Randomized Survival Probability Residual for Assessing Parametric Survival Models

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the School of Public Health
University of Saskatchewan
Saskatoon

By
Tingxuan Wu

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Director of School of Public Health
Health Sciences Building E-Wing, 104 Clinic Place
University of Saskatchewan
Saskatoon, Saskatchewan S7N 2Z4
Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

Abstract

Traditional residuals for diagnosing accelerated failure time models in survival analysis, such as Cox-Snell, martingale and deviance residuals, have been widely used. However, examining those residuals are often only made visually, which can be subjective. Therefore, lack of objective measure of examining model adequacy has been a long-standing issue that needs to be addressed for survival analysis. In this thesis, a new type of residual is proposed called Normal-transformed Randomized Survival Probability (NRSP) residual. A comprehensive review of the traditional residuals including Cox Snell and deviance residuals is firstly presented highlighting their disadvantages for examining model adequacy. We then introduce NRSP residual. Simulation studies were conducted to compare the performance of NRSP residuals with the traditional residuals. Our simulation studies demonstrated that NRSP residuals are approximately normally distributed when the fitted model is correctly specified, and has great statistical power in detecting model inadequacies. We also apply NRSP residuals to a real dataset to check the goodness-of-fit of three plausible models.

Acknowledgements

I would like to first and foremost express my sincere gratitude to my supervisors, Dr. Longhai Li and Dr. Cindy Feng, for their academic and financial support as well as their encouragements and patience throughout the period of my study. Without their enlightening instruction, impressive kindness and patience, I could not have completed my thesis. Their guidance and encouragement enlighten me not only in this thesis but also in my future study. It is a great honour to work under their supervision.

I would like to thank the School of Public Health and the Department of Mathematics and Statistics at the University of Saskatchewan for academic and financial support to my MSc study. I am grateful to all of the professors, graduate students, and staff in the School of Public Health and the Department of Mathematics and Statistics.

I would especially like to express very profound gratitude to my parents for their deep love and strong support during my whole life. I would also like to thank my sister, Jiangqing Ni, for taking care of my my daily life during my MSc study. I love them forever.

Thanks for everything.

Contents

Permission to Use i

Abstract ii

Acknowledgements iii

Contents iv

List of Tables vi

List of Figures vii

List of Abbreviations xi

1 Introduction 1

2 Methodology 4
- 2.1 A brief introduction to survival analysis . . . 4
- 2.2 Accelerated failure time regression model . . . 6
  - 2.2.1 Weibull AFT regression model . . . 7
  - 2.2.2 Log-normal AFT regression model . . . 8
- 2.3 Traditional residuals for checking models . . . 9
  - 2.3.1 Pearson residuals . . . 9
  - 2.3.2 Cox-Snell residuals . . . 9
  - 2.3.3 Martingale residuals . . . 11
  - 2.3.4 Deviance residuals . . . 11
- 2.4 Problems with traditional residuals . . . 11
  - 2.4.1 Illustrative examples for Cox-Snell residuals and deviance residuals . 12
  - 2.4.2 Example 1: Assessing distributional assumption for survival time . . 12
  - 2.4.3 Example 2: Assessing functional form of covariate effect for survival time 13

3 Normal-transformed Randomized Survival Probability residual 17
- 3.1 Definition of Normal-transformed Randomized Survival Probability (NRSP) residual . . . 17
- 3.2 Illustrative example . . . 18

4 Simulation studies 22
- 4.1 Assessing distributional assumption for a survival model . . . 23
  - 4.1.1 Results of a single simulation scenario . . . 23
  - 4.1.2 Power analysis . . . 29
  - 4.1.3 Model comparisons . . . 32
- 4.2 Assessing functional form of the covariate effect . . . 32

4.2.1 Results of a single simulation scenario 33
4.2.2 Power analysis 38
4.2.3 Model comparisons 41

5 Real data analysis **42**

6 Conclusion and future work **50**

Bibliography **51**

List of Tables

2.1 Distributions for commonly used parametric survival time and the associated error term for AFT models. 7

5.1 Variable definitions in the breast cancer study. 43
5.2 Parameter estimates of the Weibull, log-normal and log-logistic models in Breast Cancer Study. 46
5.3 Percentages of P-values smaller than 0.05 for the SW test of the NRSP residuals and AIC comparisons for Weibull, Log-normal and Log-logistic models in the breast cancer data analysis. 49

List of Figures

2.1 Unmodified Cox-Snell (UCS) residuals and deviance residuals for the true model (left panel) and the wrong model (right panel) for the first example in section 2.4.2. The green triangles correspond to the event times and the red circles correspond to the censored times. 15

2.2 Unmodified Cox-Snell (UCS) residuals and deviance residuals for the true model (left panel) and the wrong model (right panel) for the second example in section 2.4.3. The green triangles correspond to the event times and the red circles correspond to the censored times. 16

3.1 RSP for the true model (first row) and the wrong model (second row). The left panels are randomized survival functions and right panels are the randomized histogram of RSPs 19

3.2 Unmodified survival probability (USP) for the true model (first row) and the wrong model (second row). The left panels are survival functions and right panels are the histogram of USPs. 20

3.3 Modified survival probability (MSP) for the true model (first row) and the wrong model (second row). The left panels are survival functions and right panels are the histogram of MSPs. 21

4.1 Performance of the NRSP residuals in detecting distributional assumption of a sample dataset of size $n = 800$ and a percentage of censorship $c = 50\%$. The panels in the first row present the NRSP residuals for the true model: Weibull distribution. The panels in the second row present the NRSP residuals for the wrong model: Log-normal distribution. The first two columns display the scatter plots and QQ plots of the NRSP residuals, respectively. The third and fourth columns present the histograms of the SW and KS p-values for the NRSP residuals over 1000 randomly generated datasets from the true model. The green triangles correspond to the event times and the red circles correspond to the censored times. 26

4.2 Performance of the NMSP residuals in detecting distributional assumption of a sample dataset of size $n = 800$ and a percentage of censorship $c = 50\%$. The panels in the first row present the NMSP residuals for the true model: Weibull distribution. The panels in the second row present the NMSP residuals for the wrong model: Log-normal distribution. The first two columns display the scatter plots and QQ plots of the NMSP residuals, respectively. The third and fourth columns present the histograms of the SW and KS p-values for the NMSP residuals over 1000 randomly generated datasets from the true model. The green triangles correspond to the event times and the red circles correspond to the censored times. 27

4.3 Performance of the deviance residuals in detecting distributional assumption of a sample dataset of size $n = 800$ and a percentage of censorship $c = 50\%$. The panels in the first row present the deviance residuals for the true model: Weibull distribution. The panels in the second row present the deviance residuals for the wrong model: Log-normal distribution. The first two columns display the scatter plots and QQ plots of the deviance residuals, respectively. The third and fourth columns present the histograms of the SW and KS p-values for the deviance residuals over 1000 randomly generated datasets from the true model. The green triangles correspond to the event times and the red circles correspond to the censored times. 28

4.4 Comparison of the type I errors and powers of the SW tests for the NRSP, NMSP, and deviance residuals. Response variable is simulated from the true model at varying sample sizes of $n = 100$, 200, 400, 600, 800 and 1000, and the percentage of censorship $c = 20\%$ (black circles), 50% (red triangles) and 80% (green crosses). True model: Weibull model. Wrong model: Log-normal model. 30

4.5 Comparison of the type I errors and powers of the KS tests for the NRSP, NMSP, and deviance residuals. Response variable is simulated from the true model at varying sample sizes of $n = 100$, 200, 400, 600, 800 and 1000, and the percentage of censorship $c = 20\%$ (black circles), 50% (red triangles) and 80% (green crosses). True model: Weibull model. Wrong model: Log-normal model. 31

4.6 AIC for true model (Weibull regression) and wrong model (Log-normal regression) at varying sample sizes of $n = 100$, 200, 400, 600, 800 and 1000, and the percentage of censorship $c = 20\%$ (black circles), 50% (red triangles) and 80% (green crosses). 32

4.7 Performance of the NRSP residuals in detecting distributional assumption of a sample dataset of size $n = 800$ and a percentage of censorship $c = 50\%$. The panels in the first row present the NRSP residuals for the true model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 f(x) + \epsilon_i$. The panels in the second row present the NRSP residuals for the wrong model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$. The first two columns display the scatter plots and QQ plots of the NRSP residuals, respectively. The third and fourth columns present the histograms of the SW and KS p-values for the NRSP residuals over 1000 randomly generated datasets from the true model. The green triangles correspond to the event times and the red circles correspond to the censored times. 35

4.8 Performance of the NMSP residuals in detecting distributional assumption of a sample dataset of size $n = 800$ and a percentage of censorship $c = 50\%$. The panels in the first row present the NMSP residuals for the true model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 f(x) + \epsilon_i$. The panels in the second row present the NMSP residuals for the wrong model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$. The first two columns display the scatter plots and QQ plots of the NMSP residuals, respectively. The third and fourth columns present the histograms of the SW and KS p-values for the NMSP residuals over 1000 randomly generated datasets from the true model. The green triangles correspond to the event times and the red circles correspond to the censored times. 36

4.9 Performance of the deviance residuals in detecting distributional assumption of a sample dataset of size $n = 800$ and a percentage of censorship $c = 50\%$. The panels in the first row present the deviance residuals for the true model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 f(x) + \epsilon_i$. The panels in the second row present the deviance residuals for the wrong model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$. The first two columns display the scatter plots and QQ plots of the deviance residuals, respectively. The third and fourth columns present the histograms of the SW and KS p-values for the deviance residuals over 1000 randomly generated datasets from the true model. The green triangles correspond to the event times and the red circles correspond to the censored times. 37

4.10 Comparison of the type I errors and powers of the SW tests for the NRSP, NMSP, and deviance residuals. Response variable is simulated from the true model at varying sample sizes of $n = 100$, 200, 400, 600, 800 and 1000, and the percentage of censorship $c = 20\%$ (black circles), 50% (red triangles) and 80% (green crosses). True model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 f(x) + \epsilon_i$. Wrong model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$. 39

4.11 Comparison of the type I errors and powers of the KS tests for the NRSP, NMSP, and deviance residuals. Response variable is simulated from the true model at varying sample sizes of $n = 100$, 200, 400, 600, 800 and 1000, and the percentage of censorship $c = 20\%$ (black circles), 50% (red triangles) and 80% (green crosses). True model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 f(x) + \epsilon_i$. Wrong model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$. 40

4.12 AIC for true model (Weibull model $\log(T_i) = \beta_0 + \beta_1 f(x) + \epsilon_i$) and wrong model (Weibull model $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$) at varying sample sizes of $n = 100$, 200, 400, 600, 800 and 1000, and the percentage of censorship $c = 20\%$ (black circles), 50% (red triangles) and 80% (green crosses). 41

5.1 Cox-Snell residuals for Weibull, Log-logistic, and Lognormal models. The green triangles correspond to the event times and the red circles correspond to the censored times. 44

5.2 NRSP residuals for the Weibull, Log-logistic, and Lognormal AFT models fitted to the breast cancer patients dataset. The panels in the first two columns present the scatter plots and QQ plots of the NRSP residuals versus the fitted values, respectively. The green triangles correspond to the event times and the red circles correspond to the censored times. The third column presents the frequencies of the p-values of the SW normality test for 1000 replicated NRSP residuals. 45
5.3 NMSP residuals for the Weibull, Log-logistic, and Lognormal AFT models fitted to the breast cancer patients dataset. The panels in the first two columns present the scatter plots and QQ plots of the NMSP residuals versus the fitted values, respectively. The green triangles correspond to the event times and the red circles correspond to the censored times. 47
5.4 Deviance residuals for the Weibull, Log-logistic, and Lognormal AFT models fitted to the breast cancer patients dataset. The panels in the first two columns present the scatter plots and QQ plots of the deviance residuals versus the fitted values, respectively. The green triangles correspond to the event times and the red circles correspond to the censored times. 48

List of Abbreviations

AFT	Accelerated Failure Time
GOF	Goodness of Fit
CDF	Cumulative Distribution Function
PDF	Probability Density Function
AIC	Akaike's Information Criterion

1. Introduction

Examining model adequacy is a critical step in model building to ensure the validity of the statistical inference. Model checking and assessing the overall goodness of fit (GOF) are typically based on residuals. Residuals for survival models are different from the residuals for generalized linear models, due to censored observations, which makes model diagnostics very difficult. Therefore, visual inspection of residuals should be supplemented by a numerical goodness-of-fit test to detect inadequacies of a fitted model [1].

Cox-Snell[2], martingale [3], and deviance [4] residuals have often been used to diagnose survival models. Of those, Cox-Snell residuals are most widely used in the analysis of survival data. Cox-Snell residual is defined as the negative logarithms of the estimated survivor function for an individual with the observed survival time. It is proven that if a model fits observed data well without censoring, Cox-Snell residuals are distributed exponentially with unit mean. If the observed survival time for an individual is censored, the corresponding Cox-Snell residual is also censored. Cox-Snell residual is therefore quite dissimilar to those of residuals used in linear regression analysis in the sense that they cannot be negative and consequently are not symmetrically distributed around zero [1]. To assess model fit, Cox-Snell residuals are commonly plotted against values of negative logarithms of the estimated survivor function. A straight line with unit slope and zero intercept indicates that model fits a dataset well. Using this criterion, a model is evaluated both with respect to its graphics and goodness of fit. Sometimes, the Kolmogorov-Smirnov (KS) [5] goodness of fit test may be used to evaluate if Cox-Snell residuals follow a unit exponential distribution, but Kolmogorov-Smirnov test is poorly calibrated in AFT models [6]. An index plot of martingale and deviance residuals may be used to highlight individuals whose survival time is not well fitted by survival model [1]. Moreover, the common deficiency among all of these residuals are not sensitive to the violation of the survival model [7]. Therefore, it is difficult to determine whether a model effectively fits a dataset by using these traditional residuals. Martingale residual is a slight modification of Cox-Snell residual and is defined as the difference between the negative

logarithms of the estimated survivor function assigned to an individual with the observed survival time and its observed status. Martingale residuals take values between negative infinite and unity, with the residual for censored observations being negative. They are not symmetrically distributed, even when the model effectively fits the data [7]. Transforming martingale residual to achieve a more normal shaped distribution is helpful, and one such transformation is motivated by the deviance residuals found in generalized linear models literature [3]. The deviance residuals are much more symmetrically distributed around zero when the fitted model is appropriate. These traditional residuals have been proposed for use to diagnose the accelerated failure time (AFT) models in survival analysis.

In this thesis, a novel residual called Normal-transformed Randomized Survival Probability (NRSP) is proposed for diagnosing AFT models. The key idea of NRSP residual is to randomize the survival probability for censored observations into a uniform random number. NRSP residuals are computationally easy because the only information needed for computing them is survival function of the response variable. More specifically, it only requires inverting fitted survival function for each response variable and finding the corresponding standard normal quantile. For an event data, NRSP residuals are defined by taking the probit transformation of the survival function of the response variable in the AFT model. As a result, NRSP residuals are exactly normal under the true model when the parameters are known. We propose examining normality of the NRSP residuals based on Shaprio-Wilk (SW) test as the overall goodness-of-fit (GOF) test. The RSP residual uses the concept of randomization in randomized quantile residuals [8] [9] which is defined for diagnosing models for discrete response variables.

The purpose of this thesis is to demonstrate how to diagnose the AFT models in survival analysis using the NRSP residuals. Two simulation scenarios are considered including (1) misspecification of distribution assumption of survival time and (2) misidentification of functional form of covariate effect. Our simulation studies show that NRSP residuals are normal distributed under the true model, and in GOF tests, probabilities of rejecting the true model (type 1 error rates) are close to the nominal level 0.05, and powers of rejecting the wrong models are generally high. The research demonstrates the superiority of NRSP residuals when detecting model inadequacy as well as their insensitivity to the rate of censorship in

contrast to the traditional residuals.

2. Methodology

2.1 A brief introduction to survival analysis

Survival time is the duration from time origin till time of the event of interest. There are three requirements to determine survival time: (i) A time origin must be unambiguously defined, (ii) a scale for measuring the passage of time must be agreed and (iii) the definition of event of interest must be entirely clear.

Censoring is an important issue in survival analysis, representing a particular type of missing data. There are three types of censoring: 1) right censoring, 2) left censoring, and 3) interval censoring. Right censoring refers to the scenario when the event occurs after the observed survival time (follow up time); for example, event of interest occurs after the end of the study loss to follow-up during study period, or withdrawal from the study due of any reason. Left censoring occurs if the event of interest occurs before a time point, but do not know when it exactly happened. In this case the actual survival time is less than the observed censoring time. Yet another type of censoring is interval censoring, the event occurs within an interval of time. The most commonly encountered form of censoring is right censoring, which will be the focus of this thesis.

The actual survival time of an individual, t, can be regarded as the observed value of a random variable, T, that can take on any non-negative value. Let C denote the censoring time, that is, the time beyond which the study subject cannot be observed. The survival time starts at time origin and continues until the event of interest X or a censoring time C, whichever comes first. $T = min(X; C)$ is the follow-up time, and $\delta = I(X \leqslant C)$ is an indicator for status at the end of follow-up, The observed data are denoted by $(T\ ;\ \delta)$. The survival function is defined as the probability that the survival time is greater or equal to t,

$$S(t) = P(T \geqslant t), \tag{2.1}$$

For continuous time T (only consider continuous survival time in this thesis), the cumulative

density function (CDF) and the probability density function (PDF) of T are:

$$F(t) = 1 - S(t), \tag{2.2}$$

$$f(t) = F'(t) = -S'(t), \tag{2.3}$$

The hazard function gives the instantaneous failure rate at t given that the individual has survived up to time t,

$$h(t) = \lim_{\Delta T \to 0} \frac{P(t \leq T \leq t + \Delta T | T \geqslant t)}{\Delta T}, \tag{2.4}$$

The relationship between $h(t)$ and $S(t)$ is given by:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{\partial \log S(t)}{\partial t}, \tag{2.5}$$

$$S(t) = \exp\left(-\int_0^t h(w)dw\right) = \exp(-H(t)), \tag{2.6}$$

where $H(t) = \int_0^t h(w)dw$ is called cumulative hazard function, which can be derived from survival function by $H(t) = -\log S(t)$. From equation (2.1) the PDF of T can be written as

$$f(t) = h(t)S(t) = h(t)\exp\left(-\int_0^t h(w)dw\right), \tag{2.7}$$

If one of these functions is known, the other two are determined. One of these functions can be chosen as the basis of statistical analysis according to the particular situations. These three functions give mathematically equivalent specification of the distributions of the survival time t.

There are three approaches for regression in survival analysis:

- Non-parametric regression
- Semi-parametric regression
- Parametric regression

This thesis focuses on parametric regression models, i.e., accelerated failure time (AFT) models [1, 10].

2.2 Accelerated failure time regression model

The AFT model is one type of survival models commonly used in practice, in which explanatory variables measured on an individual are assumed to act multiplicatively on the time-scale [1].

Let $X = (X_1, X_2, ..., X_p)$ be the set of covariates. The survival function based on an AFT model is written as $S(t|X) = S_0(\frac{t}{\exp\{\eta(X)\}})$, where $S_0(t)$ is the baseline survival function and $\exp(\eta(X))$ is an "acceleration factor" that is a ratio of survival times corresponding to any fixed value of $S(t)$. The acceleration factor is given according to the formula $\exp(\eta(X)) = \exp(a_1X_1 + a_2X_2 + ... + a_pX_p)$.

In AFT model, the explanatory variables impact on survival by a time invariant factor, the acceleration factor. That is, the covariate effects are assumed to be constant and multiplicative on the time scale. According to the relationship of survival function and hazard function, the hazard function for an individual with covariates $X_1, X_2, ..., X_p$ is given by

$$h(t) = \exp(-\eta(X))h_0(\frac{t}{\exp(\eta(X))}), \tag{2.8}$$

The corresponding log-linear form of the AFT model with respect to time is given by

$$\log T_i = \mu + a_1X_{1i} + a_2X_{2i} + ... + a_pX_{pi} + \sigma\epsilon_i, \tag{2.9}$$

where μ denotes the intercept, σ is scale parameter and ϵ_i is a random variable, assumed to follow certain particular distribution. $a_1, \cdots, a_p$ represent the effects of the covariates on the survival time. Positive values indicate that the survival time increases with increasing values of the explanatory variable, and vice versa. For each distribution of ϵ_i, there is a corresponding distribution for T_i and the AFT models are named for the distribution of T_i rather than the distribution of ϵ_i or $\log T_i$ [1, 11–13]. Table 2.1 shows the distributions for commonly used parametric survival time and the associated error term for AFT models.

The survival and hazard functions for an AFT model are given by

$$S_i(t) = S_{\epsilon_i}(\frac{\log t - \mu - a_1X_{1i} - ... - a_pX_{pi}}{\sigma}), \tag{2.10}$$

$$h_i(t) = \frac{1}{\sigma t}h_{\epsilon_i}(\frac{\log t - \mu - a_1X_{1i} - ... - a_pX_{pi}}{\sigma}). \tag{2.11}$$

Table 2.1: Distributions for commonly used parametric survival time and the associated error term for AFT models.

Distribution of T_i	Distribution of ϵ_i
Exponential	Extreme value(one parameter)
Weibull	Extreme value(two parameters)/Gumbel
Log-logistic	Logistic
Log-normal	Normal

2.2.1 Weibull AFT regression model

Suppose survival time T_i has a Weibull distribution $W(\lambda, \gamma)$ with scale parameter λ and shape parameter γ, the baseline survival function and baseline hazard function in a Weibull regression model are given by

$$S_0(t) = \exp(-(\lambda t)^\gamma) \tag{2.12}$$

$$h_0(t) = \lambda\gamma t^{\gamma-1} \tag{2.13}$$

According to the equation (2.8), the hazard function for ith individual is :

$$h_i(t) = \lambda\gamma(t)^{\gamma-1} \exp(-(a_1X_{1i} + a_2X_{2i} + ... + a_pX_{pi}))^\gamma \tag{2.14}$$

where $\eta_i = a_1X_{1i}+a_2X_{2i}+...+a_pX_{pi}$ is the linear component of the model, in which X_{ji} is the value of the jth explanatory variable, X_j, $j = 1, 2, ..., p$ for the ith individual $i = 1, 2, ..., n$ [1]. The AFT representation of the survival function and hazard function for a Weibull regression model can be derived as follows. According to the equation (2.6), the AFT representation of the survival function of a Weibull model is given by

$$S_i(t) = \exp\{-\exp\{\frac{\log t - \mu - a_1X_{1i} - a_2X_{2i} - ... - a_pX_{pi}}{\sigma}\}\} \tag{2.15}$$

Based on equations (2.1) and (2.9), the AFT version of hazard function of a Weibull model is

$$h_i(t) = \frac{1}{\sigma}t^{\frac{1}{\sigma}-1} \exp\left(\frac{-\mu - a_1X_{1i} - ... - a_pX_{pi}}{\sigma}\right) \tag{2.16}$$

2.2.2 Log-normal AFT regression model

If survival times are assumed to have a log-normal distribution, the baseline survival function and baseline hazard function are given by

$$S_0(t) = 1 - \Phi(\frac{\log t - \mu}{\sigma}) \tag{2.17}$$

$$h_0(t) = \frac{\phi(\frac{\log t}{\sigma})}{\sigma t[1 - \Phi(\frac{\log t}{\sigma})]} \tag{2.18}$$

where ϕ and Φ are PDF and CDF of the standard normal distribution, μ and σ are unknown parameters. Under the AFT model, the survival function and hazard function for the ith individual are given by

$$S_i(t) = S_0(\exp(-\eta_i)t) = 1 - \Phi(\frac{\log t - \eta_i - \mu}{\sigma}) \tag{2.19}$$

$$h_i(t) = \frac{1}{\sigma t} \frac{\frac{1}{\sqrt{2\pi}} \exp(\log t - \eta_i - \mu)}{1 - \Phi(\log t - \eta_i - \mu)} \tag{2.20}$$

where $\eta_i = a_1 X_{1i} + a_2 X_{2i} + ... + a_p X_{pi}$ is the linear combination of the values of p explanatory variables for the ith individual. Therefore,

$$\log(T_i) \sim N(\mu + \eta_i, \sigma) \tag{2.21}$$

The log-normal distribution model has AFT property [1, 7].

2.3 Traditional residuals for checking models

2.3.1 Pearson residuals

Naive form of residual to adopt in AFT modelling is a standardized residual by applying Pearson's residuals [14] to $\log(t_i)$ defined by

$$r_i^s = \frac{\log t_i - \hat{\mu} - \hat{\alpha_1} x_{1i} - \hat{\alpha_2} x_{2i} - ... - \hat{\alpha_p} x_{pi}}{\hat{\sigma}} \tag{2.22}$$

where t_i is the observed survival time of the ith individual, and $\hat{\mu}, \hat{\sigma}, \hat{\alpha_j}, j = 1, 2, ..., p$, are the estimated parameters in the fitted AFT model [1].

2.3.2 Cox-Snell residuals

The Cox-Snell residuals are the estimated values of the negative logarithms of the survivor function for the ith individual with the observed survival time t_i [1, 2, 7].The estimated survivor function for the ith individual, is given by

$$\hat{S}_i(t) = S_\epsilon(\frac{\log t_i - \hat{\mu} - \hat{\alpha_1} x_{1i} - \hat{\alpha_2} x_{2i} - ... - \hat{\alpha_p} x_{pi}}{\hat{\sigma}}) \tag{2.23}$$

where $S_\epsilon(\epsilon)$ is the survivor function of ϵ in the AFT model, $\hat{\alpha_j}$ is the estimated coefficient of x_{ji} and $\hat{\mu}$, $\hat{\sigma}$ are the estimated values of μ and σ. The Cox-Snell residuals for parametric model are defined by

$$r_i^c = \hat{H}_i(t_i) = -\log \hat{S}_i(t_i) \tag{2.24}$$

Later, when we want to distinguish this original Cox-Snell residual with other variants of Cox-Snell residual, we will refer to it as "Unmodified Cox-Snell (UCS) residual". The main property of the UCS residual is that if the model fits the data and there is no censored times, r_i^c follows a standard exponential distribution, $r_i^c \sim \exp(1)$ with the PDF is $f(r_i^c) = \exp(-r_i^c)$. Hence, a straight line in graphing with unit slope and zero intercept indicates that the model fit data well, and only using this criterion are evaluated both graphically and goodness of fit. A r_i^c for the censored observations is

$$r_i^c = \hat{H}(t_i^*) = -\log \hat{S}(t_i^*) \tag{2.25}$$

where t_i^* is the right censoring time of the ith individual. If the model is truly fitted the cumulative hazard function of unit exponential distribution increases linearly with time, so the greater the value of survival time, the grater the value of the UCS residuals. Hence, the residual for the ith individual at the actual (unknown) failure time will be greater than the residual evaluated at the observed censored survival time. To count for underestimation of the UCS residuals for the censored observations, the following modification of the UCS residuals was proposed to make the residuals for censored observations compatible with a positive constant Δ for the uncensored observations, which can be called the excess residual [1]. Since r_i^c has a unit exponential distribution, the excess residual will also have a unit exponential distribution. The expected value of Δ is therefore unity the censored observation,

that is

$$r^{c\prime}_i = \begin{cases} r^c_i & \text{uncensored observations,} \\ r^c_i + \Delta & \text{censored observations.} \end{cases} \tag{2.26}$$

According to the equations (2.26), We will refer these residuals with $\Delta = 1$ as "Modified Cox-Snell (MCS) residual"; to be more explicit, MCS residuals can be expressed as:

$$r^{c\prime}_i = \begin{cases} r^c_i = -\log \hat{S}(t_i) & \text{uncensored observations,} \\ r^c_i + 1 = -\log \hat{S}(t^*_i) + 1 & \text{censored observations.} \end{cases} \tag{2.27}$$

where t_i is event time and t^*_i is the right censoring time of the ith individual. If the random survival time t_i is event time of the ith individual, it has survival function $S_i(t_i)$. If the random survival time t^*_i is the right censoring time of the ith individual, it has survival function $S(t^*_i)$. According to the equation (2.27), $r^{c\prime}_i = -\log \hat{S}(t^*_i) + 1 = -\log \hat{S}(t^*_i) + \log e = -\log\left(\frac{\hat{S}(t^*_i)}{e}\right)$. This is equivalent to modify the survival function at t^*_i as follows:

$$S'_i(t_i) = \begin{cases} S_i(t_i) & \text{uncensored observations,} \\ \frac{S_i(t^*_i)}{e} & \text{censored observations.} \end{cases} \tag{2.28}$$

Based on the MCS residuals, modified survival probability could be transformed to normal. We name this type of residual as Normal-transformed modified survival probability (NMSP) residual, which is defined as

$$r^{c*}_i = \Phi^{-1}(S'_i(t_i)). \tag{2.29}$$

2.3.3 Martingale residuals

The martingale residuals [3] provide a measure of the difference between the number of predicted of death by the model, and the number of observed failure in the interval $(0, t_i)$, which is either 1 or 0. The martingale residuals are defined by

$$r^M_i = \delta_i - r^c_i \tag{2.30}$$

where δ_i is the event indicator for the ith observation, so that δ_i is unity if that observation is an event and zero if censored, and r^c_i is the Cox-Snell residual. The martingale residuals for a parametric AFT model sum to zero, but are not symmetrically distributed about zero [1].

2.3.4 Deviance residuals

The deviance residuals [4, 15] can be regarded as an attempt to make the martingale residuals symmetrically distributed about zero, and are defined by

$$r_i^D = \text{sgn}(r_i^M)[-2(r_i^M + \delta_i \log(\delta_i - r_i^M))]^{\frac{1}{2}} \tag{2.31}$$

where r_i^M is the martingale residual for the ith individual, the function $sgn(.)$ is the sign function [1].

2.4 Problems with traditional residuals

In practice, analyzing traditional residuals of AFT models mostly relies on visual judgement. According to the extensive studies [1, 7], Cox-Snell residuals tend to have low power for detecting model misspecification. In addition, none of martingale and deviance residuals follows a particular distribution, although they are asymptotically distributed normally if there is no censored times [16]. When censoring occurs, there is not a clearly defined null distributions [1, 3, 4]. This implies that no objective criterion to test model inadequacy and the overall GOF test using KS test is not well-calibrated [6].

2.4.1 Illustrative examples for Cox-Snell residuals and deviance residuals

For Cox-Snell residual, if the model fits the data well, r_i^c follows an unit exponential distribution with density function $f(r_i^c) = \exp(-r_i^c)$. Let $S(r_i^c)$ denote the survival function of the Cox-Snell residual then

$$S(r_i^c) = \exp(-r_i^c) \tag{2.32}$$

which implies that

$$H(r_i^c) = -\log S(r_i^c) = r_i^c. \tag{2.33}$$

Let $\hat{S}(r_i^c)$ denote the Kaplan-Meier estimate of $S(r_i^c)$, a plot of r_i^c against $-\log \hat{S}(r_i^c)$ is expected to show a straight line with zero intercept and unit slope.

Martingale residuals are not symmetrically distributed around zero even when the fitted model is true, in fact they are approximately exponentially distributed [7] and they take value in the interval $(-\infty, 1)$. This makes plots based on these residuals difficult to interpret. To overcome this deficiency, Therneau et al. introduced deviance residuals which are much more symmetrically distributed around zero when the fitted model is appropriate, however they are not necessarily sum to zero. Deviance residuals are asymptotically normal [3]. An index plot of the deviance residuals can be used to identify individuals whose survival time is not well fitted by the model, such observation may be termed outliers. Plot of deviance residuals against covariates, to see if covariate effects are appropriately modeled.

2.4.2 Example 1: Assessing distributional assumption for survival time

Firstly, we will provide an illustrative example to assess distributional assumption using traditional diagnosis tools. The true model is a Weibull regression and a wrong model is the Lognormal regression. The survival data simulation consists of two parts, simulating real life times from Weibull distribution $T_i^* \sim W(\gamma, \lambda)$ where the shape parameter γ is set as 1.74 and the scale parameter λ is set as 1. The censored time is simulated from an exponential distribution $C_i \sim \exp(\theta)$ where the parameter $\theta = 0.22$ is chosen to yield a specified percentage of censorship [17–19]. The coding of the survival status of an individual, is denoted as a binary indicator, d_i such that zero denotes a censored observation and one denotes an event observation, the d_i equal to one if $T_i^* < C_i$ and d_i is zero if $T_i^* \geq C_i$. The time period in which an individual is in the study is known as the observed time T. The observed survival time T_i is T_i^* if $d_i = 1$, and T_i is C_i if $d_i = 0$.

We consider a dichotomized covariate $x \sim Bern(p)$ from Bernoulli distribution, where $p = 0.5$. We simulate data from a Weibull regression model: $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$ with the Weibull distribution where ϵ_i has standard extreme value distribution, and set $\beta_0 = 2, \beta_1 = 1$ with size $n = 800$ and censorship $c = 80\%$ samples. Then, we fit the Weibull and Lognormal AFT regression models to the simulated data.

The panels of the first row of Figure 2.1 display the Cox-Snell residuals r_i^c against

$-\log \hat{S}(r_i^c)$ under true and wrong AFT models. Under the true model, a portion of plotted points are not on a straight line, which is difficult to determine whether it has approximately unit slope; similarly, under the wrong model, more than half of plotted points are not on a straight line, which has not approximately unit slope. Both plots indicate that the two models fail to fit the data adequately, and the right of plot is worser than the left one. The panels the second row of Figure 2.1 display the index plots of deviance residuals for true and wrong models. Under the true model, all the residuals of censored data are below zero and the most of residuals of event data are randomly scattered between -2 and 3. Similarly, under the wrong model, all of the residuals for the censored data are below zero and the residuals of event data are randomly scattered with residual bounded between -1 and 4. Both residuals again fail to distinguish true and wrong models.

2.4.3 Example 2: Assessing functional form of covariate effect for survival time

In this section, we will provide an illustrative example to assess the functional form of covariate effect using traditional diagnosis tools. The response variable is simulated from a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 f(x) + \epsilon_i$. Then, a wrong model assuming $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$ is considered. The data simulation is same as section 2.4.2 with different parameter values: Weibull distribution $T_i^* \sim W(\gamma, \lambda)$ where the shape parameter γ is set as 1.8 and the scale parameter λ is set as 1, and exponential distribution $C_i \sim \exp(\theta)$ where the parameter $\theta = 2.6$ chosen to yield a specified percentage of censorship. We simulate a covariate $x \sim$ Uniform $(0, \frac{3\pi}{2})$ from a uniform distribution, and $f(x) = sin(2x)$. The value of coefficients set $\beta_0 = 2$, $\beta_1 = 5$ with size $n = 800$ and censorship $c = 80\%$.

The panels of the first row of Figure 2.2 display plots of Cox-Snell residuals r_i^c against $-\log \hat{S}(r_i^c)$ under true and wrong models. Under the true model, a portion of plotted points are not on a straight line, which is difficult to determine whether it has approximately unit slope; under the wrong model, most of points are not on a straight line. The panels of the second row of Figure 2.2 display plots of deviance residuals against covariates under true and wrong models. Under the true model, the residuals of event data are randomly scattered

and the residuals of censor data are clustered between -1 and 0; under the wrong model, the deviance residuals clearly indicate a sin function trend. As a result, it is challenging to differentiate true and wrong models based on either Cox-Snell or deviance residuals.

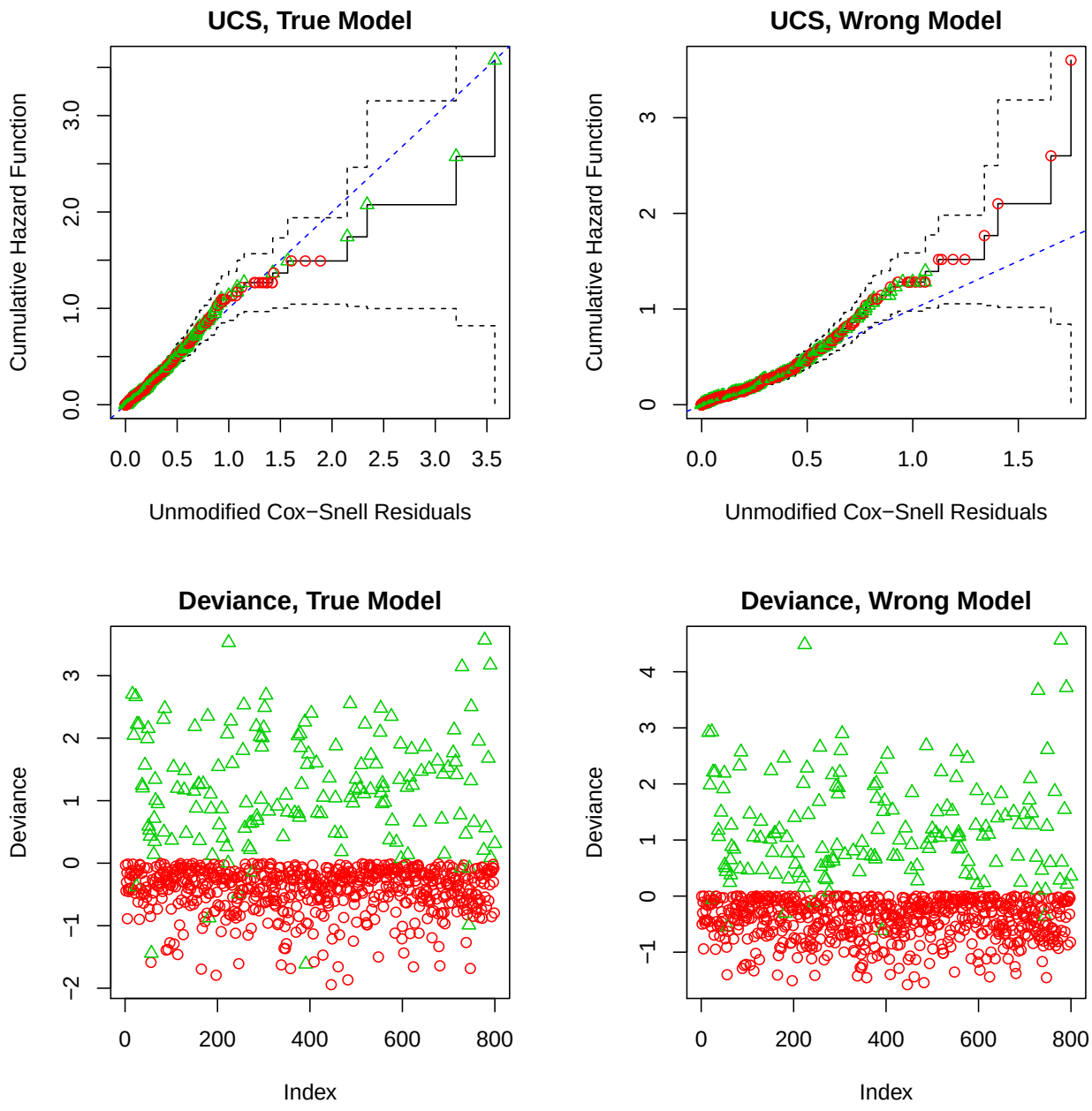

Figure 2.1: Unmodified Cox-Snell (UCS) residuals and deviance residuals for the true model (left panel) and the wrong model (right panel) for the first example in section 2.4.2. The green triangles correspond to the event times and the red circles correspond to the censored times.

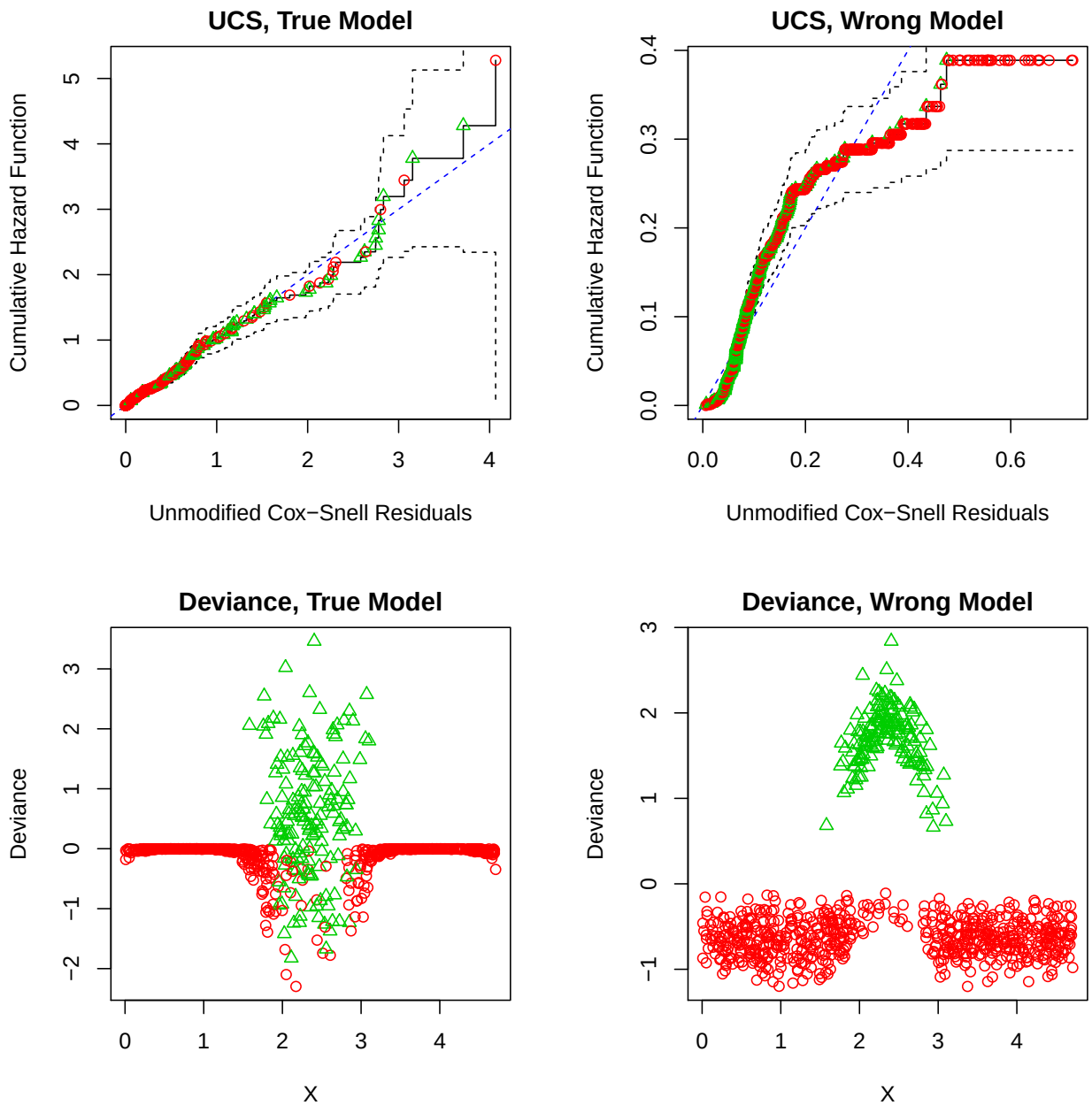

Figure 2.2: Unmodified Cox-Snell (UCS) residuals and deviance residuals for the true model (left panel) and the wrong model (right panel) for the second example in section 2.4.3. The green triangles correspond to the event times and the red circles correspond to the censored times.

3. Normal-transformed Randomized Survival Probability residual

3.1 Definition of Normal-transformed Randomized Survival Probability (NRSP) residual

The key idea of NRSP residual is to randomize the survival probability for censored observations into a uniform random number between 0 and $S_i(t_i^*)$. The innovation of in our method is to replace the e in the MCS residual to be a random number from unif [0,1]. Let $S(T_i)$ be the survival function if the random survival time T_i has survival function $S_i(T_i)$, then $S_i^*(T_i; u_i)$ is defined as follows:

$$S^*(T_i; u_i) = \begin{cases} S_i(T_i^*) & T_i^* < C_i (T_i \text{ is event time}) \\ u_i S_i(C_i) & T_i^* \geq C_i (T_i \text{ is censore time}) \end{cases} \tag{3.1}$$

where u_i is a uniform random variable on (0, 1]. Then, the NRSP residuals are defined as

$$q_i = q(T_i; u_i) = \Phi^{-1}(S^*(T_i; u_i)) \tag{3.2}$$

where $\Phi()$ is the cumulative distribution function (CDF) of a standard normal distribution.

As it can be seen from the definition, the NRSP residual has a straightforward definition for all distributions. The only information that is necessary for computing NRSP residual of AFT models is knowing the cumulative incidence function or survival function of the survival time variable, which is a great advantage over, for instance, deviance residuals which requires derivation of the saturated model or Cox-Snell residuals which is not necessarily exponentially distributed with unit mean for censored observations.

3.2 Illustrative example

To demonstrate the idea of Randomized Survival Probability (RSP) for diagnosing AFT model, an illustrative example is presented in this section. The scenario of the illustrative example is similar to the example presented in section 2.4.2, in which the survival times are simulated from a Weibull regression model with the parameter set as $\gamma = 1.784$ and the censoring times are simulated from an exponential distribution with parameter set as $\theta = 0.08$, at sample size of 2000 and the percentage of censoring is about 50 %. For the ease of visualization, we randomly sampled a subset of 400 data points from the simulated sample.

The key idea of RSP can be conveyed graphically such that the survival probability for the events should fall along a theoretical survival line, and the censor data should be randomized into a uniform distribution from 0 to 1 between the discontinuity gap of survival function in the survival curve. Now suppose the data are simulated from Weibull AFT regression models. Overall, the random numbers converted with S^* are uniformly distributed on (0,1] under the true model. As depicted in the first row of Figure 3.1, $S_i^*(T_i; u_i)$ is uniformly distributed between 0 and 1. However, under the log-normal AFT regression model, $S_i^*(T_i; u_i)$ is concentrated on middle from the second row of Figure 3.1, indicating that $S_i^*(T_i; u_i)$ is not uniformly distributed.

On the other hand, both of the event data and censor data should be on the theoretical survival line in survival curve for the unmodified survival probability. As depicted in Figure 3.2, survival probability is not uniformly distributed between 0 and 1 under the true and wrong models. Under the modified survival probability, the survival probabilities of the event data are still on the theoretical survival line, but the survival probabilities of the censored data are devide by e. Figure 3.3 shows that survival probability is not uniformly distributed between 0 and 1 under both the true and wrong models.

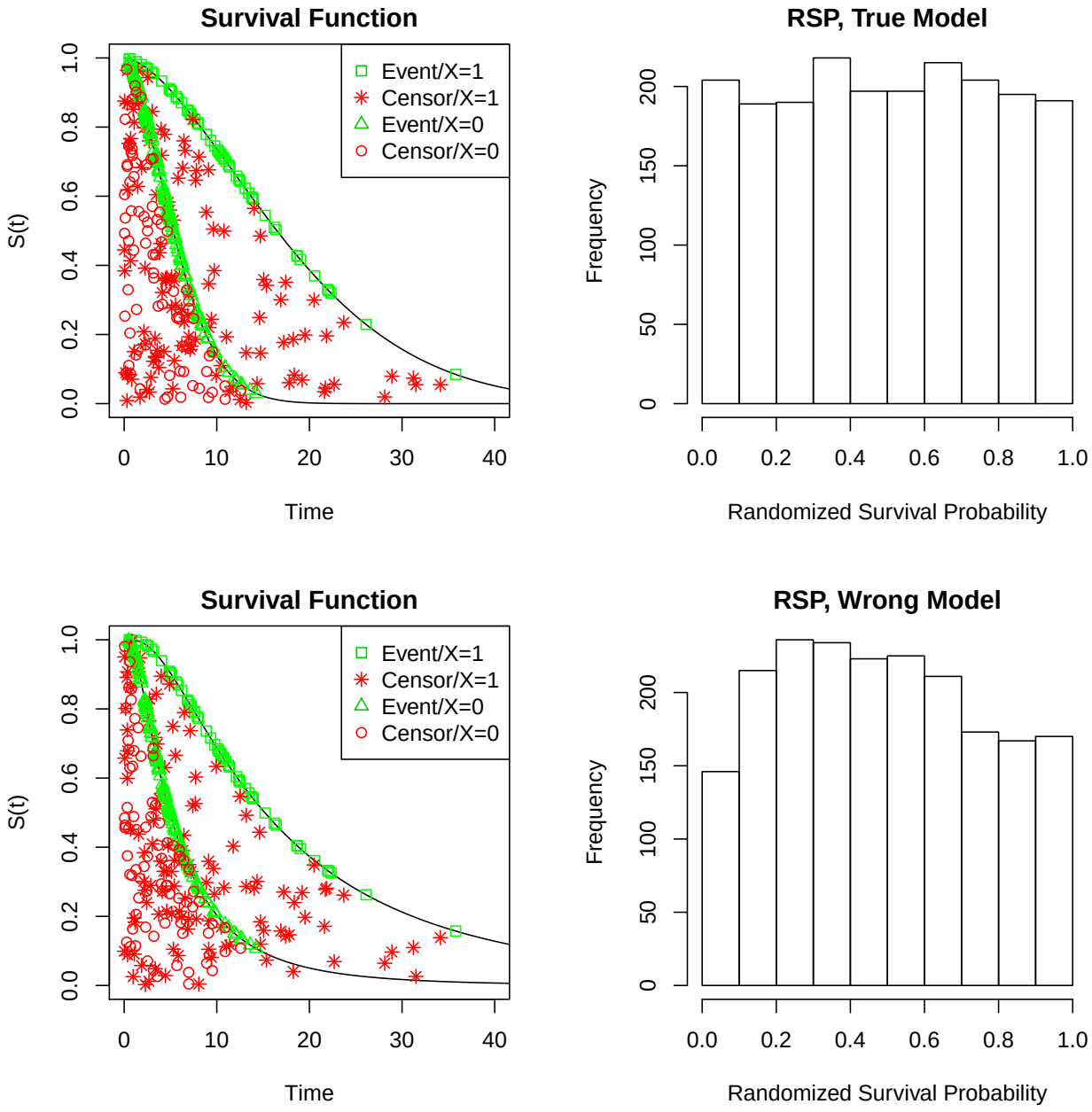

Figure 3.1: RSP for the true model (first row) and the wrong model (second row). The left panels are randomized survival functions and right panels are the randomized histogram of RSPs .

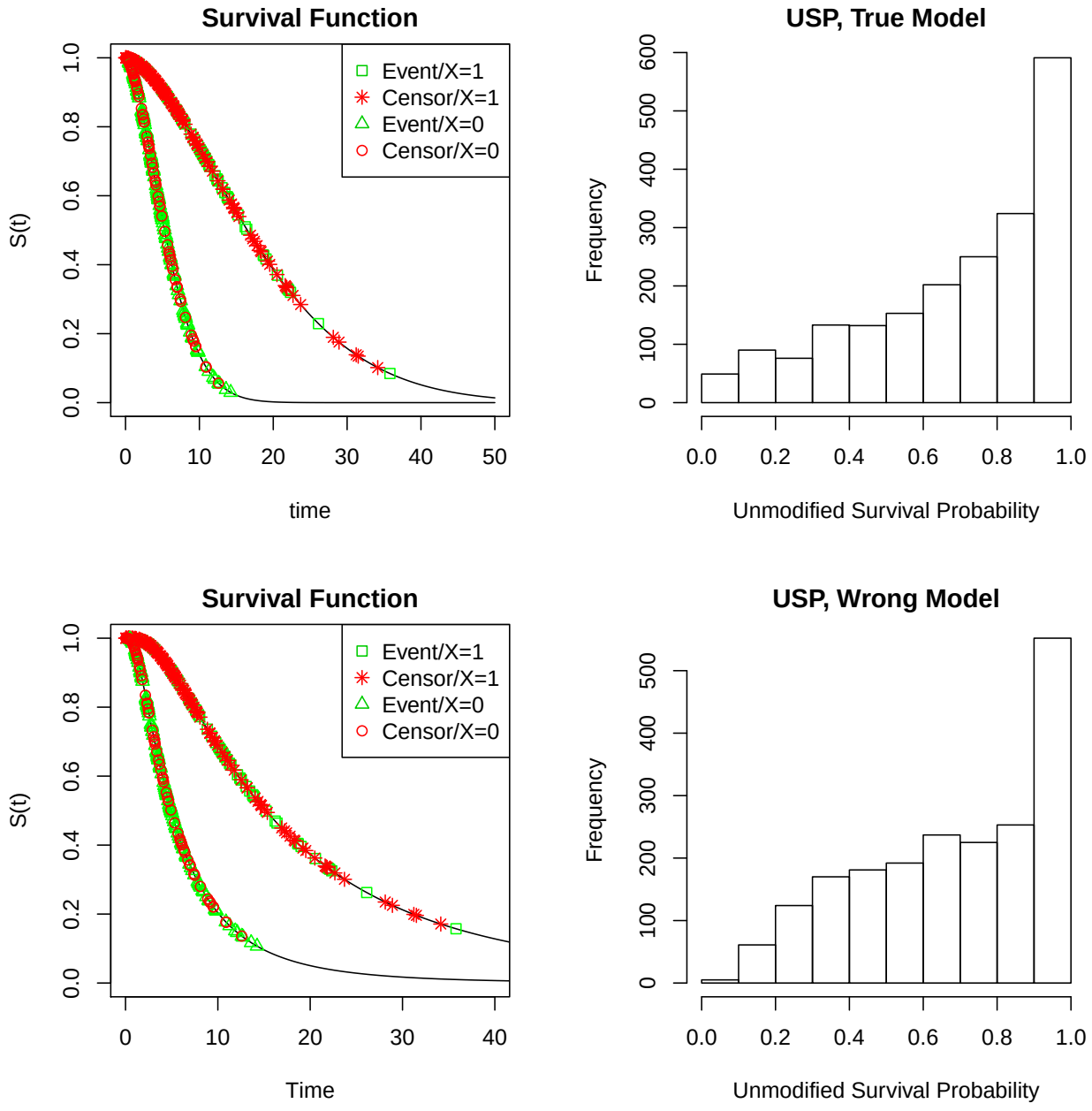

Figure 3.2: Unmodified survival probability (USP) for the true model (first row) and the wrong model (second row). The left panels are survival functions and right panels are the histogram of USPs.

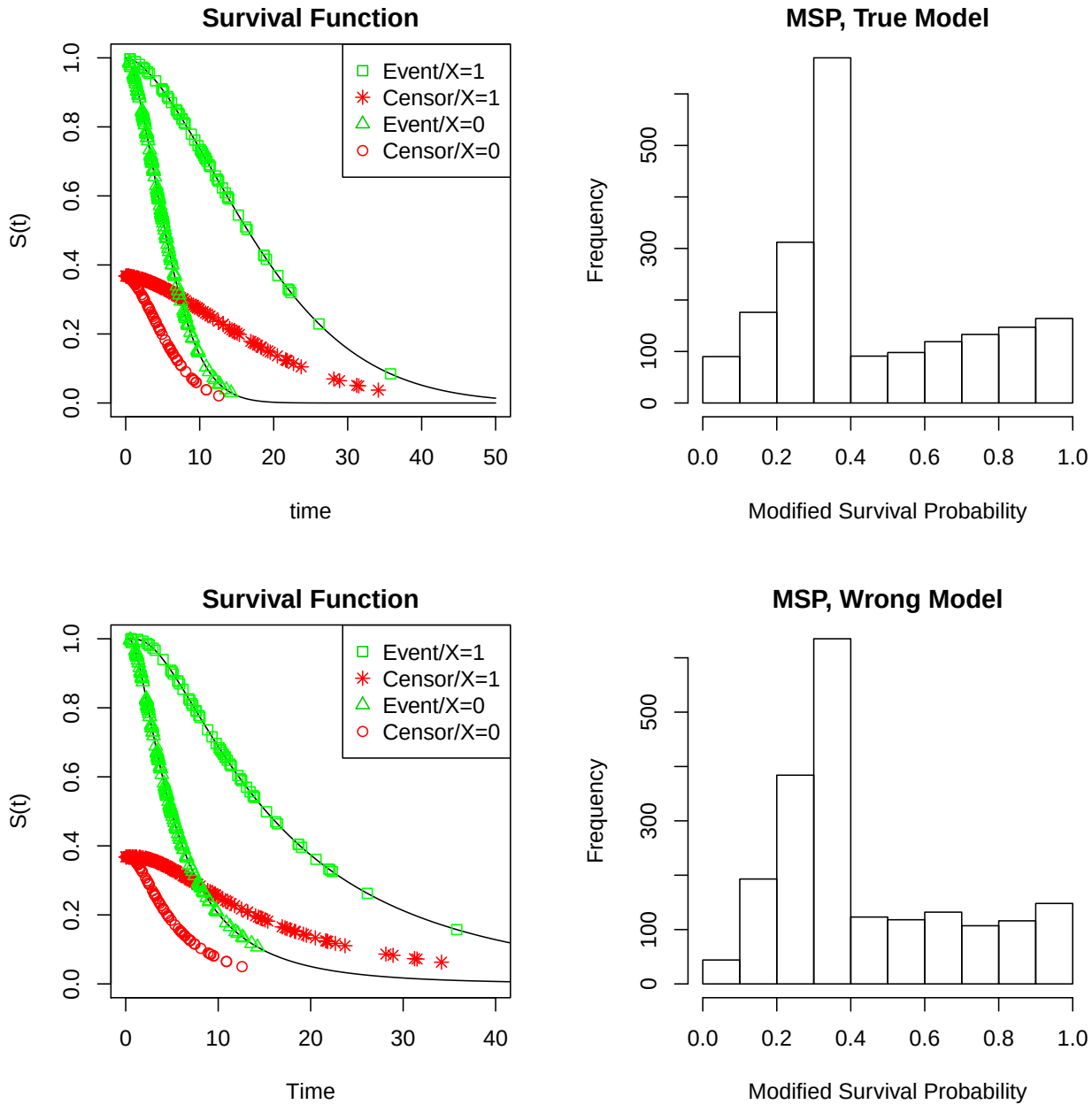

Figure 3.3: Modified survival probability (MSP) for the true model (first row) and the wrong model (second row). The left panels are survival functions and right panels are the histogram of MSPs.

4. Simulation studies

In this Chapter, we will investigate the performance of NRSP residuals by comparing them with NMSP and deviance residuals via simulation studies. The simulation consists of two scenarios by assessing (1) the distributional assumption for the survival response variable and (2) the functional form of covariate effect. For each simulation setting, the performance of NRSP, NMSP and deviance residuals are compared for identifying model misspecification. The Shapiro-Wilk (SW) test and Kolmogorov–Smirnov (KS) test for testing normality of residuals are the GOF test used in the current study. Then, the GOFs are assessed based on a comparison with the mis-specified models using NRSP, NMSP and deviance residuals. This experiment is replicated by simulating 1000 datasets from the true model simultaneously to assess the performance of overall GOF test by testing the normality of the residuals. The histogram of normality test p-values are presented for comparing the performance of various types of residuals. Furthermore, to gain more insights of the finite-sample performance, a power analysis is performed by setting the sample sizes $n = 100, 200, 400, 600, 800, 1000$ and the percentage of censorship $c = 20\%, 50\%, 80\%$. The null and alternative hypotheses are defined as H_0: the model fits the data well versus H_a: the model does not fit the data well. Under each simulation scenario, the type I error rate and statistical power are examined. The type I error rate is defined as the probability of rejecting the true model under the true model. The statistical power is defined as the probability of rejecting a wrong model. Ideally, a desirable GOF test should give a type I error close to the nominal level while providing high statistical power. We also compare the model selection performance of NRSP with Akaike's information criterion (AIC) [1]. The smaller the value of AIC, the better the model. Based on the AIC values from the 1000 replicated samples, the percentage of the difference value greater than 4 [20] and the mean of the difference value are calculated to measure the differences between the two models for each setting.

4.1 Assessing distributional assumption for a survival model

In this simulation setting, we will assess the power of NRSP residual in detecting misspecification of distributional assumption of survival time in comparison with traditional residuals, where the survival data are generated in a similar way as the illustrative example presented in section 2.4.2. The sample of size is set as $n = 800$ and the parameter of the exponential distribution is set as $\theta = 0.08$ and the shape parameter of Weibull distribution is set as $\gamma = 1.784$. The parameters are selected to give a percentage of censorship c equal to 50%. For the simulated dataset, the true and wrong models are fitted and then different types of residuals are computed. To examine the normality of the NRSP, NMSP and deviance residuals, the quantile-quantile (QQ) plots are presented. We further presented the GOF tests, i.e., the SW and KS tests to test the normality of the residuals. Moreover, power analysis is conducted by varying the percentage of censoring at various sample size for a more in-depth investigation of the proposed method.

4.1.1 Results of a single simulation scenario

In this Section, the performance of the NRSP residuals with respect to detecting distributional assumption is evaluated based on a single dataset. The panels of the first column of Figure 4.1 display the NRSP residuals against the fitted values under the true (Weibull) and wrong (Log-normal) models. Under the true model, NRSP residuals are randomly scattered without exhibiting any pattern and the standardized residuals are mostly within -3 to 3. Conversely, under the wrong model, NRSP residuals are clustered in the middle with residuals scattered mostly in [-2,4]. The panels in the second column of Figure 4.1 present the QQ plots of the NRSP residuals under the true and wrong models. Under the true model, the points in the QQ plot align almost perfectly on the diagonal line. Under the wrong model, however, the points deviate from the diagonal line in both the upper and lower tails. To examine the sensitivity of the overall GOF due to randomization, the current study replicates this experiment by simulating 1000 datasets from the true model and then we apply

the SW and KS tests to evaluate the normality of the NRSP residuals. The panels in the third column of Figure 4.1 presents the histograms of 1000 SW p-values under the true and wrong models. The p-values of the SW test for the NRSP residuals under the true model are uniformly distributed, indicating the well-calibration of this overall GOF test. In contrast, under the wrong model, the p-values of the SW test for the NRSP residuals are concentrated around zero, implying that the wrong model will be rejected most of times at a small nominal threshold, such as 0.05. Thus, the overall GOF test via the SW test for the NRSP residuals confirms the great power in detecting the wrong model. The panels in the fourth column of Figure 4.1 present the histograms of 1000 KS p-values under the true and wrong models. The p-values of the KS test for the NRSP residuals are skewed right from 0 to 1 under the true model, which indicates the KS test is too conservative in rejecting the true model. Similarly, the p-values of the KS test for the NRSP residuals are highly skewed left under the wrong model; this indicates KS test is too conservative in rejecting the wrong model.

The current research also demonstrates that the performance of the NMSP and deviance residuals with regard to their ability to detect distributional assumptions in linearity covariate effects is evaluated based on a single dataset. The panels of the first column of Figure 4.2 and Figure 4.3 display the NMSP and the deviance residuals against the fitted values under the true (Weibull) and wrong (Log-normal) models. Under the true model, the NMSP residuals of the event data are randomly scattered and do not exhibit any pattern, and the residuals of the censor data are clustered between -1 and 0 with residual bounded in [-2,3]. Similarly, under the wrong model, the NMSP residuals of the event data are randomly scattered and do not exhibit any pattern; however, a few outliers and the residuals of the censor data are clustered between -1 and 0 with residual bounded in [-2,5]. Meanwhile, deviance residuals of the censored data are below zero and the residuals of event data are randomly scattered under both of the true and wrong models. The QQ plots for the NMSP and deviance residuals are depicted in the panels of the second column of Figure 4.2 and Figure 4.3. The NMSP and deviance residuals do not follow a normal distribution under either the true or wrong models, which indicates that the NMSP and deviance residuals fail to correctly diagnose the true model. We also simulated 1000 datasets from the true model, The SW test and KS test are applied to evaluate the normality of the NMSP and deviance residuals. The panels in the

third and fourth columns of Figure 4.2 and Figure 4.3 indicate both of the SW p-values and KS p-values are all concentrated around zero under the true and wrong models, implying that all of the true and wrong models will be rejected most of times at a small nominal threshold, such as 0.05. Therefore, the NMSP and deviance residuals fail to distinguish models.

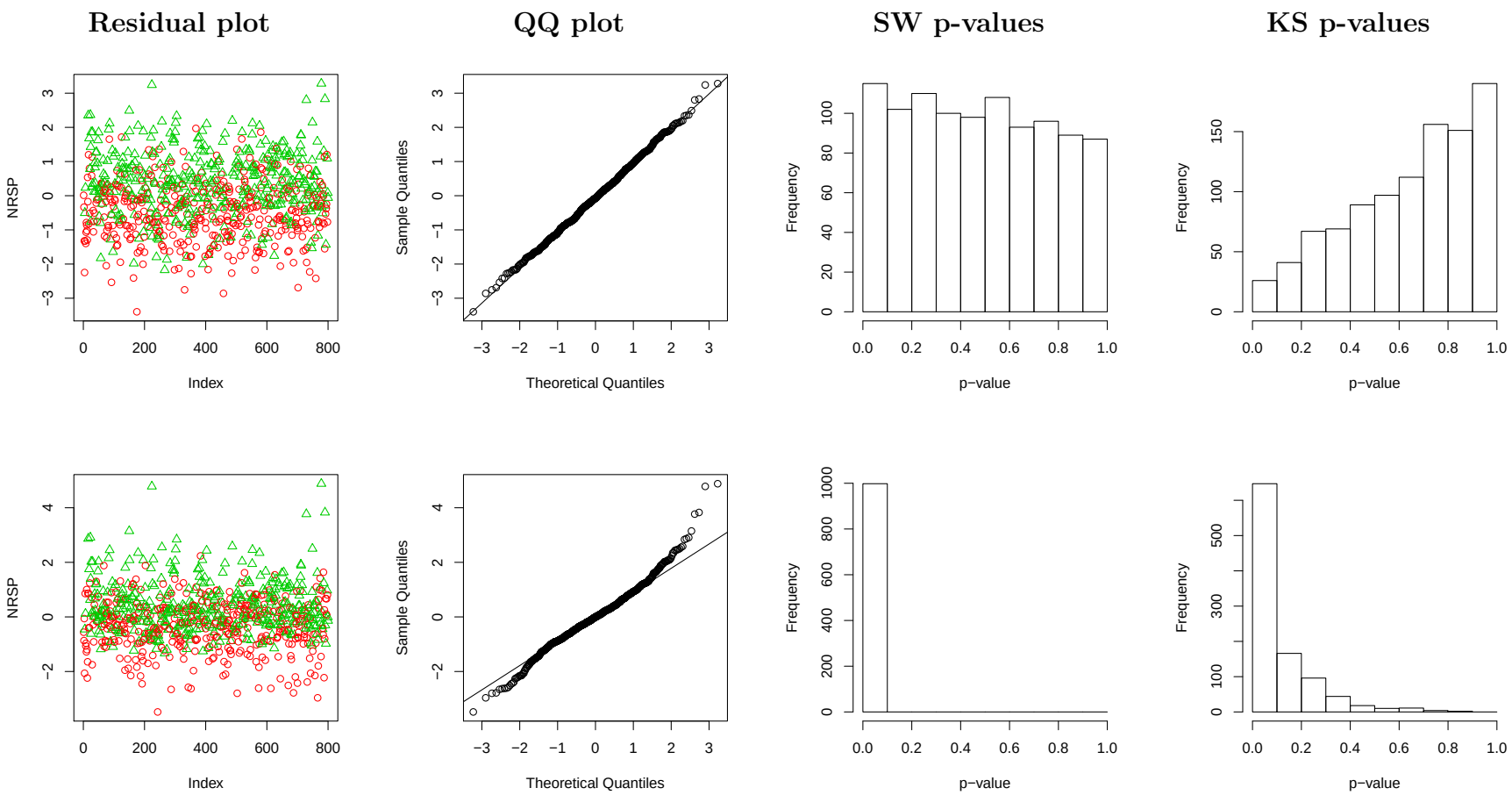

Figure 4.1: Performance of the NRSP residuals in detecting distributional assumption of a sample dataset of size $n = 800$ and a percentage of censorship $c = 50\%$. The panels in the first row present the NRSP residuals for the true model: Weibull distribution. The panels in the second row present the NRSP residuals for the wrong model: Log-normal distribution. The first two columns display the scatter plots and QQ plots of the NRSP residuals, respectively. The third and fourth columns present the histograms of the SW and KS p-values for the NRSP residuals over 1000 randomly generated datasets from the true model. The green triangles correspond to the event times and the red circles correspond to the censored times.

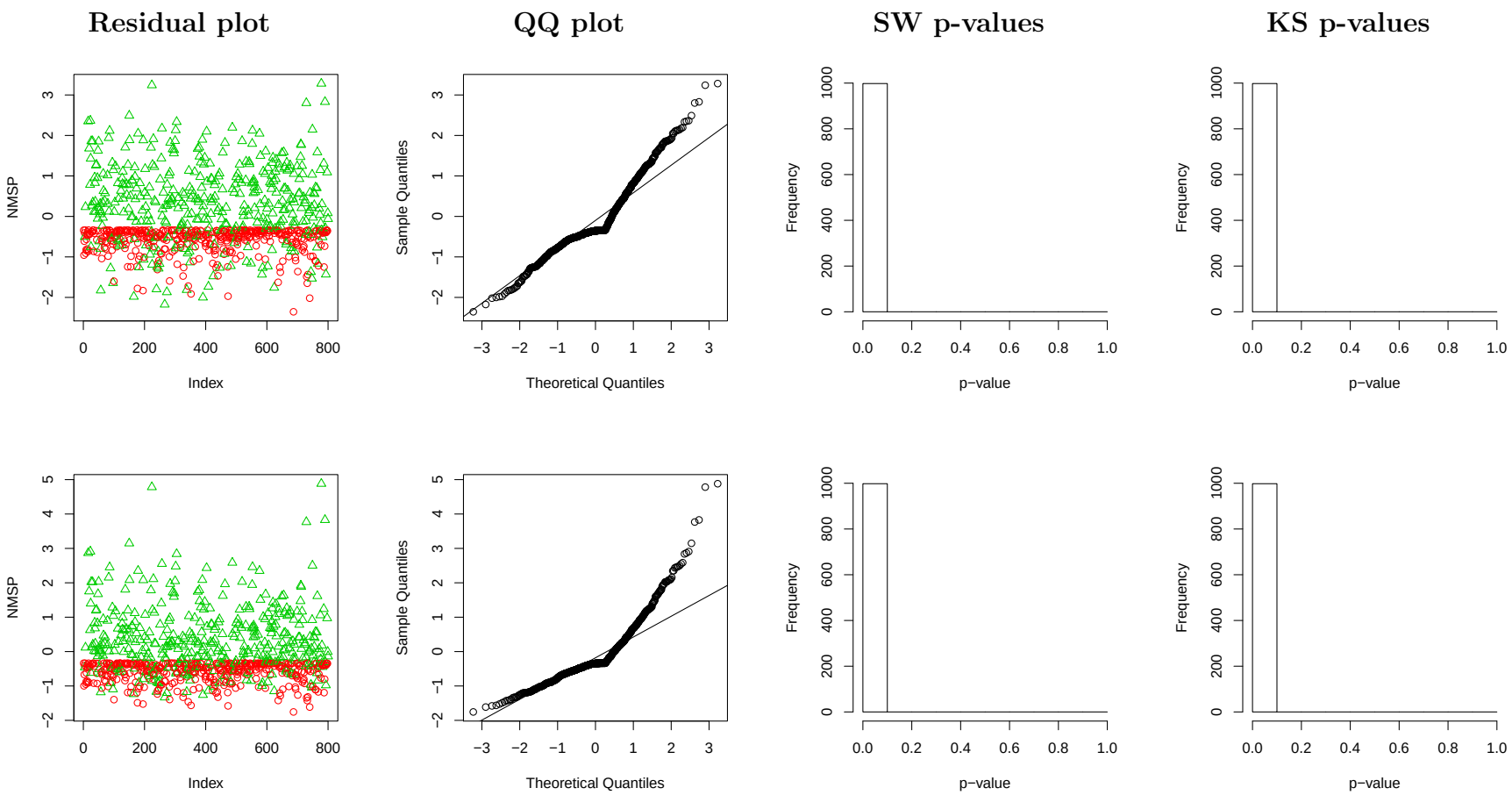

Figure 4.2: Performance of the NMSP residuals in detecting distributional assumption of a sample dataset of size $n = 800$ and a percentage of censorship $c = 50\%$. The panels in the first row present the NMSP residuals for the true model: Weibull distribution. The panels in the second row present the NMSP residuals for the wrong model: Log-normal distribution. The first two columns display the scatter plots and QQ plots of the NMSP residuals, respectively. The third and fourth columns present the histograms of the SW and KS p-values for the NMSP residuals over 1000 randomly generated datasets from the true model. The green triangles correspond to the event times and the red circles correspond to the censored times.

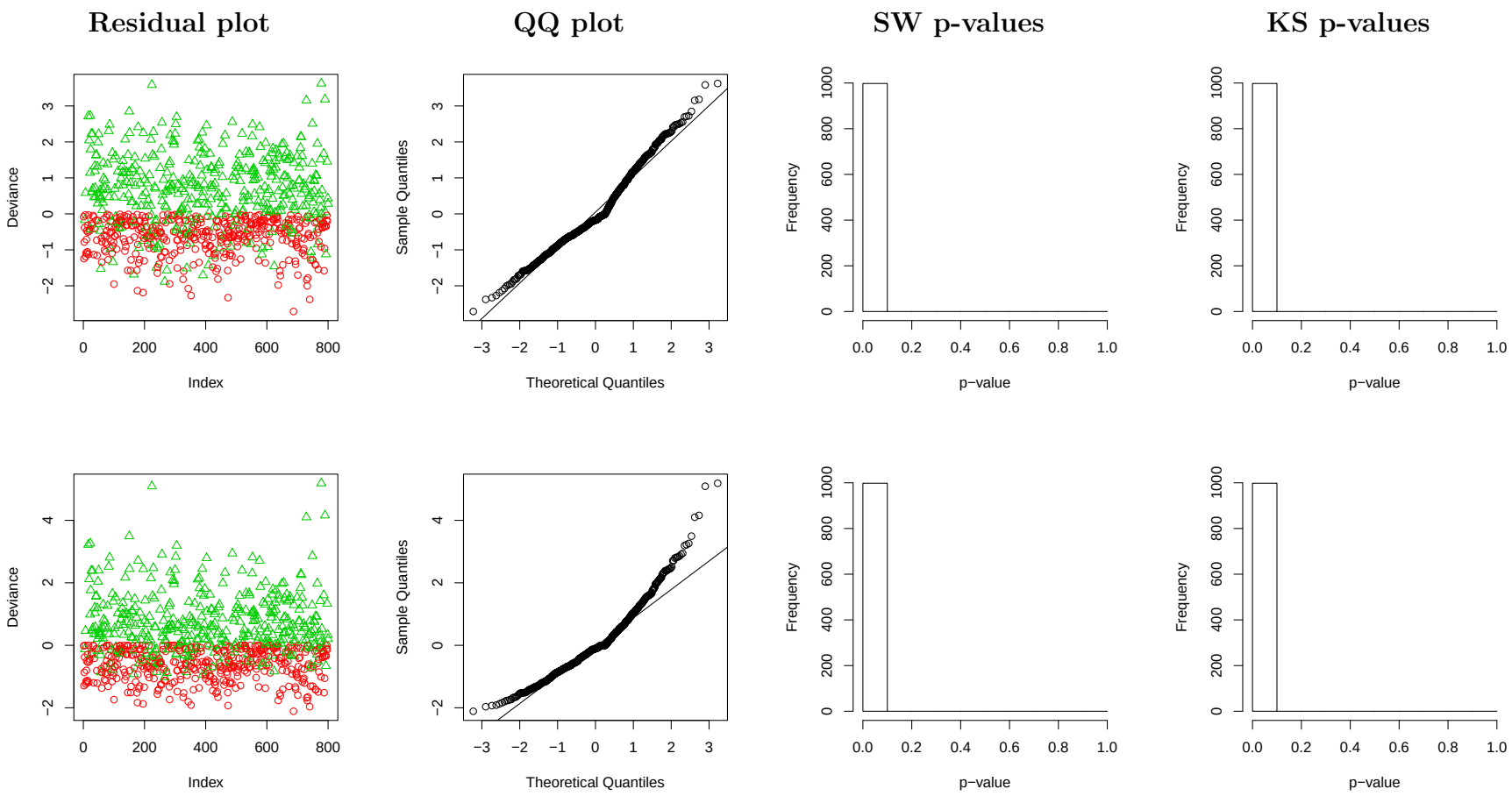

Figure 4.3: Performance of the deviance residuals in detecting distributional assumption of a sample dataset of size $n = 800$ and a percentage of censorship $c = 50\%$. The panels in the first row present the deviance residuals for the true model: Weibull distribution. The panels in the second row present the deviance residuals for the wrong model: Log-normal distribution. The first two columns display the scatter plots and QQ plots of the deviance residuals, respectively. The third and fourth columns present the histograms of the SW and KS p-values for the deviance residuals over 1000 randomly generated datasets from the true model. The green triangles correspond to the event times and the red circles correspond to the censored times.

4.1.2 Power analysis

To further evaluate the performance the NRSP residuals, power analysis is performed by setting the sample sizes at $n = 100, 200, 400, 600, 800, 1000$ and the percentage of censorship $c = 20\%, 50\%$ and 80%. We examine the probability of rejecting the true model (type I errors) and the probability of rejecting the wrong model (statistical power). As shown in Figure 4.4, the type I errors are consistently retained at a nominal level 0.05 for all scenarios based on the SW p-values for the NRSP residuals. In contrast, the type I errors for the SW tests for NMSP and deviance residuals are significantly above 0.05 as their SW p-values are incorrectly distributed near 0 when the true model is fitted. Thus, these results indicate the superior performance of SW test for NRSP residuals as the GOF test for model checking as compared to traditional survival residuals. The first row of Figure 4.5 indicates that the type I errors of the KS test are consistently lower than nominal level 0.05 for all scenarios, moreover, a portion of scenarios give SW p-values for NRSP residuals close to zeros. This provides further evidence that KS test is too conservative. In contrast, the type I errors for the KS tests for the NMSP and deviance residuals are significantly above 0.05. Furthermore, Figure 4.4 demonstrate that the NRSP residuals, the NMSP and deviance residuals have high statistical power. Although significantly high power results are obtained for all of residuals, the high type I errors make the GOF test based on SW test is impractical and undesirable. The second row of Figure 4.5 indicates that the statistical power of KS test for the NRSP residuals under wrong model are between 0 to 0.05 when sample size is small, or the percentage of censorship is larger, which means KS test does not reject some wrong models.

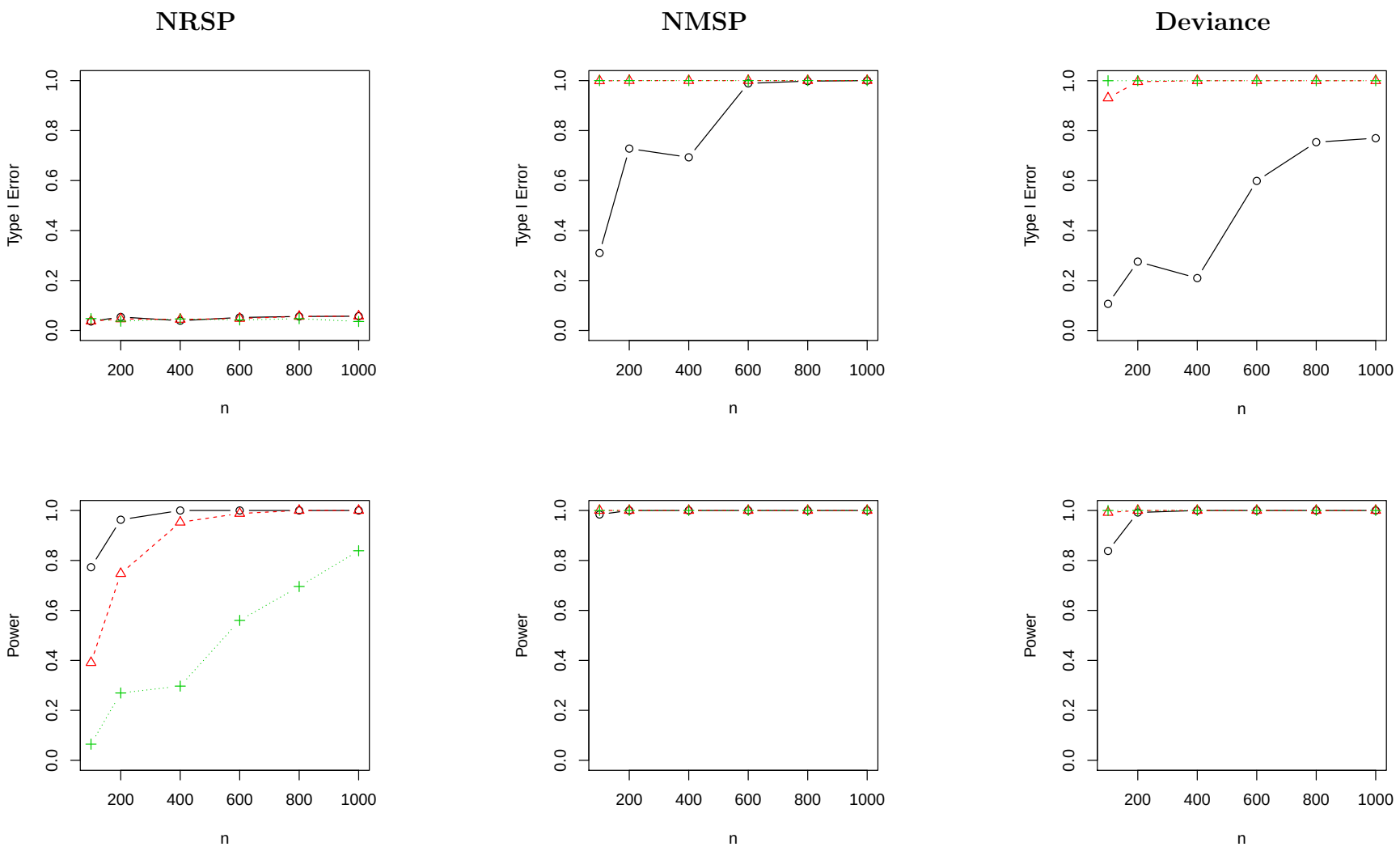

Figure 4.4: Comparison of the type I errors and powers of the SW tests for the NRSP, NMSP, and deviance residuals. Response variable is simulated from the true model at varying sample sizes of $n = 100$, 200, 400, 600, 800 and 1000, and the percentage of censorship $c = 20\%$ (black circles), 50% (red triangles) and 80% (green crosses). True model: Weibull model. Wrong model: Log-normal model.

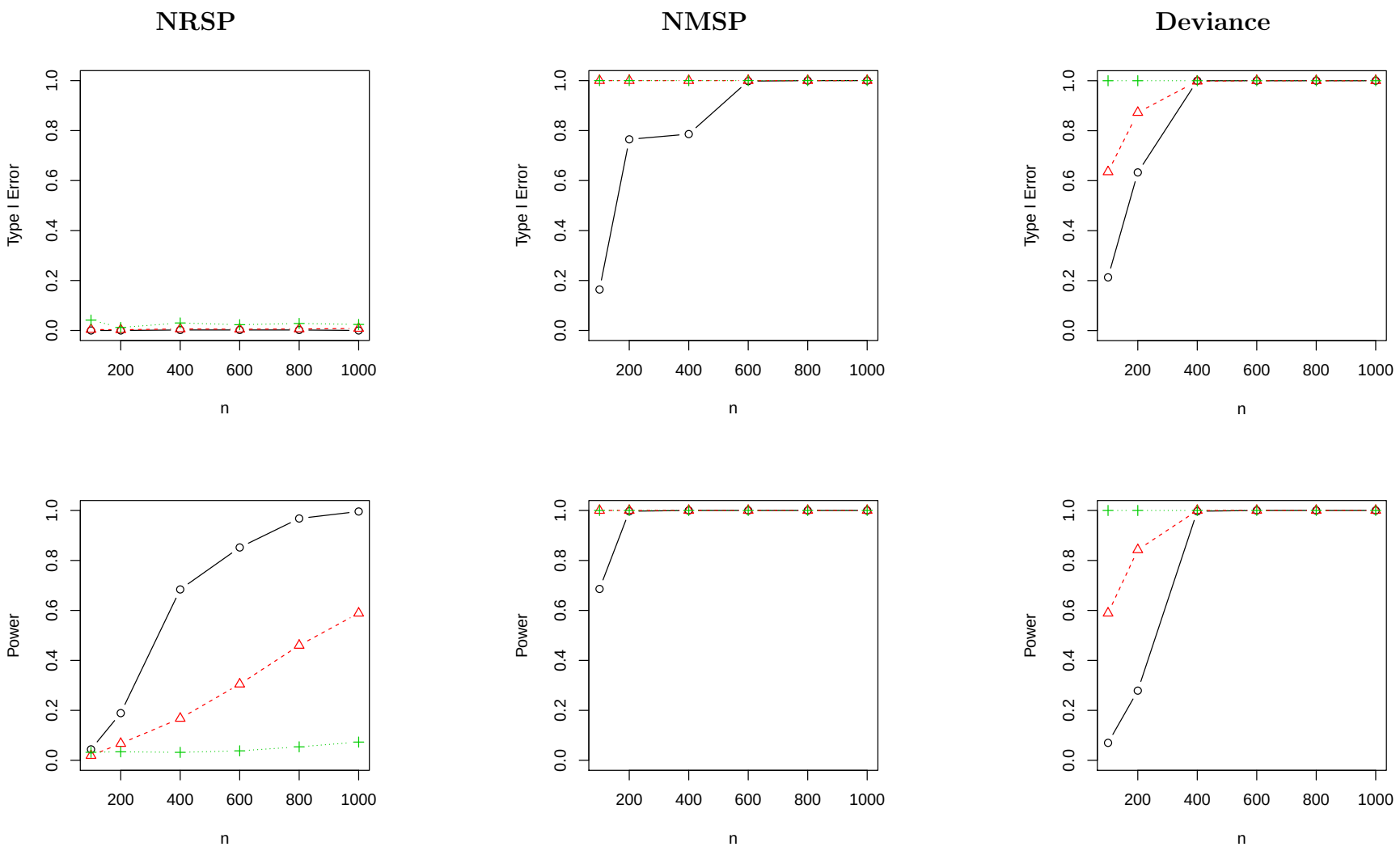

Figure 4.5: Comparison of the type I errors and powers of the KS tests for the NRSP, NMSP, and deviance residuals. Response variable is simulated from the true model at varying sample sizes of $n = 100$, 200, 400, 600, 800 and 1000, and the percentage of censorship $c = 20\%$ (black circles), 50% (red triangles) and 80% (green crosses). True model: Weibull model. Wrong model: Log-normal model.

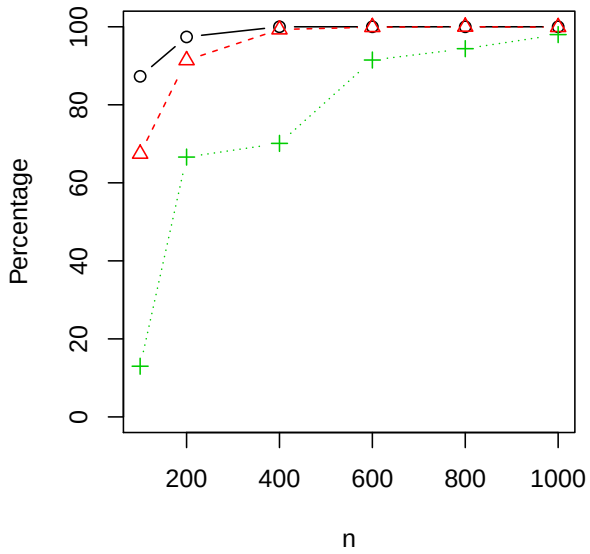

Figure 4.6: AIC for true model (Weibull regression) and wrong model (Log-normal regression) at varying sample sizes of $n = 100$, 200, 400, 600, 800 and 1000, and the percentage of censorship $c = 20\%$ (black circles), 50% (red triangles) and 80% (green crosses).

4.1.3 Model comparisons

To confirm the performance of the proposed residual diagnosis tool in comparison with traditional residuals in survival analysis, we further compare the true and wrong models based on AIC in all the simulation settings. The percentage of the difference value of AIC greater than 4 [20], and the mean of the difference value of AIC are between the true and wrong models based on 1000 replicated samples. Figure 4.6 presents the percentage of the difference value as greater than 4, which increases with the increased the sample size and decreased censorship.

4.2 Assessing functional form of the covariate effect

We simulated a dataset of size $n = 800$, and adjusted the parameter of exponential distribution $\theta = 0.024$ so that a percentage of censorship was $c = 50\%$. Models with two different functional forms are diagnosed by NRSP, NMSP and deviance residuals. To examine the normality of different residuals, QQ plots are presented, and we further presented histogram

the p-values of the normality tests, i.e., of SW and KS tests. The power analysis and the AIC also be used to confirm the consistency of results. The same approach was implemented to investigate the performance of the NRSP and the NMSP and deviance residuals with respect to their ability to assess the functional form of the covariate effect under the same distribution assumption in survival model. The simulation survival data and the model fit the data are same in second illustrative example in section 2.4.3.

4.2.1 Results of a single simulation scenario

The performance of the NRSP residuals with respect to their ability to detect the functional form of the covariate effect is evaluated based on a single simulated dataset. The panels of the first column of Figure 4.7 display the NRSP residuals against the covariate under the true and wrong models. The NRSP residuals confirms that the true model fits the data well with residuals randomly scattered without exhibiting any pattern and being bound mostly between -3 and 3; whereas, the wrong model does not fit the data well as residuals are scattered as a sin functional trend. The panels of the second column of Figure 4.7 present the QQ plots of the NRSP residuals under the true and wrong models. Under the true model, the QQ plot almost perfectly aligns with the diagonal line, but under the wrong model, the QQ plot deviates from the diagonal line in the upper tail. The p-values of SW test and KS test based on the 1000 repeated samples, as displayed in the panels in the third column of Figure 4.7 and the fourth column of Figure 4.7. The third column presents the p-values of the SW test for the NRSP residuals under the true model, which are uniformly distributed, indicating the effective calibration of this overall GOF test. In contrast, the p-values of the SW test for the NRSP residuals are distributed around zero under the wrong model, implying that the wrong model will be rejected most of times at a small nominal threshold, such as 0.05. Thus, the overall GOF test via the SW test for the NRSP residuals confirms its ability to detect wrong model. The fourth column presents the p-values of the KS test for the NRSP residuals, which are right skewed from 0 to 1 under the true model; this indicates KS test is too conservative in rejecting the true model, though the KS p-values under the wrong model are distributed around 0.

The performance of the NMSP and deviance residuals with respect to their ability to

detect the functional form of the covariate effect firstly was evaluated based on a single simulation setting. The panels of the first column of Figure 4.8 and Figure 4.9 display the NMSP and the deviance residuals against the covariate under the true and wrong models, respectively. Under the true model, the NMSP and deviance residuals of the event data are randomly scattered, and the residuals of the censor data are clustered between -1 and 0. Under the wrong model, the NMSP and deviance residuals clearly indicate a sin function trend. The QQ plots for the NMSP and deviance residuals are depicted in the panels of the second column of Figure 4.8 and Figure 4.9, which show that the points deviate from the diagonal line under both the true and wrong models. Hence, NMSP and deviance residuals fail to correctly diagnose the true model. The panels in the third and fourth columns of Figure 4.8 and Figure 4.9 are based on the repeated samples and demonstrate the SW p-values and KS p-values are all mostly distributed around 0 under both true and wrong models. Therefore, the NMSP and deviance residuals fail to distinguish models.

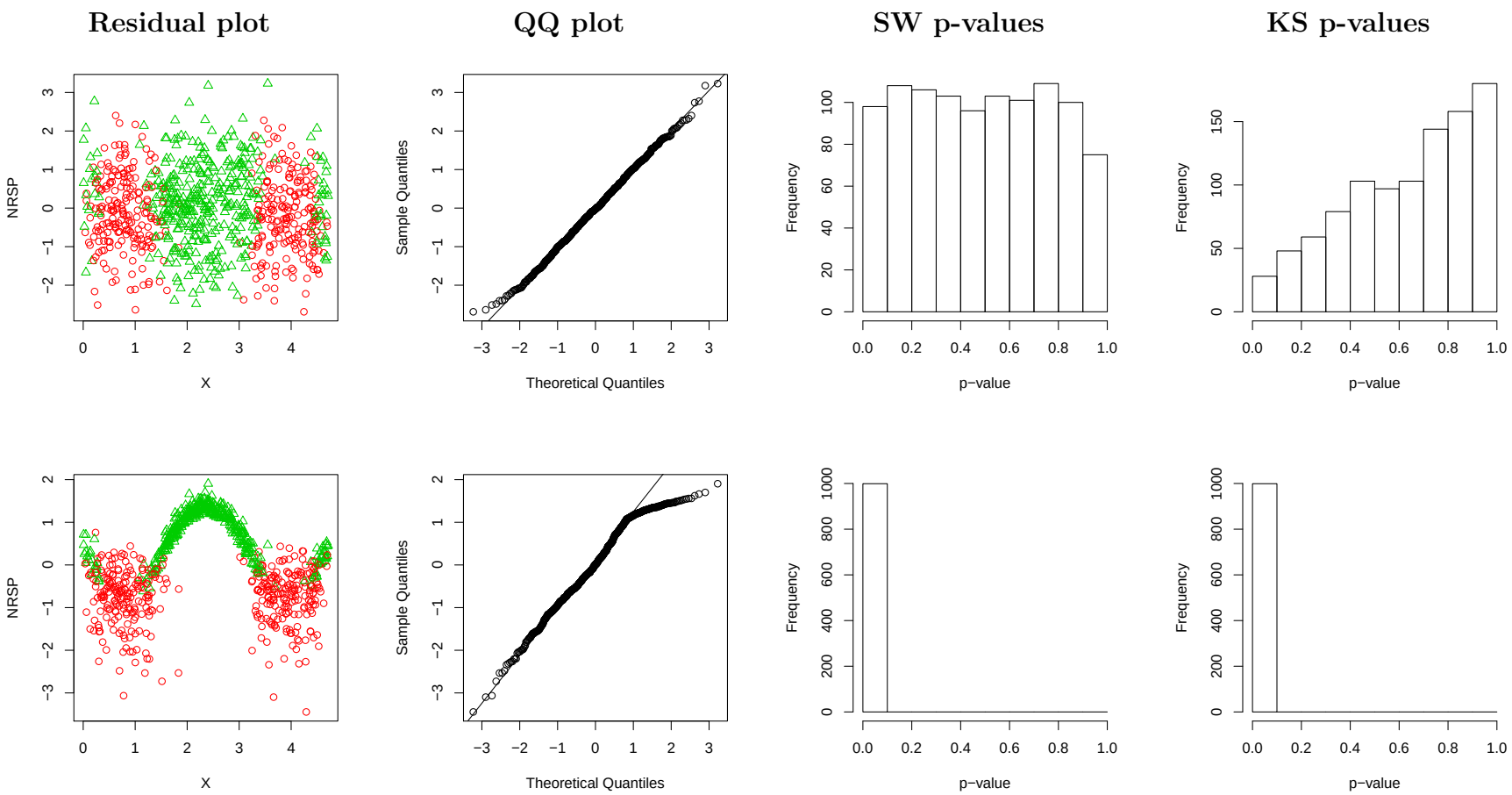

Figure 4.7: Performance of the NRSP residuals in detecting distributional assumption of a sample dataset of size $n = 800$ and a percentage of censorship $c = 50\%$. The panels in the first row present the NRSP residuals for the true model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 f(x) + \epsilon_i$. The panels in the second row present the NRSP residuals for the wrong model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$. The first two columns display the scatter plots and QQ plots of the NRSP residuals, respectively. The third and fourth columns present the histograms of the SW and KS p-values for the NRSP residuals over 1000 randomly generated datasets from the true model. The green triangles correspond to the event times and the red circles correspond to the censored times.

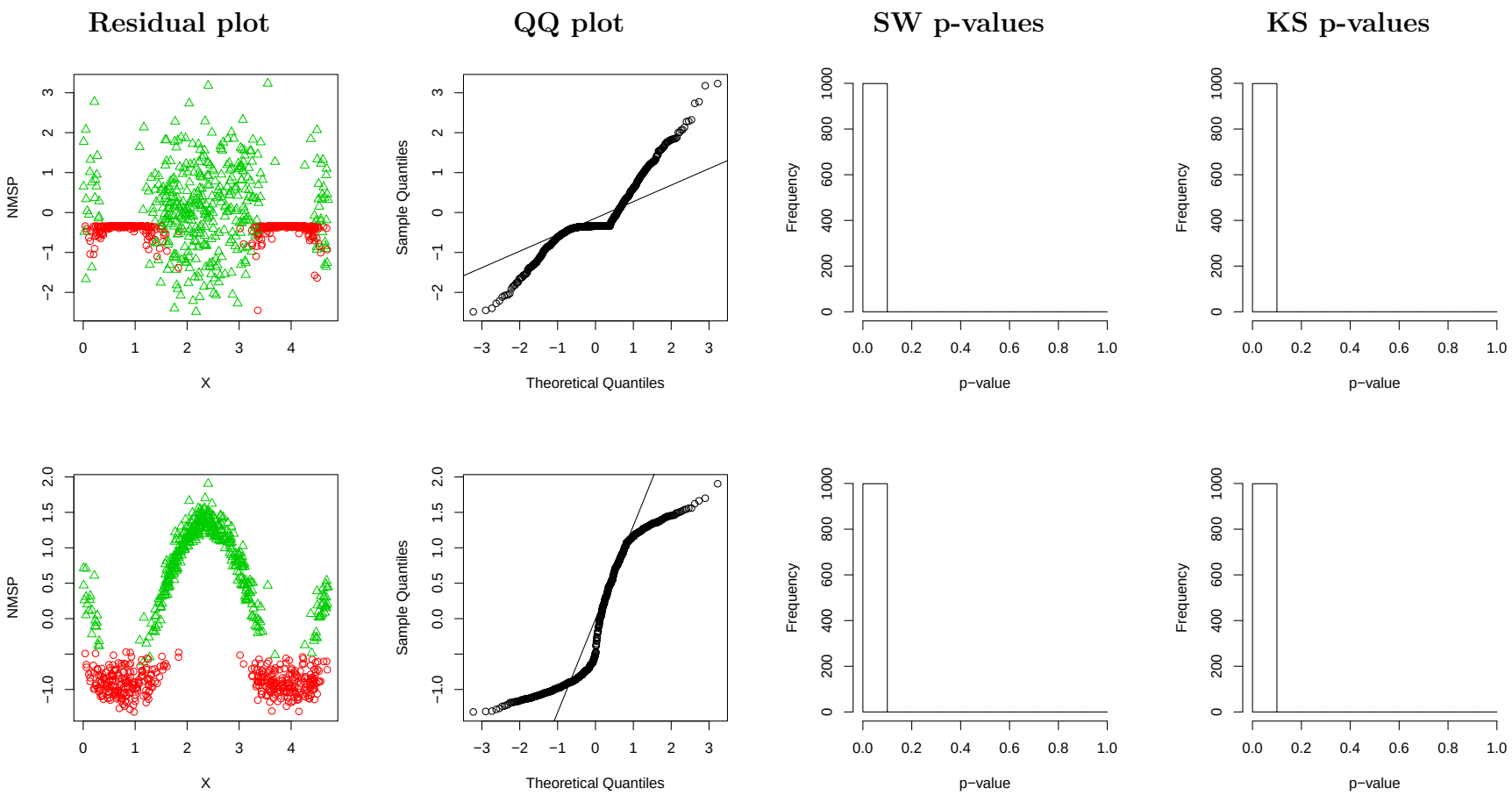

Figure 4.8: Performance of the NMSP residuals in detecting distributional assumption of a sample dataset of size $n = 800$ and a percentage of censorship $c = 50\%$. The panels in the first row present the NMSP residuals for the true model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 f(x) + \epsilon_i$. The panels in the second row present the NMSP residuals for the wrong model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$. The first two columns display the scatter plots and QQ plots of the NMSP residuals, respectively. The third and fourth columns present the histograms of the SW and KS p-values for the NMSP residuals over 1000 randomly generated datasets from the true model. The green triangles correspond to the event times and the red circles correspond to the censored times.

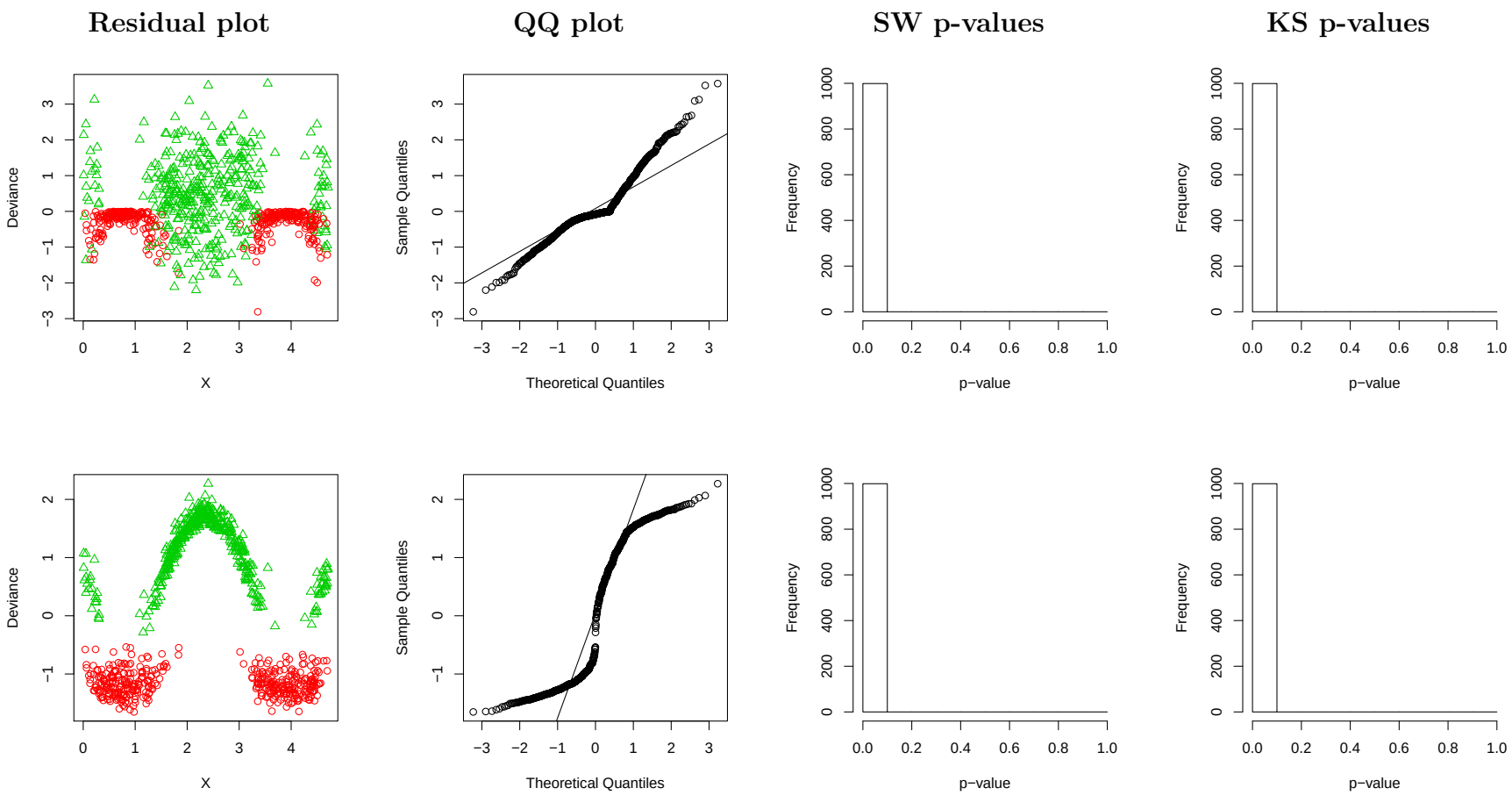

Figure 4.9: Performance of the deviance residuals in detecting distributional assumption of a sample dataset of size $n = 800$ and a percentage of censorship $c = 50\%$. The panels in the first row present the deviance residuals for the true model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 f(x) + \epsilon_i$. The panels in the second row present the deviance residuals for the wrong model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$. The first two columns display the scatter plots and QQ plots of the deviance residuals, respectively. The third and fourth columns present the histograms of the SW and KS p-values for the deviance residuals over 1000 randomly generated datasets from the true model. The green triangles correspond to the event times and the red circles correspond to the censored times.

4.2.2 Power analysis

To evaluate the finite-sample performance of the SW test for the NRSP residuals as the overall model diagnosis tool, power analysis is performed by setting the sample sizes at $n = 100, 200, 400, 600, 800, 1000$ and the percentage of censorship at $c = 20\%, 50\%$ and 80%. As shown in Figure 4.10, type I errors of the SW test for the NRSP residuals remain at the nominal level 0.05 for all scenarios. In contrast, the type I errors of the SW tests for the NMSP and deviance residuals are substantially higher than the 0.05 norminal level. Figure 4.10 also shows that statistical power at all scenarios for the NRSP, the NMSP and deviance residuals, which indicates that the NMSP and deviance residuals always reject both the wrong model and the true model all the time. As compared to the NMSP and deviance residuals, NRSP residuals is more superior with regard to its ability to identify the true model across all the considered sample sizes and percentage of censorship. To investigate the behavior of KS test for all of these residuals, the first column of Figure 4.11 indicates that the type I errors of the KS test for NRSP residuals are consistently lower than nominal level 0.05 for all scenarios, moreover, some of values are mostly close to zero under the true models. The statistical power of KS test for the NRSP residuals under wrong models are between 0 to 0.05 when sample size is small, or the percentage of censorship is larger. The type I errors and statistical power of the KS tests at all scenarios are above 0.05 in the NMSP and deviance residuals in the second and third column of Figure 4.11. This provides further evidence the unsatisfactory performance of the KS test.

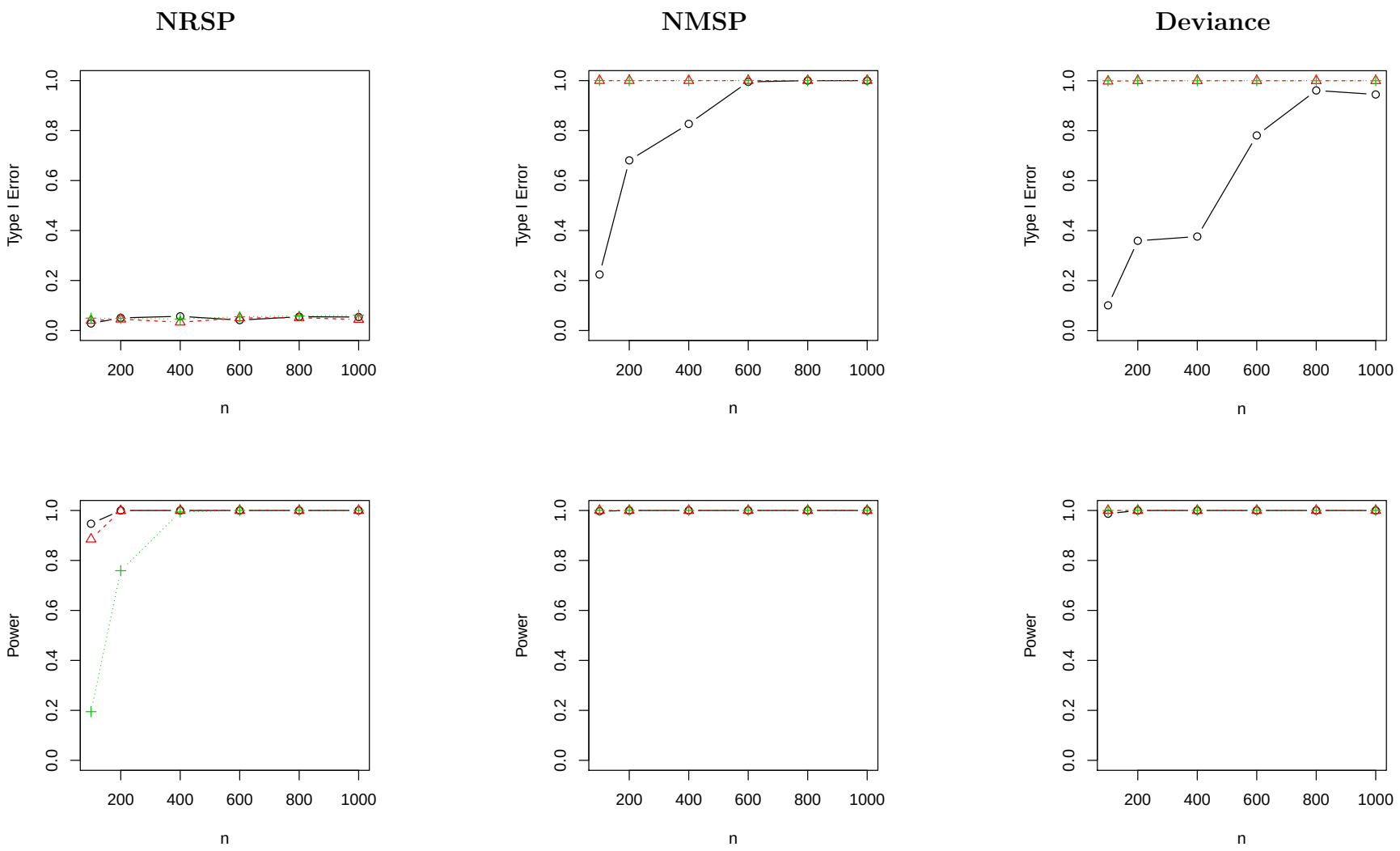

Figure 4.10: Comparison of the type I errors and powers of the SW tests for the NRSP, NMSP, and deviance residuals. Response variable is simulated from the true model at varying sample sizes of $n = 100$, 200, 400, 600, 800 and 1000, and the percentage of censorship $c = 20\%$ (black circles), 50% (red triangles) and 80% (green crosses). True model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 f(x) + \epsilon_i$. Wrong model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$.

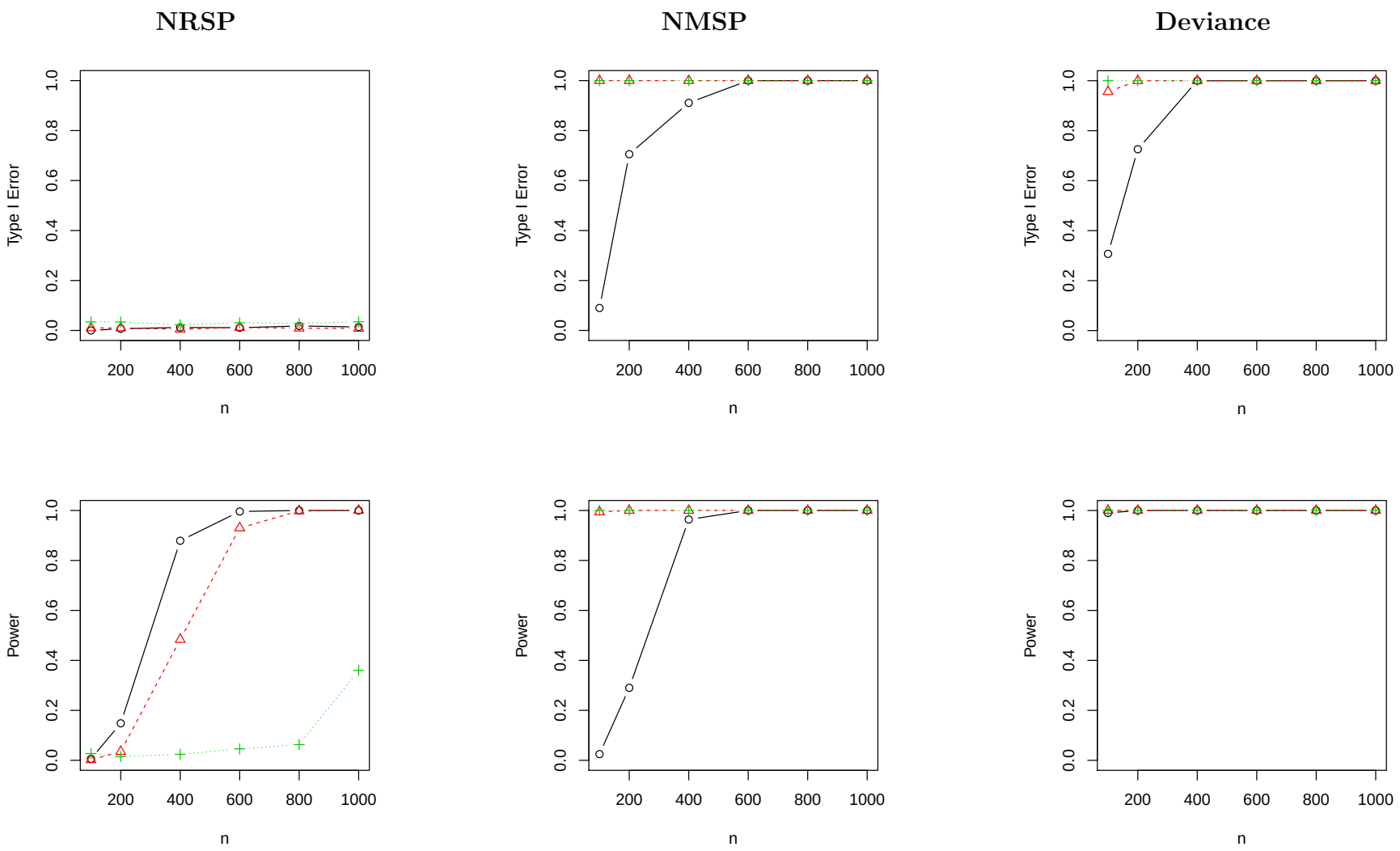

Figure 4.11: Comparison of the type I errors and powers of the KS tests for the NRSP, NMSP, and deviance residuals. Response variable is simulated from the true model at varying sample sizes of $n = 100$, 200, 400, 600, 800 and 1000, and the percentage of censorship $c = 20\%$ (black circles), 50% (red triangles) and 80% (green crosses). True model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 f(x) + \epsilon_i$. Wrong model: a Weibull AFT regression model $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$.

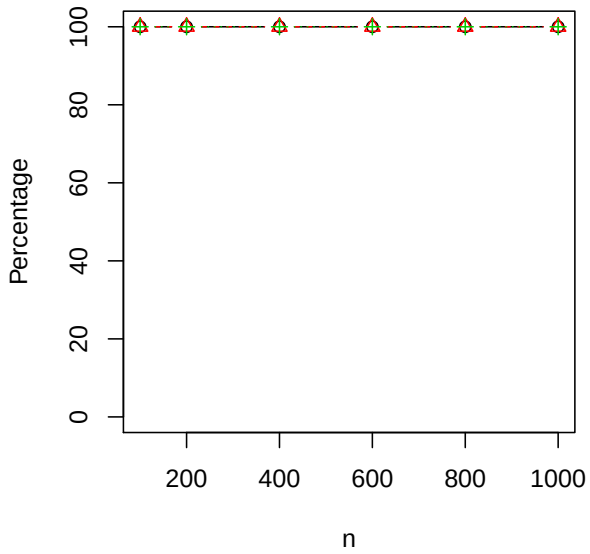

Figure 4.12: AIC for true model (Weibull model $\log(T_i) = \beta_0 + \beta_1 f(x) + \epsilon_i$) and wrong model (Weibull model $\log(T_i) = \beta_0 + \beta_1 x + \epsilon_i$) at varying sample sizes of $n = 100$, 200, 400, 600, 800 and 1000, and the percentage of censorship $c = 20\%$ (black circles), 50% (red triangles) and 80% (green crosses).

4.2.3 Model comparisons

To confirm the performance of the proposed residual diagnosis tool in comparison with traditional residuals in survival analysis, we further compare the true and wrong models base d on AIC in all the simulation settings. Figure 4.12 shows that the percentage of the difference value of AIC greater than 4 between the wrong and true models based on 1000 replicated samples. In all of the scenarios, the results are greater than 4 [20], which further confirms the current study's findings.

5. Real data analysis

In this Chapter, we will introduce a real application on the recurrence-free survival in breast cancer patients dataset [1], and apply the NRSP residual to examine the GOF of AFT models. A cohort study of breast cancer in a large number of hospitals was carried out by the German Breast Cancer Study Group to compare three cycles of chemotherapy with six cycles, and also to investigate the effect of additional hormonal treatment consisting of a daily dose of 30 mg of tamoxifen over two years [21]. The patients in the study had primary histologically proven non-metastatic node-positive breast cancer who had been treated with mastectomy. The response variable of interest is recurrence-free survival, which is the time from entry to the study until a recurrence of the cancer or death. Earlier analyses of the data had shown that recurrence-free survival was not affected by the number of cycles of chemotherapy, and so only the factor associated with whether or not a patient received tamoxifen is included in this example. In addition to this treatment factor, data were available on patient age, menopausal status, size and grade of the tumour, number of positive lymph nodes, progesterone and oestrogen receptor status. The data in this example relate to data from 41 centres and 686 patients with 56.5% censorship [22]. The variables in this dataset are presented in Table 5.1.

In this study, Weibull, Log-logistic and Lognormal AFT models with all of variables listed in Table 5.1 included as covariates are fitted to the recurrence-free survival in breast cancer dataset. We will firstly present the results based on the traditional residuals. Figure 5.1 displays residuals r_i^c against $-\log \hat{S}(r_i^c)$ under Weibull, Lognormal and Log-logistic AFT models. Under the Weibull and Lognormal models, a portion of plotted points deviate from the straight line. Similarly, under the Log-logistic model, most of the plotted points are not on a straight line. As a result, it is very challenging to distinguish which model fits the data most effectively, especially between the Weibull and Lognormal models, though the plot of Log-logistic model seems the least problematic. Moreover, there are no statistics tests with which this can be ascertained.

The NRSP residual is applied to examine the GOF of the Weibull, Log-logistic, and

Table 5.1: Variable definitions in the breast cancer study.

Variable	Definition
Time	Recurrence-free survival time (days)
Status	Event indicator (0 = censored, 1 = relapse or death)
Treat	Hormonal treatment (0 = no tamoxifen, 1 = tamoxifen)
Age	Patient age (years)
Men	Menopausal status (1 = premenopausal, 2 = postmenopausal)
Size	Tumour size (mm)
Grade	Tumour grade (1, 2, 3)
Nodes	Number of positive lymph nodes
Prog	Progesterone receptor status (femtomoles)
Oest	Oestrogen receptor status, (femtomoles)

Lognormal AFT models for the dataset. The panels in the first column of Figure 5.2 present the scatter plots of the NRSP residuals versus the fitted values for each model. The Lognormal model fits the dataset fairly well with residuals ranging between -3 and 3, as well as a random pattern present. In contrast, the Weibull model does not fit the dataset well with residuals ranging between -4 and 2. In addition, the Log-logistic model has residuals ranging between -3 and 3 with most points clustered between -2 and 2. The QQ plots of the NRSP residuals as presented in the panels of the second column of Figure 5.2, illustrate the inadequate fits of the Weibull and Log-logistic models. However, Lognormal model fits satisfactory to the data with almost all the points falling along the diagonal line.

One concern of using the NRSP residual method is the fluctuation in the residuals introduced. This is caused by the randomization of the survival probability for the censored observations. To determine impact of uncertainty due to randomization, 1000 realizations of the NRSP residuals are generated for the exact same dataset. The panels in the third column of Figure 5.2 display the histograms of 1000 replicated p-values of the SW tests. The p-values of the SW test for the NRSP residuals for the fitted Lognormal model varied between 0 and 1 with about 92.2% of the p-values being above 0.05. This confirms the adequacy of the Log-

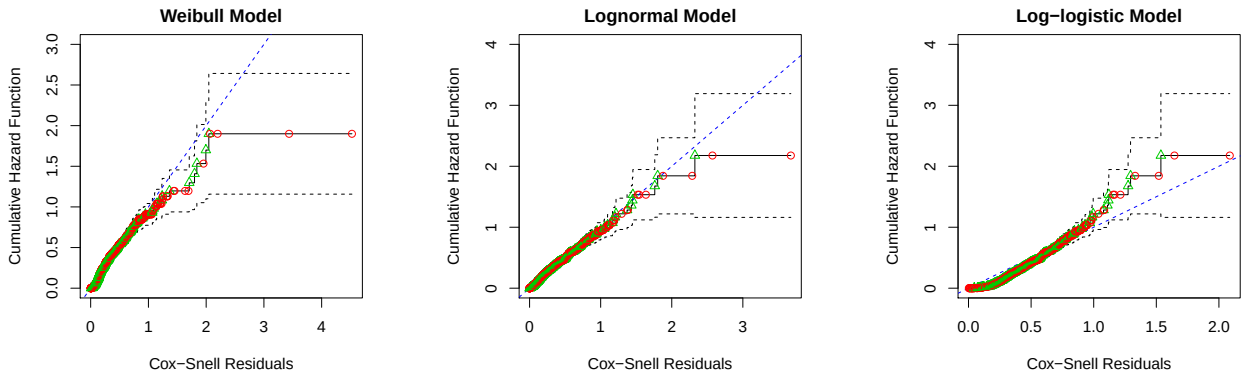

Figure 5.1: Cox-Snell residuals for Weibull, Log-logistic, and Lognormal models. The green triangles correspond to the event times and the red circles correspond to the censored times.

normal model. In contrast, under the Weibull and Log-logistic models, the p-values of the SW test for the NRSP residuals are concentrated around zero, confirming the inadequacies of both Weibull and Log-logistic models. Hence, randomization does not compromise much the statistical power of the NRSP residuals in this application.

The performance of the NMSP and deviance residuals with regard to detecting the Weibull, Log-logistic, and Lognormal AFT models for the breast cancer data analysis are also evaluated. The panels of the first column of Figures 5.3 and Figures 5.4 display the NMSP and deviance residuals against the fitted values. Under all the models, the NMSP residuals of event data are randomly scattered with residual bounded in [-1,3]. However, the residuals of censor data are clustered around -1 with residual bounded in [-2,0]. The deviance residuals perform very similarly as the NMSP residuals, with the results showing that there is no significant difference among the models. The QQ plots for the NMSP and deviance residuals are depicted in the panels of the second columns of Figures 5.3 and Figures 5.4. The NMSP and deviance residuals do not follow a normal distribution under all of models, and fail to diagnose the true model. Similarly, by the SW tests, the p-values for the NMSP and deviance residuals are very small, implying that all of models will be rejected at a small nominal threshold. Therefore, the NMSP and deviance residuals fail to distinguish models.

Table 5.2 contains the estimated regression coefficients, the corresponding standard errors and p-values for the covariates effects the Weibull, Lognormal, and Log-logistic models. The

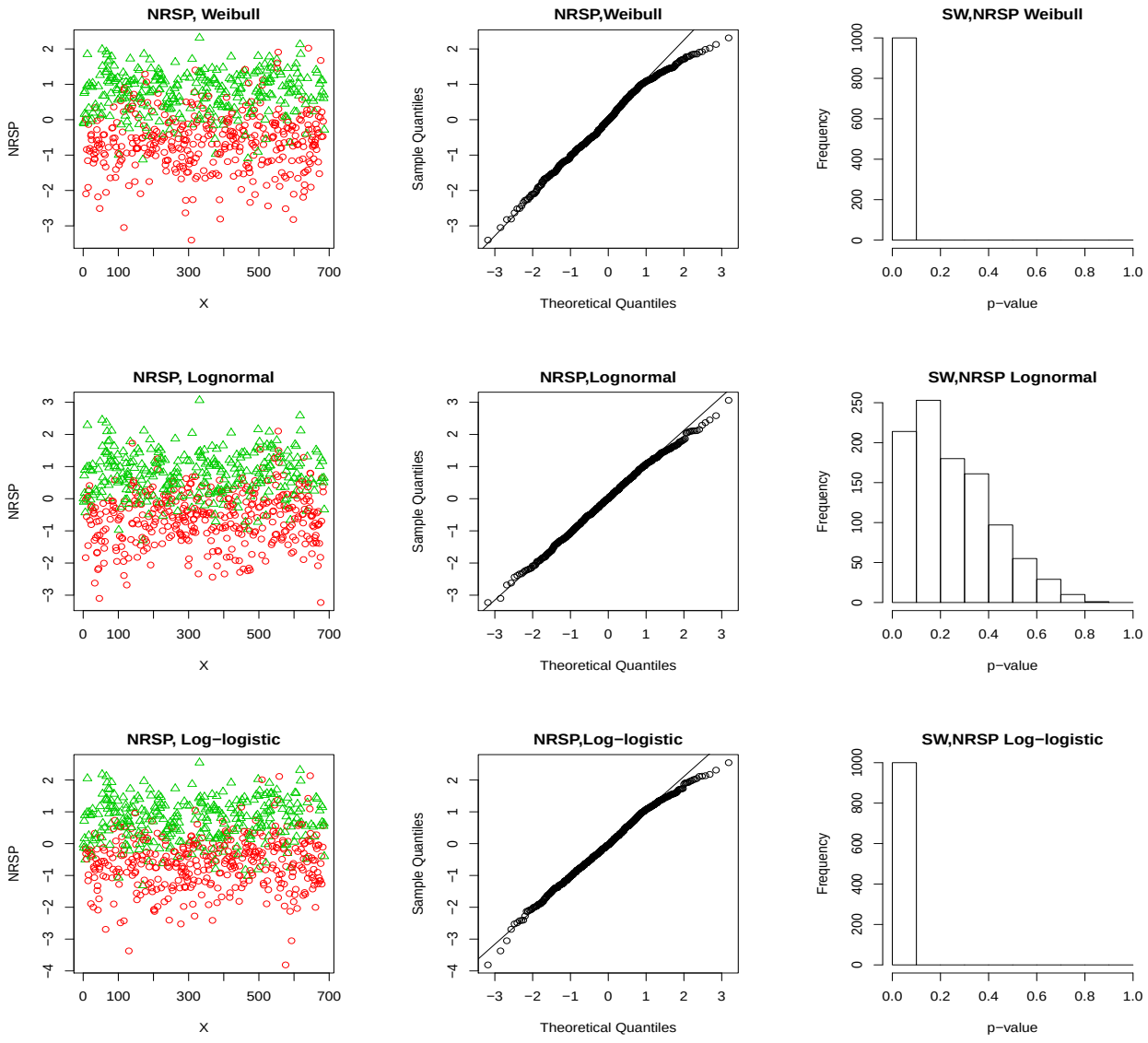

Figure 5.2: NRSP residuals for the Weibull, Log-logistic, and Lognormal AFT models fitted to the breast cancer patients dataset. The panels in the first two columns present the scatter plots and QQ plots of the NRSP residuals versus the fitted values, respectively. The green triangles correspond to the event times and the red circles correspond to the censored times. The third column presents the frequencies of the p-values of the SW normality test for 1000 replicated NRSP residuals.

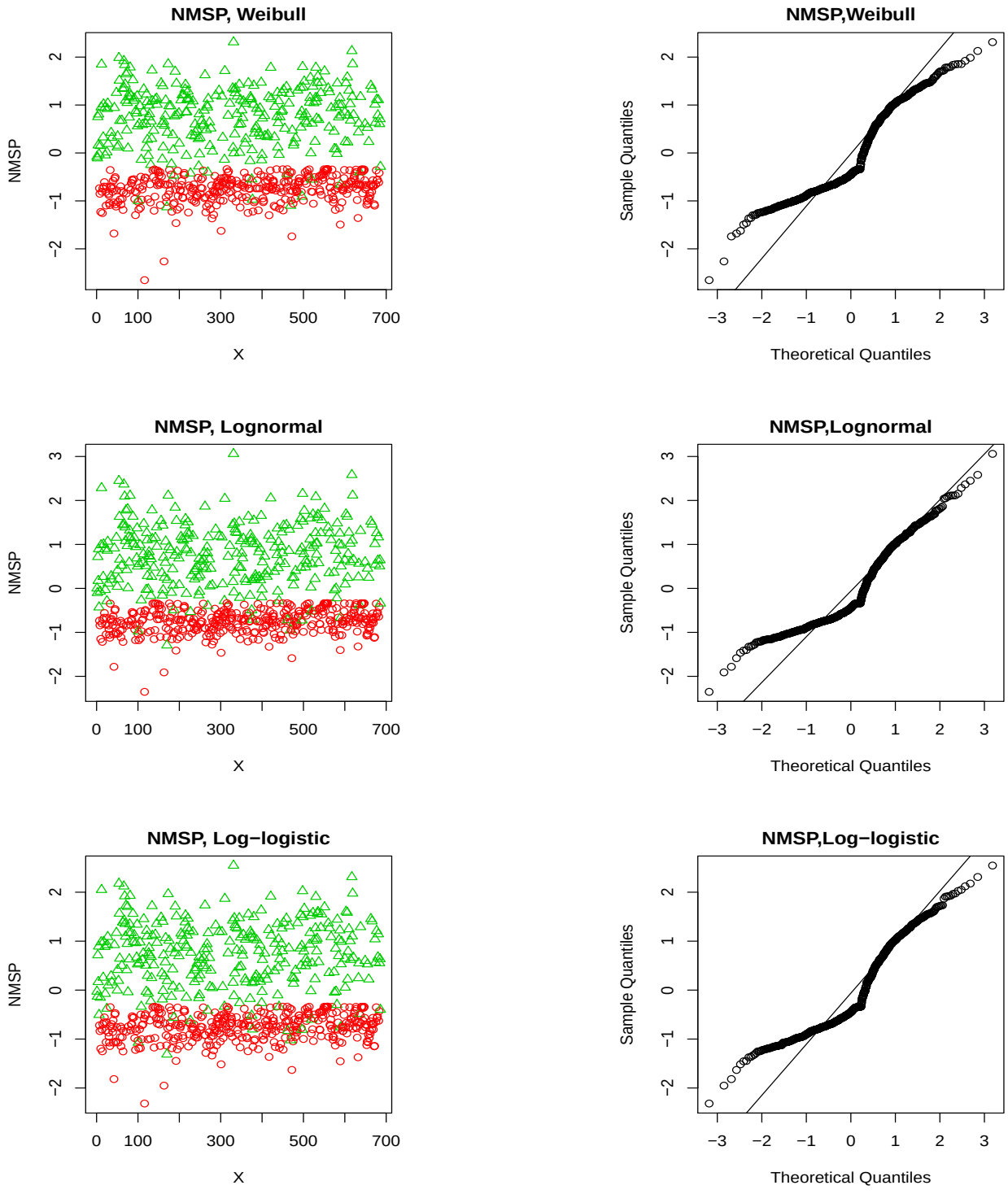

Figure 5.3: NMSP residuals for the Weibull, Log-logistic, and Lognormal AFT models fitted to the breast cancer patients dataset. The panels in the first two columns present the scatter plots and QQ plots of the NMSP residuals versus the fitted values, respectively. The green triangles correspond to the event times and the red circles correspond to the censored times.

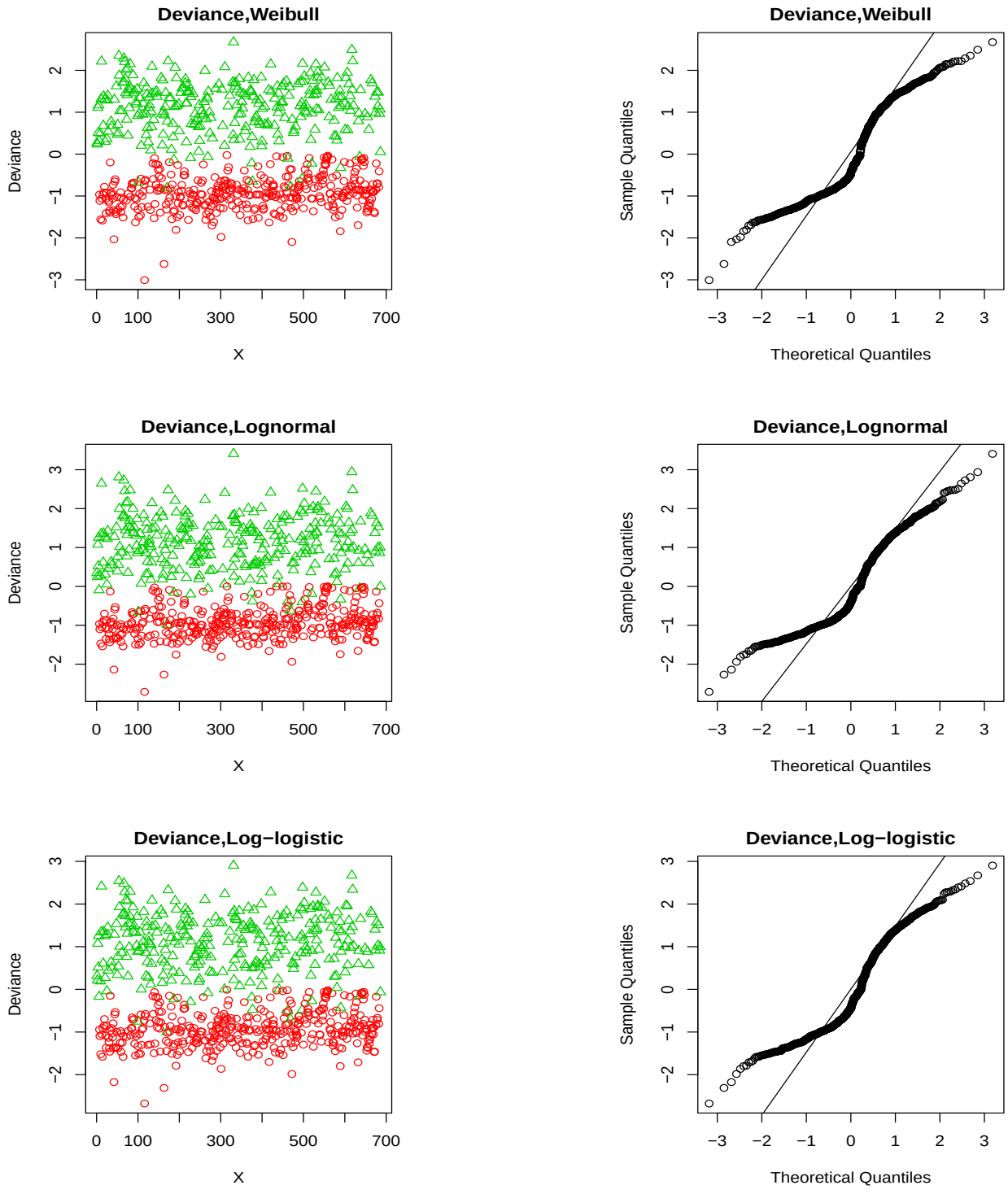

Figure 5.4: Deviance residuals for the Weibull, Log-logistic, and Lognormal AFT models fitted to the breast cancer patients dataset. The panels in the first two columns present the scatter plots and QQ plots of the deviance residuals versus the fitted values, respectively. The green triangles correspond to the event times and the red circles correspond to the censored times.

findings indicate that the choice of model distribution has a significant impact on estimating covariate effects. Table 5.3 displays the value of AIC statistic for the fitted Weibull, Log-normal, and Log-logistic models. Lognormal model yields the lowest AIC and therefore it provides a better fit to this data as compared to other models. Moreover, by repeatedly 1000 p-values of the SW test for NRSP residual, Table 5.3 shows the percentage of times that the p-values are less than 0.05 for all of three models. The results clearly demonstrate that the Lognormal model is a better model with only 7.8% of the p-values less than 0.05 for this application.

Table 5.2: Parameter estimates of the Weibull, log-normal and log-logistic models in Breast Cancer Study.

	Weibull			Log-normal			Log-logistic		
Covariates	Estimate	SE	P-value	Estimate	SE	P-value	Estimate	SE	P-value
Treat	0.261	0.093	0.005	0.309	0.097	0.002	0.321	0.097	0.001
Age	0.007	0.007	0.304	0.013	0.007	0.070	0.013	0.007	0.062
Men	-0.202	0.131	0.123	-0.260	0.143	0.070	-0.289	0.143	0.043
Size	-0.006	0.003	0.044	-0.006	0.003	0.052	-0.007	0.003	0.037
Grade	-0.211	0.076	0.006	-0.256	0.082	0.002	-0.230	0.082	0.005
Nodes	-0.039	0.005	<0.001	-0.051	0.008	<0.001	-0.052	0.008	<0.001
Prog	0.002	0.001	<0.001	0.001	<0.001	<0.001	0.002	<0.001	<0.001
Oest	<0.001	<0.001	0.635	<0.001	<0.001	0.886	<0.001	<0.001	0.862

Table 5.3: Percentages of P-values smaller than 0.05 for the SW test of the NRSP residuals and AIC comparisons for Weibull, Log-normal and Log-logistic models in the breast cancer data analysis.

Model fit	Weibull	Log-normal	Log-logistic
NRSP	100%	7.8%	99.3%
AIC	5182	5140	5154

6. Conclusion and future work

In this thesis, we proposed NRSP residual of diagnosing AFT models in survival analysis and computationally justified the normality of the proposed residual and compared its performance with the traditional residuals, including the Cox-Snell residuals and deviance residuals, through simulation studies and a real data application. This thesis reinforces that the traditional residuals are not well-calibrated and fail to assist in model diagnosis. However, NRSP residuals are well-calibrated and can be used for a wide range of distributions. It is computationally demonstrated that NRSP residuals are normally distributed, aside from the variability in the estimation of the parameters. This provides a unified way of simply plotting the NRSP residuals against predicted values or the covariates as well as their QQ-plots for visually checking the model adequacy. Meanwhile, according to the GOF test, the probabilities of rejecting the true model (type 1 error rates) are close to the nominal level 0.05, and the powers of rejecting the wrong models are high when the sample size of events is relatively large and the departure from the true model is not marginal. Another significant advantage of NRSP residuals over the traditional ones is their simple definition, which only requires knowing the CDF of the response variable. In conclusion, NRSP residual is an excellent tool that can be used to compare and diagnose AFT models in survival analysis.

For further study, random effects can be added in the AFT models. In the multicenter clinical trial, model center variation could then be added using a random effect, and there would be interest in investigating the performance NRSP residuals in different distributions [1]. Furthermore, Cox proportional hazard model is a very widely used survival model, as the baseline function can take on any forms. We will extend NRSP residuals to diagnose Cox proportional hazard models with and without random effects in the near future.

Bibliography

[1] David Collett. *Modelling Survival Data in Medical Research*. Chapman and Hall/CRC, 2015.

[2] David R. Cox and E. Joyce Snell. A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 248–275, 1968.

[3] Terry M. Therneau, Patricia M. Grambsch, and Thomas R. Fleming. Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160, March 1990. ISSN 0006-3444. doi: 10.1093/biomet/77.1.147.

[4] Terry M. Therneau and Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer Science & Business Media, November 2013. ISBN 978-1-4757-3294-8.

[5] Myles Hollander and Douglas Wolfe. *Nonparametric Statistical Methods, 2nd Edition*. Wiley-Interscience, January 1999. ISBN 978-0-471-19045-5.

[6] Elin Ansin. *An Evaluation of the Cox-Snell Residuals*. PhD thesis, 2015.

[7] Martina Müller. Goodness of fit criteria for survival data. page 29.

[8] Peter K. Dunn and Gordon K. Smyth. Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics*, 5(3):236–244, September 1996. ISSN 1061-8600. doi: 10.1080/10618600.1996.10474708.

[9] Cindy Feng, Alireza Sadeghpour, and Longhai Li. Randomized Quantile Residuals: An Omnibus Model Diagnostic Tool with Unified Reference Distribution. *arXiv:1708.08527 [stat]*, August 2017.

[10] Alessandra Nardi and Michael Schemper. Comparing Cox and parametric models in clinical studies. *Statistics in Medicine*, 22(23):3597–3610, December 2003. ISSN 1097-0258. doi: 10.1002/sim.1592.

[11] John D. Kalbfleisch and Ross L. Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, January 2011. ISBN 978-1-118-03123-0.

[12] John P. Klein and Melvin L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media, June 2013. ISBN 978-1-4757-2728-9.

[13] Jerald F. Lawless. *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, January 2011. ISBN 978-1-118-03125-4.

[14] Donald A. Pierce and Daniel W. Schafer. Residuals in Generalized Linear Models. *Journal of the American Statistical Association*, 81(396):977–986, December 1986. ISSN 0162-1459. doi: 10.1080/01621459.1986.10478361.

[15] P. McCullagh and John A. Nelder. *Generalized Linear Models, Second Edition*. CRC Press, August 1989. ISBN 978-0-412-31760-6.

[16] Alessandra Nardi and Michael Schemper. New Residuals for Cox Regression and Their Application to Outlier Screening. *Biometrics*, 55(2):523–529, June 1999. ISSN 1541-0420. doi: 10.1111/j.0006-341X.1999.00523.x.

[17] J. Qian, B. Li, and P. Chen. Generating Survival Data in the Simulation Studies of Cox Model. In *2010 Third International Conference on Information and Computing(ICIC)*, volume 04, pages 93–96, June 2010. ISBN 978-0-7695-4047-4. doi: 10.1109/ICIC.2010.294.

[18] Fei Wan. Simulating survival data with predefined censoring rates for proportional hazards models. *Statistics in Medicine*, 36(5):838–854, February 2017. ISSN 1097-0258. doi: 10.1002/sim.7178.

[19] Edson Zangiacomi Martinez, Jorge Alberto Achcar, Marcos Vinicius de Oliveira Peres, and Jose Andre Mota de Queiroz. A brief note on the simulation of survival data with a desired percentage of right-censored datas. *Journal of Data Science*, 14(4):701–712, 2016.

[20] M. Y. J. Tan and Rahul Biswas. The reliability of the Akaike information criterion method in cosmological model selection. *Monthly Notices of the Royal Astronomical Society*, 419(4):3292–3303, February 2012. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.19969.x.

[21] M Schumacher, G Bastert, H Bojar, K Hübner, M Olschewski, W Sauerbrei, C Schmoor, C Beyerle, R L Neumann, and H F Rauschecker. Randomized 2 x 2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. German Breast Cancer Study Group. *Journal of Clinical Oncology*, 12(10):2086–2093, October 1994. ISSN 0732-183X. doi: 10.1200/JCO.1994.12.10.2086.

[22] W. Sauerbrei and P. Royston. Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(1):71–94, January 1999. ISSN 1467-985X. doi: 10.1111/1467-985X.00122.