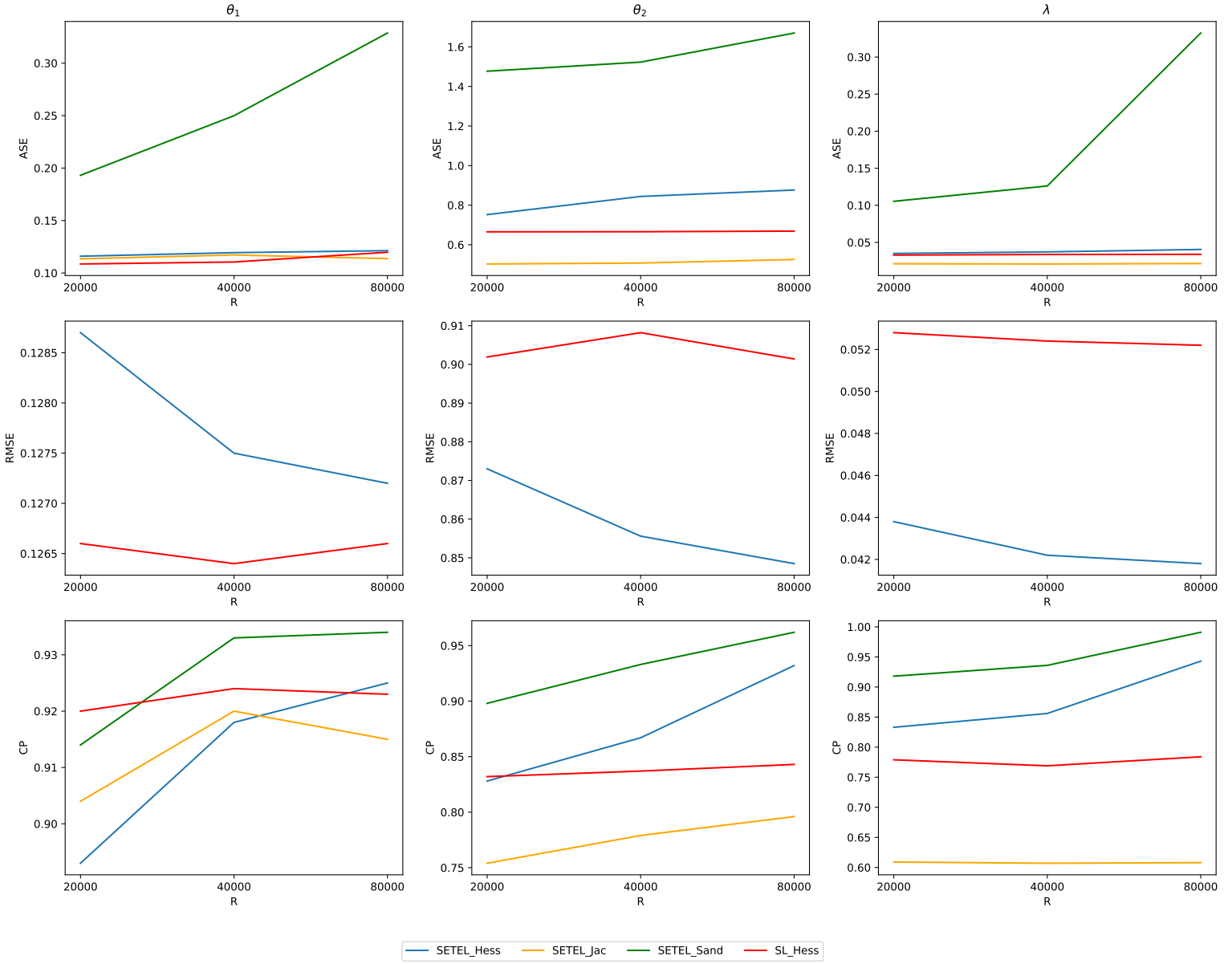
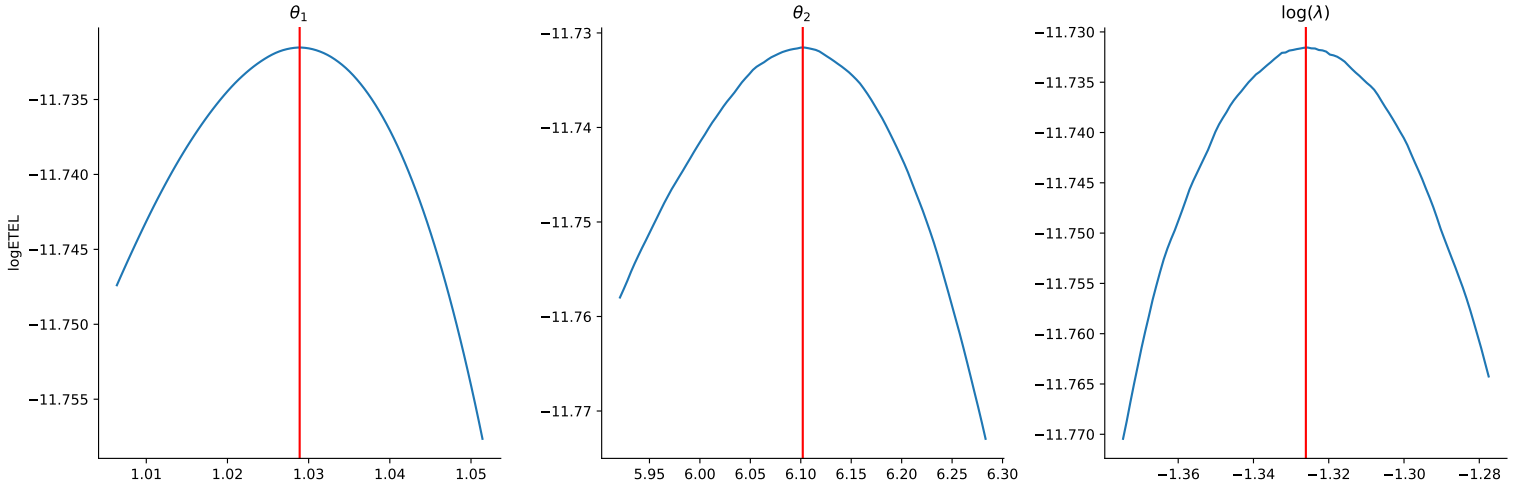
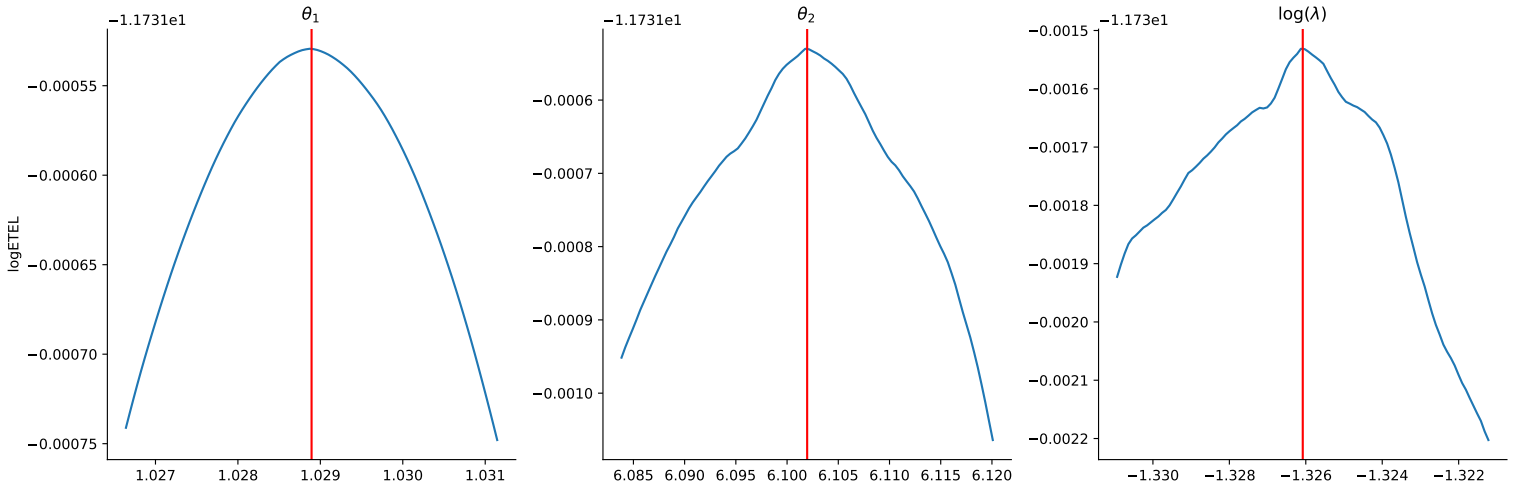


Figure 4.13: The ASE, RMSE, and 95% CP of four proposed methods evaluated at $R = 20000, 40000, 80000$ across $N_{\text{rep}} = 1000$ repetitions in the M/G/1 model.



(a) The logSETEL function evaluated within the interval $\left[\hat{\xi}_{\text{SETEL}} \pm \delta \cdot \text{SE}(\hat{\xi}_{\text{SETEL}}) \right]$, where $\delta = 0.2$.



(b) The logSETEL function evaluated within the interval $\left[\hat{\xi}_{\text{SETEL}} \pm \delta \cdot \text{SE}(\hat{\xi}_{\text{SETEL}}) \right]$, where $\delta = 0.02$.

Figure 4.14: The logSETEL function of $\xi = (\theta_1, \theta_2, \log(\lambda))$ in the M/G/1 model, generated by 100 equally-spaced grid points, within the interval $\left[\hat{\xi}_{\text{SETEL}} \pm \delta \cdot \text{SE}(\hat{\xi}_{\text{SETEL}}) \right]$ via SETEL-Hess, where $\delta = 0.2, 0.02$. Each panel varies one parameter while holding all others constant at their respective SETEL point estimates. The red vertical line indicates the SETEL point estimate of the parameter being evaluated.

4.5 Conclusion

Our work in this chapter demonstrates that synthetic empirical likelihood (SEL) is a valuable approach for addressing intractable likelihood problems compared to ABC/BSL. We explore the asymptotic properties of SEL under just-identification (when the number of summary statistics m equals the number of parameters p , i.e., $m = p$), by introducing a new likelihood approximation approach. To address the limitations of MCMC in handling the full SEL pseudo-posterior, we propose several point and variance estimators via BNA and reparametrization trick, for better uncertainty quantification. Additionally, given a fixed observed dataset, under over-identification ($m > p$), we proposed a new approach, synthetic exponentially tilted empirical likelihood (SETEL), to address cases where the moment condition cannot be satisfied. The simulation study demonstrates that SETEL outperforms SEL in such cases.

In the simulation study, we compare various methods of point and variance estimators across three models under just-identification and two models under over-identification. The results deliver several key points for the just-identification case. For the three models with increasing level of complexity, our proposed SEL-class methods, SETEL-class methods, and SL-Hess perform well, regarding the performance metrics: ASE, RMSE, and CP, highlighting the robustness of our proposed methods. Furthermore, in each of the three examples, our proposed methods exhibit similar results, because they all share the same asymptotic properties: all point estimators consistently estimate the true value θ_0 , and their associated variance estimators, when multiplied by N , consistently estimate the variance \mathbf{V}_{EL} , as established in Theorem 1. In addition, MCMC for SEL (SEL-MCMC) produces inflated variance estimates and over-coverage as R increases, whereas our proposed methods maintain stable and more accurate variance estimates and coverage probabilities as R grows. While MCMC for SL (BSL-MCMC) also achieves reasonable results, it is substantially slower than our BNA-based methods for parameter inference.

Under over-identification, the simulation results point to additional suggestions. The SEL methods suffer from either exploded variance estimates or under-coverage. In contrast, SETEL-Hess, SETEL-Jac, and SETEL-Sand yield reasonable results in terms of the three performance metrics, with SETEL-Hess demonstrates greater robustness to the irregular shape of the SETEL function caused by discontinuous summary statistics. Moreover, when the normality assumption for the summary statistics is not accurate (e.g., in the M/G/1 example), the SETEL methods exhibit more accurate point estimates (as measured by RMSE) and variance estimates (as measured by ASE and CP) compared to SL.

Overall, the SETEL and SL methods are preferred over SEL under over-identification. The SETEL methods are preferred over SL when the normality assumption of summary

statistics are not appropriate. Notably, SETEL-Hess demonstrates strong performance across all five models in our simulation study and should be prioritized for practical use.

Several directions for future work are suggested. First, similar to the asymptotic Theorem 1, the asymptotic property of the SETEL estimator under over-identification, expressed by moment conditions, can be derived. Second, subgradient methods (Boyd and Mutapcic, 2008; Duchi et al., 2011) are designed for optimizing non-differentiable but convex functions. Given that our SEL, SETEL, and SL functions may exhibit similar behavior (e.g., in SV and M/G/1 queue models), advanced stochastic optimization approaches, such as Adagrad (Duchi et al., 2011), Adam (Kingma and Ba, 2014), and RMSprop (Hinton et al., 2012), based on subgradient, may be explored to further improve computational efficiency and stability. Last, connections between the ETEL and Kullback-Leibler (KL) divergence minimization, as discussed in Schennach (2007) and Chib et al. (2018), suggest potential similarities between SETEL and KL-based ABC approaches (Jiang, 2018), which indicate further investigation.

Chapter 5

Conclusion

This thesis provides a statistical investigation to the reliability analysis of wooden products and the likelihood-free methods based on empirical likelihood. The chapters cover a comprehensive examination of the Duration of Load (DOL) effect, damage assessment in proof-loading, and the development of advanced likelihood-free techniques for complex statistical models with intractable likelihood.

Chapter 2 introduces a comprehensive framework for evaluating lumber reliability under the DOL effect. We develop a multimodel Bayesian approach that integrates three existing DOL models: the US model, the Canadian model, and the Gamma process model. Each model offers unique insights into lumber strength degradation, but individually may not capture the full complexity of the DOL effect. To address this limitation, we implement Bayesian model-averaging (BMA), which combines reliability estimates from individual models while accounting for both model selection uncertainty and parameter uncertainty. The framework includes procedures for generating residential, snow, and wind load profiles. We demonstrate the effectiveness of this methodology through analysis of a Hemlock experimental dataset, presenting estimated reliability indices with 95% intervals bands.

In Chapter 3, we investigate the potential damage effect to lumber strength caused by proof-loading testing, a useful industrial procedure for screening out weak specimens. We introduce a statistical framework that characterizes the impact of proof-loading on the strength of surviving lumber specimens and detects significant damage among these survivors. Unlike the previously applied DOL, copula models, and equal-rank method, our approach directly measures the extent of the damage effect and rigorously evaluates its significance. In the real data analysis, the proposed framework successfully identifies the damage effect, quantifies parameter uncertainty, and demonstrates a strong fit to the

experimental data. A correlation estimate is also provided if two strength properties are involved in the proof-loading testing.

In Chapter 4, we explore statistical models defined by generative processes where likelihood functions are difficult or impossible to compute, requiring likelihood-free approaches. The Canadian model, discussed in Chapter 2, represents one such case. To tackle these difficulties, recent research has proposed the synthetic empirical likelihood (SEL), a likelihood-free approach based on empirical likelihood. In this chapter, we introduce a novel method for defining the SEL estimator by approximating the likelihood using the reparametrization trick, which enhances optimization efficiency. We establish the asymptotic properties of the SEL estimator in the just-identified case and develop several variance estimators for uncertainty quantification. The framework is then extended by incorporating exponential tilting to handle over-identified scenarios, resulting in the synthetic exponentially tilted empirical likelihood (SETEL), which empirically outperforms SEL in over-identified cases. The effectiveness of these proposed methods are demonstrated through comprehensive simulation studies across a range of scenarios.

References

- Ahsan, M. N. (2021), Statistical inference for stochastic volatility models, PhD thesis, McGill University.
- An, Z., Nott, D. J. and Drovandi, C. (2020), ‘Robust bayesian synthetic likelihood via a semi-parametric approach’, *Statistics and Computing* **30**(3), 543–557.
- Andrieu, C. and Roberts, G. O. (2009), ‘The pseudo-marginal approach for efficient monte carlo computations’.
- ASTM Standard D4761 (2005), Standard test methods for mechanical properties of lumber and wood-base structural material, Technical Report DOI: 10.1520/D4761-05, ASTM International, West Conshohocken, PA.
- Barbu, C. M., Sethuraman, K., Billig, E. M. and Levy, M. Z. (2018), ‘Two-scale dispersal estimation for biological invasions via synthetic likelihood’, *Ecography* **41**(4), 661–672.
- Bartlett, F., Hong, H. and Zhou, W. (2003), ‘Load factor calibration for the proposed 2005 edition of the national building code of canada: Statistics of loads and load effects’, *Canadian Journal of Civil Engineering* **30**(2), 429–439.
- Bernton, E., Jacob, P. E., Gerber, M. and Robert, C. P. (2019), ‘Approximate bayesian computation with the wasserstein distance’, *Journal of the Royal Statistical Society Series B: Statistical Methodology* **81**(2), 235–269.
- Blum, M. G. and François, O. (2010), ‘Non-linear regression models for approximate bayesian computation’, *Statistics and computing* **20**, 63–73.
- Boyd, S. and Mutapcic, A. (2008), ‘Stochastic subgradient methods’, *Lecture Notes for EE364b, Stanford University* **97**.

- Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S. and Zhang, Q. (2018), ‘JAX: composable transformations of Python+NumPy programs’.
URL: <http://github.com/jax-ml/jax>
- Brodtkorb, P. A. and D’Errico, J. (2015), ‘numdifftools 0.9.11’, <https://github.com/pbrod/numdifftools>.
- Cai, Y. (2015), Statistical methods for relating strength properties of dimensional lumber, PhD thesis, University of British Columbia.
- Cai, Y., Joe, H. and Pan, S. (2021), ‘Estimating dependence among lumber strength properties with copula models’, *Frontiers in Applied Mathematics and Statistics* **6**, 578614.
- Chan, Y. L., Schanzenbach, D. and Hickerson, M. J. (2014), ‘Detecting concerted demographic response across community assemblages using hierarchical approximate bayesian computation’, *Molecular Biology and Evolution* **31**(9), 2501–2515.
- Chaudhuri, S. and Ghosh, M. (2011), ‘Empirical likelihood for small area estimation’, *Biometrika* **98**(2), 473–480.
- Chaudhuri, S., Ghosh, S. and Pham, K. (2024), ‘On an empirical likelihood based solution to the approximate bayesian computation problem’, *Statistical Analysis and Data Mining: The ASA Data Science Journal* **17**.
- Chaudhuri, S., Mondal, D. and Yin, T. (2017), ‘Hamiltonian monte carlo sampling in bayesian empirical likelihood computation’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **1**(79), 293–320.
- Chen, J., Variyath, A. M. and Abraham, B. (2008), ‘Adjusted empirical likelihood and its properties’, *Journal of Computational and Graphical Statistics* **17**(2), 426–443.
- Chib, S., Shin, M. and Simoni, A. (2018), ‘Bayesian estimation and comparison of moment condition models’, *Journal of the American Statistical Association* **113**(524), 1656–1668.
- De Amorim, S. and Johnson, R. A. (1986), ‘Experimental designs for estimating the correlation between two destructively tested variables’, *Journal of the American Statistical Association* **81**(395), 807–812.
- Del Moral, P., Doucet, A. and Jasra, A. (2012), ‘An adaptive sequential monte carlo method for approximate bayesian computation’, *Statistics and computing* **22**, 1009–1020.

- Drovandi, C. C. and Pettitt, A. N. (2011), ‘Estimation of parameters for macroparasite population evolution using approximate bayesian computation’, *Biometrics* **67**(1), 225–233.
- Drovandi, C. and Frazier, D. T. (2022), ‘A comparison of likelihood-free methods with and without summary statistics’, *Statistics and Computing* **32**(3), 42.
- Duchi, J., Hazan, E. and Singer, Y. (2011), ‘Adaptive subgradient methods for online learning and stochastic optimization.’, *Journal of machine learning research* **12**(7).
- Ellingwood, B. and Rosowsky, D. (1991), ‘Duration of load effects in lrfd for wood construction’, *Journal of Structural Engineering* **117**(2), 584–599.
- Emerson, S. C. and Owen, A. B. (2009), ‘Calibration of the empirical likelihood method for a vector mean’.
- Evans, J. W., Johnson, R. A. and Green, D. W. (1984), ‘Estimating the correlation between variables under destructive testing, or how to break the same board twice’, *Technometrics* **26**(3), 285–290.
- Fan, J. and Li, R. (2002), ‘Variable selection for cox’s proportional hazards model and frailty model’, *The Annals of Statistics* **30**(1), 74–99.
- Fearnhead, P. and Prangle, D. (2012), ‘Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation’, *Journal of the Royal Statistical Society Series B: Statistical Methodology* **74**(3), 419–474.
- Foschi, R., Folz, B. and Yao, F. (1989), *Reliability-based design of wood structures*, University of British Columbia.
- Foschi, R. O. and Barrett, J. D. (1982), ‘Load-duration effects in western hemlock lumber’, *Journal of the Structural Division* **108**(7), 1494–1510.
- Foschi, R. and Yao, Z. (1986), Another look at three duration of load models, in ‘Proceedings of IUFRO wood engineering group meeting, Florence, Italy, paper’, pp. 19–9.
- Frazier, D. T., Nott, D. J., Drovandi, C. and Kohn, R. (2023), ‘Bayesian inference using synthetic likelihood: asymptotics and adjustments’, *Journal of the American Statistical Association* **118**(544), 2821–2832.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013), *Bayesian data analysis*, CRC press.

- Gerhards, C. (1979), ‘Time-related effects of loading on wood strength: a linear cumulative damage theory.’, *Wood science* **11**(3), 139–144.
- Gerhards, C. and Link, C. (2007), ‘A cumulative damage model to predict load duration characteristics of lumber’, *Wood and Fiber Science* **19**(2), 147–164.
- Gilbert, B. P., Zhang, H. and Bailleres, H. (2019), ‘Reliability of laminated veneer lumber (lvl) beams manufactured from early to mid-rotation subtropical hardwood plantation logs’, *Structural Safety* **78**, 88–99.
- Green, D. W., Evans, J. W. and Johnson, R. A. (1984), ‘Investigation of the procedure for estimating concomitance of lumber strength properties’, *Wood and Fiber Science* **16**(3), 427–440.
- Green, D. W., Winandy, J. E. and Kretschmann, D. E. (1999), ‘Mechanical properties of wood’, *Wood handbook: wood as an engineering material. Madison, WI: USDA Forest Service, Forest Products Laboratory, 1999. General technical report FPL; GTR-113: Pages 4.1-4.45* **113**.
- Hall, A. R. and Inoue, A. (2003), ‘The large sample behaviour of the generalized method of moments estimator in misspecified models’, *Journal of Econometrics* **114**(2), 361–394.
- Harvey, A. C. and Shephard, N. (1996), ‘Estimation of an asymmetric stochastic volatility model for asset returns’, *Journal of Business & Economic Statistics* **14**(4), 429–434.
- Haynes, M. A., MacGillivray, H. and Mengersen, K. (1997), ‘Robustness of ranking and selection rules using generalised g-and-k distributions’, *Journal of Statistical Planning and Inference* **65**(1), 45–66.
URL: <https://www.sciencedirect.com/science/article/pii/S0378375897000505>
- Hinton, G., Srivastava, N. and Swersky, K. (2012), ‘Neural networks for machine learning lecture 6a overview of mini-batch gradient descent’, *Cited on* **14**(8), 2.
- Hoffmeyer, P. and Sørensen, J. D. (2007), ‘Duration of load revisited’, *Wood Science and Technology* **41**(8), 687–711.
- Huber, P. J. (1992), Robust estimation of a location parameter, *in* ‘Breakthroughs in statistics: Methodology and distribution’, Springer, pp. 492–518.
- Imbens, G. W. (1997), ‘One-step estimators for over-identified generalized method of moments models’, *The Review of Economic Studies* **64**(3), 359–383.

- Jiang, B. (2018), Approximate bayesian computation with kullback-leibler divergence as data discrepancy, *in* ‘International conference on artificial intelligence and statistics’, PMLR, pp. 1711–1721.
- Johnson, R. A. and Galligan, W. L. (1983), ‘Estimating the concomitance of lumber strength properties’, *Wood and Fiber Science* **15**(3), 235–244.
- Johnson, R. A. and Lu, W. (2007), ‘Proof load designs for estimation of dependence in a bivariate weibull model’, *Statistics & probability letters* **77**(11), 1061–1069.
- Karacabeyli, E. and Lum, C. (1994), ‘Failure rate and strength degradation in lumber under constant load’, *IUFRO TIMBER ENGINEERING MEETING*.
- Kass, R. E. and Raftery, A. E. (1995), ‘Bayes factors’, *Journal of the american statistical association* **90**(430), 773–795.
- Katzengruber, R., Jeitler, G., Brandner, R. and Schickhofer, G. (2006), Tensile proof loading to assure quality of finger-jointed structural timber, *in* ‘Proceedings of WCTE (CD-ROM), 9th World Conference on Timber Engineering’, pp. p6–10.
- Kingma, D. P. and Ba, J. (2014), ‘Adam: A method for stochastic optimization’, *arXiv preprint arXiv:1412.6980*.
- Kingma, D. P. and Welling, M. (2013), ‘Auto-encoding variational bayes’, *arXiv preprint arXiv:1312.6114*.
- Köhler, J. and Svensson, S. (2011), ‘Probabilistic representation of duration of load effects in timber structures’, *Engineering structures* **33**(2), 462–467.
- Lazar, N. A. (2003), ‘Bayesian empirical likelihood’, *Biometrika* **90**(2), 319–326.
- Lee, S. (2016), ‘Asymptotic refinements of a misspecification-robust bootstrap for gel estimators’, *Journal of econometrics* **192**(1), 86–104.
- Li, Y. and Lam, F. (2016), ‘Reliability analysis and duration-of-load strength adjustment factor of the rolling shear strength of cross laminated timber’, *Journal of wood science* **62**(6), 492–502.
- Liang, H. and Chen, J. (2023), ‘Global consistency of empirical likelihood’, *arXiv preprint arXiv:2303.16410*.

- Liu, C.-S. and Liang, H.-Y. (2022), ‘Bayesian empirical likelihood of quantile regression with missing observations’, *Metrika* pp. 1–29.
- Maasoumi, E. and Phillips, P. C. (1982), ‘On the behavior of inconsistent instrumental variable estimators’, *Journal of Econometrics* **19**(2-3), 183–201.
- Madsen, B. (1976), ‘In-grade testing: degree of damage due to proof loading of lumber in bending.’, *Structural Research Series Report. Department of Civil Engineering University of British Columbia (Canada). no. 17.* .
- Madsen, H. O., Krenk, S. and Lind, N. C. (2006), *Methods of structural safety*, Courier Corporation.
- Marcondes, D. (2020), ‘Parameter estimation in dynamical systems via statistical learning: a reinterpretation of approximate bayesian computation applied to covid-19 spread’, *arXiv preprint arXiv:2007.14229* .
- Margossian, C. C. (2019), ‘A review of automatic differentiation and its efficient implementation’, *Wiley interdisciplinary reviews: data mining and knowledge discovery* **9**(4), e1305.
- Marin, L. A. and Woeste, F. E. (1981), ‘Reverse proof loading as a means of quality control in lumber manufacturing’, *Transactions of the ASAE* **24**(5), 1273–1277.
- Marjoram, P., Molitor, J., Plagnol, V. and Tavaré, S. (2003), ‘Markov chain monte carlo without likelihoods’, *Proceedings of the National Academy of Sciences* **100**(26), 15324–15328.
- Martin, G. M., McCabe, B. P., Frazier, D. T., Maneesoonthorn, W. and Robert, C. P. (2019), ‘Auxiliary likelihood-based approximate bayesian computation in state space models’, *Journal of Computational and Graphical Statistics* **28**(3), 508–522.
- Mengersen, K. L., Pudlo, P. and Robert, C. P. (2013), ‘Bayesian computation via empirical likelihood’, *Proceedings of the National Academy of Sciences* **110**(4), 1321–1326.
- Murphy, J., Ellingwood, B. and Hendrickson, E. (1987), ‘Damage accumulation in wood structural members under stochastic live loads’, *Wood and fiber science* **19**(4), 453–463.
- Newey, W. K. and Smith, R. J. (2004), ‘Higher order properties of gmm and generalized empirical likelihood estimators’, *Econometrica* **72**(1), 219–255.

- Owen, A. (1988), ‘Empirical likelihood ratio confidence intervals for a single functional’, *Biometrika* **75**(2), 237–249.
- Owen, A. B. (2001), *Empirical likelihood*, Chapman and Hall/CRC.
- Price, L. F., Drovandi, C. C., Lee, A. and Nott, D. J. (2018), ‘Bayesian synthetic likelihood’, *Journal of Computational and Graphical Statistics* **27**(1), 1–11.
- Priddle, J. W., Sisson, S. A., Frazier, D. T., Turner, I. and Drovandi, C. (2022), ‘Efficient bayesian synthetic likelihood with whitening transformations’, *Journal of Computational and Graphical Statistics* **31**(1), 50–63.
- Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. and Feldman, M. W. (1999), ‘Population growth of human y chromosomes: a study of y chromosome microsatellites.’, *Molecular biology and evolution* **16**(12), 1791–1798.
- Qin, J. and Lawless, J. (1994), ‘Empirical likelihood and general estimating equations’, *The Annals of Statistics* pp. 300–325.
- Quach, H., Barreiro, L. B., Laval, G., Zidane, N., Patin, E., Kidd, K. K., Kidd, J. R., Bouchier, C., Veuille, M., Antoniewski, C. et al. (2009), ‘Signatures of purifying and local positive selection in human mirnas’, *The American Journal of Human Genetics* **84**(3), 316–327.
- Raftery, A. E., Madigan, D. and Hoeting, J. A. (1997), ‘Bayesian model averaging for linear regression models’, *Journal of the American Statistical Association* **92**(437), 179–191.
- Rao, J. and Wu, C. (2010), ‘Bayesian pseudo-empirical-likelihood intervals for complex surveys’, *Journal of the Royal Statistical Society Series B: Statistical Methodology* **72**(4), 533–544.
- Richard, L. and Burden, J. (2001), ‘Douglas faires. numerical analysis . thomson learning’, *Inc. Aug* .
- Schennach, S. M. (2005), ‘Bayesian exponentially tilted empirical likelihood’, *Biometrika* **92**(1), 31–46.
- Schennach, S. M. (2007), ‘Point estimation with exponentially tilted empirical likelihood’, *The Annals of Statistics* pp. 634–672.
- Shephard, N. and Andersen, T. G. (2009), Stochastic volatility: origins and overview, *in* ‘Handbook of financial time series’, Springer, pp. 233–254.

- Sisson, S. A., Fan, Y. and Beaumont, M. (2018), *Handbook of approximate Bayesian computation*, CRC Press.
- Smith, I., Landis, E. and Gong, M. (2003), *Fracture and fatigue in wood*, John Wiley & Sons.
- Steiner, S. and Wesolowsky, G. (1995), ‘Estimating the correlation between destructively measured variables using proof-loading’, *Technometrics* **37**(1), 94–101.
- Strickler, M., Pellerin, R., Talbott, J. et al. (1970), ‘Experiments in proof loading structural end-jointed lumber.’, *Forest Products Journal* **20**(2), 29–35.
- Suddarth, S. K., Woeste, F. E. and Galligan, W. L. (1978), *Differential reliability: Probabilistic engineering applied to wood members in bending/tension*, Vol. 302, Department of Agriculture, Forest Service, Forest Products Laboratory.
- Sueishi, N. (2022), ‘Large sample justifications for the bayesian empirical likelihood’, *Econometric Theory* pp. 1–31.
- Taylor, S. J. (1994), ‘Modeling stochastic volatility: A review and comparative study’, *Mathematical finance* **4**(2), 183–204.
- Tseng, S.-T. and Yu, H.-F. (1997), ‘A termination rule for degradation experiments’, *IEEE Transactions on Reliability* **46**(1), 130–133.
- van der Vaart, E., Beaumont, M. A., Johnston, A. S. and Sibly, R. M. (2015), ‘Calibration and evaluation of individual-based models using approximate bayesian computation’, *Ecological Modelling* **312**, 182–190.
- Vankov, E. and Ensor, K. B. (2014), ‘Stochastic volatility filtering with intractable likelihoods’, *arXiv preprint arXiv:1405.4323* .
- Woeste, F., Green, D., Tarbell, K. and Marin, L. (1986), ‘Proof loading to assure lumber strength’, *Wood and fiber science* **19**(3), 283–297.
- Wong, S. W. (2020), ‘Calibrating wood products for load duration and rate: A statistical look at three damage models’, *Wood Science and Technology* **54**(6), 1511–1528.
- Wong, S. W. and Zidek, J. V. (2019), ‘The duration of load effect in lumber as stochastic degradation’, *IEEE Transactions on Reliability* **68**(2), 410–419.

- Wood, L. W. (1951), *Relation of strength of wood to duration of load*, number 1916, US Department of Agriculture, Forest Service, Forest Products Laboratory.
- Wood, L. W. et al. (1960), 'Relation of strength of wood to duration of load'.
- Wood, S. N. (2010), 'Statistical inference for noisy nonlinear ecological dynamic systems', *Nature* **466**(7310), 1102–1104.
- Yang, C.-H., Zidek, J. V. and Wong, S. W. (2019), 'Bayesian analysis of accumulated damage models in lumber reliability', *Technometrics* **61**(2), 233–245.
- Yang, Y. and He, X. (2012), 'Bayesian empirical likelihood for quantile regression', *The Annals of Statistics* **40**(2), 1102–1131.

APPENDICES

Appendix A

Derivation of formulas in Chapter 2

A.1 Derivation of equation (2.9)

Based on the property that the distribution for the maximum of the N_s segment ground snow load G_s equals that for the annual ground snow load G , we have

$$F_G(g) = [F_{G_s}(g)]^{N_s}, \quad (\text{A.1})$$

where $F_G(g), F_{G_s}(g)$ are the cumulative distribution function of G and G_s , respectively.

Using the law of total probability, the cumulative distribution function of G_s equals to

$$\begin{aligned} F_{G_s}(g) &= P(G_S < g|\text{no snow})P(\text{no snow}) + P(G_S < g|\text{snow})P(\text{snow}) \\ &= P(0 < g|\text{no snow})P(\text{no snow}) + P(G_S < g|\text{snow})P(\text{snow}) \\ &= 1 \cdot (1 - p_e) + P(G_S < g|\text{snow}) \cdot p_e. \end{aligned} \quad (\text{A.2})$$

Combine (A.1) and (A.2) yields

$$F_G(g) = [(1 - p_e) + P(G_S < g|\text{snow}) \cdot p_e]^{N_s}, \quad (\text{A.3})$$

Plug (A.3) into (2.5), we have (2.9).

Appendix B

Derivation of formulas in Chapter 3

B.1 Histograms of the Hemlock dataset

B.2 Derivation for the detailed expression of the likelihood (3.10) in the bivariate scenario

Given the joint PDF $f_{X,Y}(x, y)$ and the damage function,

$$Y^* = \left[Y - \alpha \cdot \frac{\mu_Y}{\mu_X} \left(\frac{l}{\eta} - X \right) \right] \cdot \mathbf{I}\{l > \eta \cdot X\} + Y \cdot \mathbf{I}\{l \leq \eta \cdot X\},$$

the goal is to calculate $\int_l^\infty f_{X,Y^*}(x, y^*; \boldsymbol{\beta}, \alpha, \eta) dx$.

Note that $\int_l^\infty f_{X,Y^*}(x, y^*; \boldsymbol{\beta}, \alpha, \eta) dx = \frac{d}{dy^*} P(X > l, Y^* \leq y^*)$. We first consider $P(X > l, Y^* \leq y^*)$, which follows

$$\begin{aligned} & P(X > l, Y^* \leq y^*) \\ &= P(X > l, h(Y, \alpha) < y^* \mid X < \frac{l}{\eta}) P(X < \frac{l}{\eta}) + P(X > l, Y < y^* \mid X \geq \frac{l}{\eta}) P(X \geq \frac{l}{\eta}) \\ & \text{(using law of total probability)} \\ &= P(l < X < \frac{l}{\eta}, Y < h^{-1}(y^*, \alpha)) + P(X \geq \frac{l}{\eta}, Y < y^*) \end{aligned} \tag{B.1}$$

$$= \int_l^{\frac{l}{\eta}} \int_{-\infty}^{h^{-1}(y^*, \alpha)} f_{X,Y}(x, y; \boldsymbol{\beta}) dx dy + \int_{\frac{l}{\eta}}^\infty \int_{-\infty}^{y^*} f_{X,Y}(x, y; \boldsymbol{\beta}) dx dy, \tag{B.2}$$

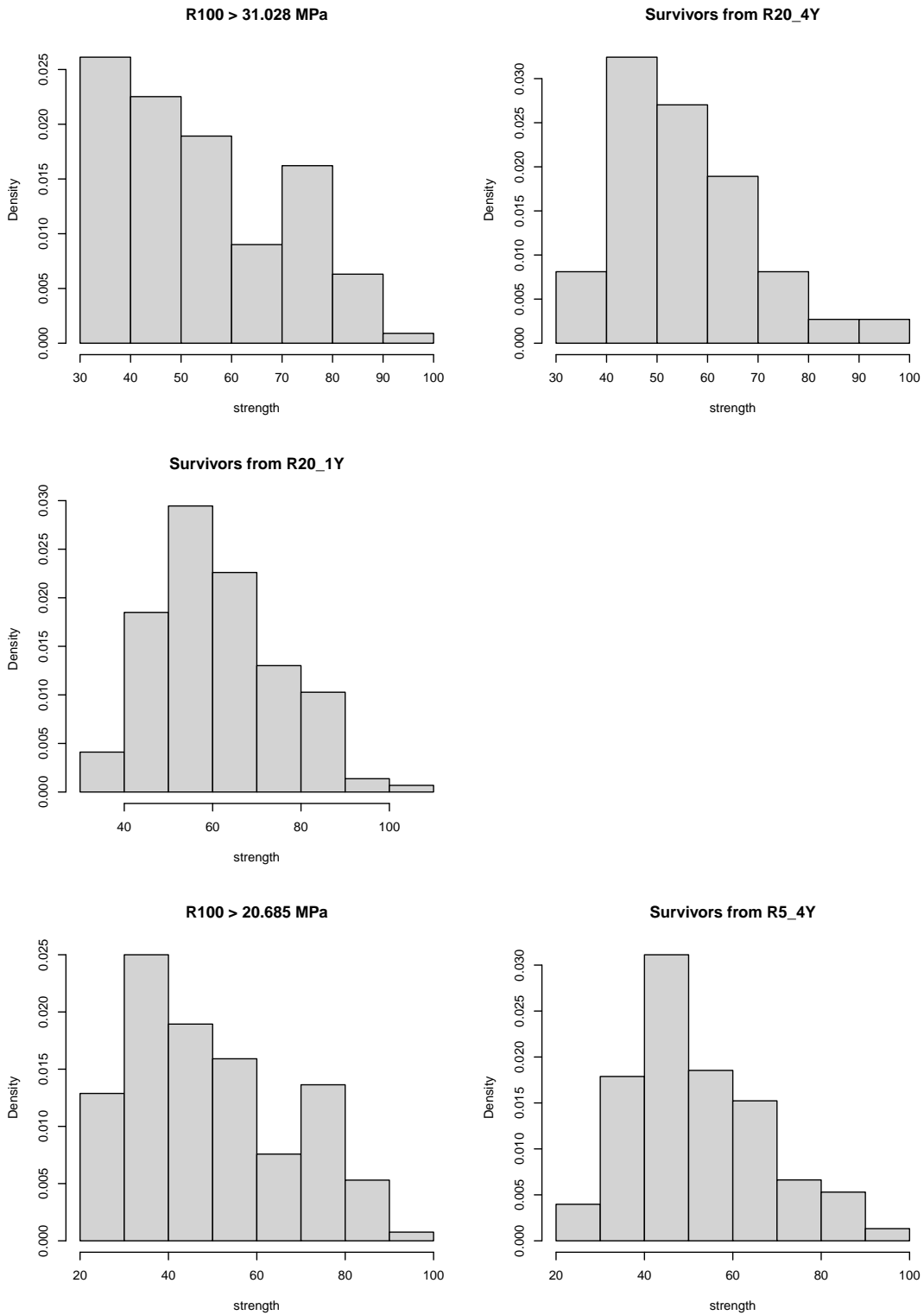


Figure B.1: Histograms of strength distributions for truncated subsets of R100 (pieces with strength greater than 31.0275 MPa or 20.68 MPa) and for survivors from the proof-loading groups R20_1Y, R20_4Y, and R5_4Y from the Hemlock dataset.

where $h^{-1}(y^*, \alpha) = y^* + \alpha \cdot \frac{\mu_Y}{\mu_X} \left(\frac{l}{\eta} - x \right)$.

Then taking the derivative, we have

$$\begin{aligned}
& \int_l^\infty f_{X,Y^*}(x, y^*; \boldsymbol{\beta}, \alpha, \eta) dx \\
&= \frac{d}{dy^*} P(X > l, Y^* \leq y^*) \\
&= \frac{d}{dy^*} h^{-1}(y^*, \alpha) \int_l^{\frac{l}{\eta}} f_{X,Y}(x, h^{-1}(y^*, \alpha); \boldsymbol{\beta}) dx + \int_{\frac{l}{\eta}}^\infty f_{X,Y}(x, y^*; \boldsymbol{\beta}) dx \\
&= \int_l^{\frac{l}{\eta}} f_{X,Y}(x, h^{-1}(y^*, \alpha); \boldsymbol{\beta}) dx + \int_{\frac{l}{\eta}}^\infty f_{X,Y}(x, y^*; \boldsymbol{\beta}) dx, \tag{B.3}
\end{aligned}$$

where $h^{-1}(y^*, \alpha) = y^* + \alpha \cdot \frac{\mu_Y}{\mu_X} \left(\frac{l}{\eta} - x \right)$.

B.3 Derivation for the detailed expression of the likelihood (3.13) in the univariate scenario

Given the $Y \sim N(\mu, \sigma^2)$ and the damage model

$$Y^* = \left[Y - \alpha \cdot \left(\frac{l}{\eta} - Y \right) \right] \cdot \mathbf{I}\{l > \eta \cdot Y\} + Y \cdot \mathbf{I}\{l \leq \eta \cdot Y\}, \tag{B.4}$$

the goal here is to find the detailed expression the likelihood $L_p(\boldsymbol{\beta}, \alpha, \eta; y^*)$, where $\boldsymbol{\beta} = (\mu, \sigma)$.

Note that the likelihood is equivalent to the PDF of Y^* , which is

$$L_p(\boldsymbol{\beta}, \alpha, \eta; y^*) = f_{Y^*}(y^*; \boldsymbol{\beta}, \alpha, \eta), \tag{B.5}$$

where $f_{Y^*}(y^*; \boldsymbol{\beta}, \alpha, \eta)$ is the PDF of Y^* .

Before investigating the PDF of Y^* , we introduce the truncated normal distribution. Let $\text{TN}(\mu, \sigma^2, a, b)$ represent the truncated normal distribution with the original normal $N(\mu, \sigma^2)$ truncated in the interval (a, b) . The PDF of a truncated normal distribution $\text{TN}(\mu, \sigma^2, a, b)$ is given by:

$$f_{\text{TN}}(\tilde{y}; \mu, \sigma, a, b) = \begin{cases} \frac{\frac{1}{\sigma}\phi\left(\frac{\tilde{y}-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)} & \text{for } a \leq \tilde{y} \leq b, \\ 0 & \text{otherwise,} \end{cases} \quad (\text{B.6})$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and CDF of standard normal distribution respectively, and \tilde{y} is the value at which the PDF of the truncated normal distribution $\text{TN}(\mu, \sigma^2, a, b)$ is evaluated. The corresponding CDF is $F_{\text{TN}}(\tilde{y}; \mu, \sigma, a, b)$.

We next discuss the distribution of Y and CDF of Y^* . The proof-loading background indicates that the strengths of all survivors satisfy $Y^* > l$. Then the corresponding Y in the survivors satisfies $Y > h^{-1}(l, \alpha)$. This indicates that the original strength Y in the survivors follows the truncated normal distribution, which is

$$Y \sim \text{TN}(\mu, \sigma^2, h^{-1}(l, \alpha), \infty). \quad (\text{B.7})$$

It yields the PDF and CDF of Y , which are $f_Y(y; \boldsymbol{\beta}, \alpha, \eta) = f_{\text{TN}}(y; \mu, \sigma, h^{-1}(l, \alpha), \infty)$ and $F_Y(y; \boldsymbol{\beta}, \alpha, \eta) = F_{\text{TN}}(y; \mu, \sigma, h^{-1}(l, \alpha), \infty)$.

Given the distribution of Y , the CDF of Y^* is

$$\begin{aligned} F_{Y^*}(y^*; \boldsymbol{\beta}, \alpha, \eta) &= P(Y^* < y^* \mid Y < l/\eta)P(Y < l/\eta) + P(Y^* < y^* \mid Y \geq l/\eta)P(Y \geq l/\eta) \\ &= P((h(Y, \alpha) < y^*, Y < l/\eta) + P(Y < y^*, Y \geq l/\eta) \\ &= P(Y < h^{-1}(y^*, \alpha), Y < l/\eta) + P(l/\eta \leq Y < y^*) \\ &= P(Y < \min(h^{-1}(y^*, \alpha), l/\eta)) + P(l/\eta \leq Y < y^*), \end{aligned} \quad (\text{B.8})$$

where $h^{-1}(y^*, \alpha) = (\alpha + 1)^{-1}(y^* + \alpha \cdot l/\eta)$. Organize it and we have

$$F_{Y^*}(y^*; \boldsymbol{\beta}, \alpha, \eta) = \begin{cases} F_Y(h^{-1}(y^*, \alpha); \boldsymbol{\beta}, \alpha, \eta) & \text{for } l < y^* < l/\eta, \\ F_Y(y^*; \boldsymbol{\beta}, \alpha, \eta) & \text{for } y^* \geq l/\eta. \end{cases} \quad (\text{B.9})$$

Taking the derivative of CDF of Y^* , $F_{Y^*}(y^*; \boldsymbol{\beta}, \alpha, \eta)$, we have the PDF of Y^* , $f_{Y^*}(y^*; \boldsymbol{\beta}, \alpha, \eta)$, which is

$$f_{Y^*}(y^*; \boldsymbol{\beta}, \alpha, \eta) = \begin{cases} (\alpha + 1)^{-1}f_Y(h^{-1}(y^*, \alpha); \boldsymbol{\beta}, \alpha, \eta) & \text{for } l < y^* < l/\eta, \\ f_Y(y^*; \boldsymbol{\beta}, \alpha, \eta) & \text{for } y^* \geq l/\eta. \end{cases} \quad (\text{B.10})$$

Plugging in $f_Y(y; \boldsymbol{\beta}, \alpha, \eta) = f_{\text{TN}}(y; \mu, \sigma, h^{-1}(l, \alpha), \infty)$ into Equation (B.10) gives Equation (3.13).